

RESEARCH ARTICLE

Non-canonical genomic driver mutations of urethane carcinogenesis

Siqi Li[‡], Christopher M. Counter^{‡*}

Department of Pharmacology & Cancer Biology, Duke University, Durham, NC, United States of America

[‡] Current address: Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States of America* count004@mc.duke.edu

Abstract

The carcinogen urethane induces pulmonary tumors in mice initiated by an incredibly specific Q₆₁L/R oncogenic mutation in the proto-oncogene *Kras*. Previous Whole-Exome Sequencing of urethane-induced tumors revealed a bias towards A→T/G and G→A substitutions. Subsequent ultra-sensitive Maximum-Depth Sequencing of *Kras* shortly after urethane exposure suggest a further refinement to **CA→CT/G** substitutions. As C₁₈₂AA→C₁₈₂T/GA substitutions in *Kras* result in Q₆₁L/R mutations, the extreme bias of urethane towards these genomic driver mutations can be ascribed to the specificity of the carcinogen for **CA→CT/G** substitutions. However, we previously found that changing rare codons to common in the *Kras* gene to increase protein expression shifted mutations in urethane-induced tumors away from *Kras*, or when detected in *Kras*, to G12D mutations that are usually rarely detected in such tumors. Moreover, the loss of p53 partially reversed this effect, generating tumors with either Q₆₁L/R or G12D oncogenic *Kras* mutations, or no *Kras* mutations, presumably due to other genomic driver mutations. Determining the origin of these G12D and other unknown non-canonical genomic driver mutations would provide critical insight into the extreme bias of carcinogens for specific genomic driver mutations. We thus compared the types of Single Nucleotide Variations detected by previously performed Maximum-Depth Sequencing immediately after urethane exposure to the mutation signatures derived from Whole Exome Sequencing of urethane-induced tumors. This identified two types of non-canonical mutations. First, a V₆₃₇E oncogenic mutation in the proto-oncogene *Braf* that conforms to the mutation signature of urethane, suggesting that the mutational bias of the carcinogen may account for this non-canonical mutation, similar to that for canonical Q₆₁L/R mutations in *Kras*. Second, G₁₂D and Q₆₁H mutations in *Kras* that did not fit this mutation signature, and instead shared similarity with Single Nucleotide Variations detected by Maximum-Depth Sequencing from normal cells, suggesting that perhaps these mutations were pre-existing. We thus posit that when canonical *Kras* mutations are selected against that the carcinogen may instead promote the expansion of pre-existing genomic driver mutations, although admittedly we cannot rule out other mechanisms. Interrogating the mutation signatures of human lung cancers similarly identified *KRAS* genomic driver mutations that failed to match the mutation signature of the tumor. Thus, we also speculate that the selection for non-canonical genomic driver mutations during urethane

OPEN ACCESS

Citation: Li S, Counter CM (2022) Non-canonical genomic driver mutations of urethane carcinogenesis. PLoS ONE 17(4): e0267147. <https://doi.org/10.1371/journal.pone.0267147>

Editor: Shawn Ahmed, University of North Carolina at Chapel Hill, UNITED STATES

Received: December 3, 2021

Accepted: April 3, 2022

Published: April 28, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0267147>

Copyright: © 2022 Li, Counter. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw sequencing data for WES datasets was deposited to NCBI Sequence Read Archive (SRA) under accession number PRJNA663179. MDS datasets were previously generated [7,11] and the raw sequencing data was deposited to NCBI Sequence

Read Archive (SRA) under accession number PRJNA561927 and PRJNA663179.

Funding: This work was supported by the National Cancer Institute (USA) grants R01CA123031 and P01CA203657 as well as an internal voucher grant from the Duke University School of Medicine to CMC. The Duke Cancer Institute Sequencing and Genomic Technologies Shared Resource that performed sequencing for this study is supported by P30CA014236. There was no additional external funding received for this study.

Competing interests: CMC is co-Founder of Merlon Inc, a member of the External Advisory Board for the University of Colorado Cancer Center, has a cross appointment with Duke-NUS, is an ex officio of the executive team for the Cancer Biology Training Consortium, has previously consulted for the Guidepoint Network in an ad hoc fashion and received licensing reimbursement from Humacyte Inc in the past. These relationships did not provide any salary or research support, did not play any role in the study design, data collection and analysis, decision to publish, and do not alter our adherence to PLOS ONE policies on sharing data and materials.

carcinogenesis may reflect the process by which discordance between genomic driver mutations and mutational signatures arises in human cancers.

Introduction

Intraperitoneal injections of urethane, a carcinogen detected in fermented and alcoholic products [1], induces pulmonary tumors in mice characterized by either a $Kras^{Q61L}$ or $Kras^{Q61R}$ genomic driver mutation, depending on the mouse strain [2]. Urethane is known to induce ethenodeoxyadenosine adducts [3, 4]. Consistent with this type of DNA damage, Whole Exome Sequencing (WES) of urethane-induced tumors revealed a bias towards $A \rightarrow T$ transversions as well as $A \rightarrow G$ and $G \rightarrow A$ transitions [5]. Capitalizing upon the development of the ultra-sensitive Maximum Depth Sequencing (MDS) to detect *de novo* mutations in bacteria [6], we previously adapted this assay to screen for Single Nucleotide Various (SNV) in the *Kras* gene of murine lungs shortly after the animals were exposed to urethane. This analysis revealed a more restricted mutation signature of a C followed by an $A \rightarrow T$ transversion, and to a lesser extent, an $A \rightarrow G$ transition [7]. Importantly, $CA \rightarrow CT/G$ substitutions convert codon $C_{182}AA$ encoding Q_{61} in *Kras* to the $C_{182}TA$ ($Q_{61}L$) and proportionally the less common $C_{182}GA$ ($Q_{61}R$) oncogenic mutations found in urethane-induced tumors of the A/J strain of mice. Moreover, previous WES analysis of urethane-induced tumors failed to identify other highly recurrent co-operating mutations, arguing that $Kras^{Q61L/R}$ mutations are indeed the genomic driver mutations in urethane carcinogenesis [7]. Collectively, these data support a model whereby the mutational preference of this carcinogen underlies the extreme bias for the $Kras^{Q61L/R}$ genomic driver mutations characteristic of urethane-induced tumors [5, 7, 8].

We previously reported that the *Kras* gene of mice (and humans) is enriched in rare codons, and further, that this bias towards rare codons reduces protein levels [9]. We further found that converting 27 rare codons to their common counterparts in the third coding exon of the endogenous *Kras* murine gene (termed *Kras^{ex3op}*), which does not contain the sites for oncogenic mutations, not only reduced the number of urethane-induced tumors with a *Kras* genomic driver mutation, but in those tumors in which *Kras* was mutated, a completely different oncogenic mutation ($G_{12}D$) was recovered [10, 11]. We suggest that these latter mutations arose due to selection of the less active $G_{12}D$ mutation in the more highly expressed *Kras^{ex3op}* allele to avoid oncogenic stress. Indeed, high oncogenic RAS expression is known to induce senescence [12], a $G_{12}D$ mutation exhibits lower levels of both active GTP-bound *Kras* [10, 13] and activation of downstream target genes [11] compared to a $Q_{61}R$ mutation, and loss of the tumor suppressor *Trp53* to suppress oncogenic stress generates urethane-induced tumors with a $Q_{61}R$ (in addition to $G_{12}D$) mutation in the *Kras^{ex3op}* allele [11]. Of special note, a $G_{12}D$ mutation arises from a $GG_{35}T \rightarrow GAT$ transversion that does *not* match the aforementioned $CA \rightarrow CT/G$ mutation signature of urethane-induced tumors.

Given that these mutations were discordant with the mutation signature of urethane, and that some tumors arose without an oncogenic mutation in *Kras*, we sought to identify the types of non-canonical mutations induced by urethane and then their potential origins, to elucidate how the extreme bias of urethane for $Kras^{Q61L/R}$ mutations is reprogrammed. We thus compared the SNV previously detected by MDS in the *Kras* gene shortly after urethane exposure against the mutation signature derived by WES analysis of urethane-induced tumors. This identified two types of non-canonical genomic driver mutations. First, genomic driver mutations that tracked with the mutation signature of urethane, suggesting that the mutagenic

activity of the carcinogen induced these mutations. Second, genomic driver mutations that tracked with SNVs detected in normal cells, speculatively suggesting that perhaps the carcinogen may instead have acted to promote a pre-existing driver mutation.

Materials and methods

Generation of WES datasets

Urethane induced tumors were isolated at the time of necropsy from 23 urethane-induced lung tumors from *Sftpc*^{CreER/CreER}; *Trp53*^{fl/fl} mice [11] injected with tamoxifen with none (6 tumors), one (10 tumors), or two (7 tumors) *Kras*^{ex3op} alleles (S1 Table). As previously described [11], mice (6–8 weeks old, male and female) were injected intraperitoneally with tamoxifen (Sigma T5648) dissolved in corn oil at a dose of 0.25 mg/g body weight every other day for a total of four doses. One week later all mice were injected intraperitoneally once with urethane (Sigma U2500) dissolved in sterile PBS at a dose of 1 mg/g body weight. Animals were maintained under pathogen-free conditions and visually monitored and weighed weekly. Approximately 12 months after urethane injection, mice were humanely euthanized by CO₂ asphyxiation followed by cervical dislocation, after which the lung tumors were harvested. Genomic DNA was isolated from the tumors as previously described [11]. 50 µg of this DNA was sheared to ~400 bp fragments using the Covaris E210 system. Libraries were generated with the Twist Mouse Exome Panel (Twist Bioscience 101242), hybridized, and enriched according to the manufacturer's instructions. Libraries were pooled and sequenced through 150 bp paired-end reads on an Illumina NovaSeq 6000 S-Prime flow cell. Analysis of WES data was performed on the Galaxy platform [14]. Sequencing adaptor was trimmed from the raw sequencing reads using Trimmomatic [15]. Trimmed reads were aligned to GRCm38/mm10 version of the *Mus musculus* genome using BWA-MEM [16]. Successfully mapped reads with a minimal mapping quality of 1, and for which the mate read has also been mapped were selected using the tool 'Filter BAM datasets on a variety of attributes' [17]. PCR duplicates were removed using RmDup [18]. Genomic regions targeted by the Twist Mouse Exome Panel were selected using the tool 'Slice BAM by genomic regions' [19] and bed dataset 'Twist_Mouse_Exome_Target_Rev1_7APR20.bed' from Twist Bioscience website (<https://www.twistbioscience.com/resources/bed-file/twist-mouse-exome-panel-bed-file>). Pileup file was generated from the BAM dataset using the tool 'Generate pileup from BAM dataset' [20]. On average, samples were sequenced with a sequencing depth of 90× with 74% of the exome covered by >20×. The raw sequencing data for WES datasets was deposited to NCBI Sequence Read Archive (SRA) under accession number PRJNA663179. Animal studies were approved in writing by the Duke University Institutional Animal Care Committee.

Generation of MDS datasets

MDS datasets were previously generated [7, 11] and the raw sequencing data was deposited to NCBI Sequence Read Archive (SRA) under accession number PRJNA561927 and PRJNA663179.

Variant identification, prioritization, and validation

Copy number alterations were identified using CNVkit-v0.9.10 [21]. SNVs were identified using VarScan [22]. GRCm38/mm10 served as the reference during calling and the minimum read depth for variants was set at 8. Called variants were annotated using SnpEff eff [23]. SNVs with variant allele frequency (VAF) < 5% were discarded. Variant calls from different samples were combined into a merged set using 'bcftools merge' [20]. The following SNVs were

considered as SNPs and removed: SNPs reported by the Mouse Genome Project of the Sanger Institute in the 129S1_SvImJ strain background [24]; SNVs appearing together in the same group of samples; SNVs clustering together by genomic coordinates in individual samples. High-likelihood cancer genomic driver genes were initially prioritized from a list of 460 previously defined cancer genes [25]. To focus on clonal genomic driver mutations, non-synonymous SNVs in the selected cancer genes were further filtered by VAF higher than the mean-3SD of the VAFs of *Kras* mutations determined from *Kras*-mutant tumors (15.7%). SNVs with VAF of 100% were considered as a SNP and removed.

To expand the list of putative genomic driver mutations, non-synonymous SNVs in all genes were filtered by VAF between the mean-3SD and mean+3SD of the VAFs of *Kras* mutations determined from *Kras* mutant tumors (15.7%-72.2%). SNVs clustering in the same gene in the same sample were removed. SNVs were further limited to genes for which the human homolog has single non-synonymous SNV occurring in at least five cancer patients in the ICGC data portal [26]. SNVs with VAF outside the 150% or 50% range of the VAF of *Kras* or *Braf*^{V637E} mutations in *Kras*- or *Braf*^{V637E}-mutant samples were removed. SNVs occurring in more than eight samples were considered as SNPs and removed. SNVs were further limited to amino acid positions that are conserved and mutated in human and/or mutated to the same amino acid in human to shorten the list of putative genomic driver mutations for each tumor to ~5. This list of SNVs were compiled with the SNVs selected from defined cancer genes above that occur at amino acid positions mutated in cancer patients. SNVs with VAF < 65% of the VAF of *Kras* or *Braf*^{V637E} mutations in *Kras*- or *Braf*^{V637E}-mutant tumors, or the putative genomic driver mutation with the highest VAF in the other tumors, were removed.

Validation of SNVs

SNVs were validated by Sanger sequencing of the PCR product from either genomic DNA or cDNA. SNVs were called validated if they were detected in the tumor but not in matched normal lung tissue. Sanger validation was prioritized for mutations recurrent across tumors as well as mutations with the highest VAF in each tumor. The primers used were listed in [S4 Table](#) and validation results were shown in [S2 Table](#).

Comparison of the mutation spectrum of MDS and WES

For the mutation spectrum of MDS, substitutions detected in *Kras* exon 1 and exon 2, *Hras* exon 2 in the lungs of A/J strain of mice one week after urethane or PBS exposure in a previous study [7] and well as substitutions detected in *Kras* exon 1 and exon 2 in the lungs of 129 strain of mice one week after urethane or PBS exposure in another study [11], were compiled together ([S3 Table](#)). A previously published study has shown that the C→T or G→T substitutions from the strand sequenced by MDS assay has a high mutation frequency not reflected in the complementary strand [7], thus those substitutions were considered as potential technical artifacts and removed from further analysis. Substitutions with frequency higher than 2×10^{-5} were considered as outliers and removed. Substitutions were annotated by the 96 possible trinucleotide context substitutions (6 types of substitutions \times 4 possible flanking 5' bases \times 4 possible flanking 3' bases) and the mutation frequency for each one of the 96 substitutions averaged across all nucleotide positions in all mice were plotted. The average mutation frequency calculated PBS-treated mice were multiplied by 10^6 and used as mutagenic optimality score for PBS in [Figs 1A](#) and [3B](#). The average mutation frequency of urethane-treated mice was normalized to PBS-treated mice by deducting the average mutation frequency of PBS-treated mice from that of urethane-treated mice for each of the 96 substitutions. This normalized mutation frequency was multiplied by 10^6 and used as mutagenic optimality score for

Fig 1. Identification of non-canonical genomic driver mutations. (A) Putative genomic driver mutations determined from WES of tumors, plotted based on their mutagenic optimality determined by MDS for urethane-treated normalized to PBS-treated mouse lung tissues versus for PBS-treated mouse lung tissues. Red highlight mutations occurring in at least one human cancer patient. Thick circles denote mutants validated by Sanger sequencing. (B) String analysis of proteins with a putative genomic driver mutation identified from urethane-induced mouse lung tumors. Limited to those with mutations with a high VAF at amino acid positions that are mutated in at least one human cancer patient in genes with single mutation observed in at least five human cancer patients. Dark circle highlights mutations occurring in multiple mouse lung tumors.

<https://doi.org/10.1371/journal.pone.0267147.g001>

‘urethane—PBS’ in Figs 1A and 3B. For the mutation spectrum of WES, SNVs in all tumors were annotated by the 96 possible trinucleotide context substitutions and summed in each tumor. These counts were then converted to per tumor proportions and the average across all tumors were calculated and plotted.

Mutations signature analysis of *KRAS*-mutant human lung cancer patients

Contribution to each mutation class by each mutation signature was based on the file ‘COSMIC_Mutational_Signatures_v3.1’ downloadable from COSMIC (<https://cancer.sanger.ac.uk/cosmic/signatures/SBS/index.tt>) [27]. Contribution of mutation signature in each tumor were extracted from the files “PCAWG_sigProfiler_SBS_signatures_in_samples.csv” and “TCGA_WES_sigProfiler_SBS_signatures_in_samples.csv” under synapse ID syn11804040 at Synapse (<https://www.synapse.org/>). *KRAS* mutations detected in each tumor were extracted from through ICGC Data Portal [26] by cross-referencing the donor ID matching the sample ID in the files “PCAWG_sigProfiler_SBS_signatures_in_samples.csv” and “TCGA_WES_sigProfiler_SBS_signatures_in_samples.csv” with the donor ID of *KRAS*-mutant LUADs. For each *KRAS*-mutant tumor, the contribution of individual signature to each substitution type in that tumor was calculated by 1) converting the mutation counts from each signature to per tumor proportions; 2) multiplying the calculated per tumor proportions by the percentage of contribution to different substitutions by each signature recorded in the file ‘COSMIC_Mutational_Signatures_v3.1’; 3) normalizing the calculated value so the sum of the values from all mutation signatures for each substitution in a tumor is 1. The calculated contribution of individual signature to the substitution matching the *KRAS* mutation in that tumor were plotted in Fig 5B. The signature mismatch score in Fig 5B was calculated as the contribution to the substitution matching *KRAS* mutation in that tumor minus the contribution to that tumor overall by the signature contributing the most to the substitution matching the *KRAS* mutation in that tumor. To calculate the normalized mutation burden for different substitutions in Fig 5A, the mutation counts from individual signatures was first estimated by multiplying the mutation counts for each signature in that tumor by the contribution of individual signature to each substitution type in that tumor. Counts from all signatures in that tumor were then summed for each substitution type and converted to per tumor proportions of the counts for all substitutions. The per tumor proportion for the substitution matching the *KRAS* mutation in that tumor was used as concordance score.

Results

Identification of non-canonical genomic driver mutations

To identify non-canonical genomic driver mutations of urethane carcinogenesis, we performed WES analysis on a panel of 23 tumors derived from urethane-exposed mice with none, one, or two of the aforementioned *Kras*^{ex3op} alleles upon recombination of the two *Trp53*^{fl} alleles specifically in the lung (S1 Table), as loss of p53 yields tumors with either canonical Q₆₁L/R or non-canonical G₁₂D *Kras* mutations in these backgrounds [11]. The resulting

datasets were then screened for nonsynonymous SNV in any gene with a Variant Allele Frequency (VAF) within three standard deviations of the mean of that detected for all *Kras* mutations, suggestive of a truncal mutation (S2 Table). These were then censored for having at least one mutation detected in the human counterpart from at least five human cancer patients in ICGC data portal [26], suggestive of a tumor-related gene (Fig 1A). Following this, candidate genes were screened for mutations at the same codon in human tumors, suggestive of a genomic driver mutation (S2 Table). This revealed 0 to 5 putative genomic driver mutations per tumor, so we further defined a putative initiating genomic driver mutation as having the identical mutation in more than ten human cancer patients. As expected, this analysis identified the substitutions giving rise to the canonical Q₆₁L and Q₆₁R mutations in *Kras*, as well as three other non-canonical mutations, the previously described G₁₂D mutation in *Kras*, a novel Q₆₁H mutation in *Kras*, and finally, a novel Bra^{V637E} mutation. Outside of these five mutations, a putative initiating genomic driver mutation could not be identified in the remaining tumors by these criteria. Furthermore, STRING analysis of all identified mutant proteins did not identify any obvious relationship beyond the MAPK pathway (Fig 1B). Censoring for any mutation in a panel of 460 genes associated with human cancers [25] found a CAN → CT/GN mutation with a high VAF in *Ctnnb1* and in *Spn* in one *Kras*/*Braf* mutation-negative tumor each (Fig 2), but these mutations were not conserved in human cancers. Thus, we confirm the previous identification of the non-canonical G₁₂D mutation in *Kras*, but now extend this to include another non-canonical *Kras* mutation, Q₆₁H, and finally, identify Bra^{V637E} as a potential new initiating genomic driver mutation in urethane-induced tumors, thereby expanding the mutational bias of this carcinogen to yet another oncogene activating the MAPK pathway.

Two types of non-canonical genomic driver mutations

To explore the relationship of the non-canonical mutations with urethane mutagenesis, we plotted the frequency of all possible SNVs determined by WES analysis of tumors, which we term the ‘WES mutation signature’ for ease of discussion. To determine the urethane mutation signature, we similarly plotted all SNVs from previous generated MDS sequencing data [7, 11] of *Kras* exon 1 and exon 2 as well as from *Hras* exon 2 from the lungs from mice exposed to urethane normalized to the same MDS sequencing analysis of the lungs of mice injected with the vehicle PBS (S3 Table), which we term the ‘urethane-PBS mutation signature’. To parse out contributions by other possible mutagenic events in the urethane-PBS signature, we further we plotted the SNVs from the raw MDS sequencing data for just urethane, which we term the ‘urethane mutation signature’, and for just PBS, which we term the ‘PBS mutation signature’ (Fig 3A). We next compared the optimality of all putative genomic driver mutations detected in the tumors by WES analysis to the PBS versus urethane-PBS mutation signatures. As expected, the Q₆₁L and to a lesser degree Q₆₁R mutations in *Kras* were more similar to the urethane-PBS mutation signature. Similarly, the Bra^{V637E} mutation was a better match to this signature as well. On the other hand, the G₁₂D and Q₆₁H mutations in *Kras* tracked with the PBS signature (Figs 1A and 3B). Thus, we identified genomic driver mutations both matching and not matching the mutation signature of urethane.

The non-canonical Bra^{V637E} genomic driver mutation matches the mutation signature of urethane

As noted above, a Bra^{V637E} mutation was identified in five tumors (Figs 1A and 3B), which corresponds to the oncogenic BRAF^{V600E} mutation [28] found in ~2% of human lung adenocarcinomas [26]. As a Bra^{V637E} mutation had not previously been ascribed to urethane carcinogenesis, but has been detected in liver or lung tumors induced by other carcinogens or

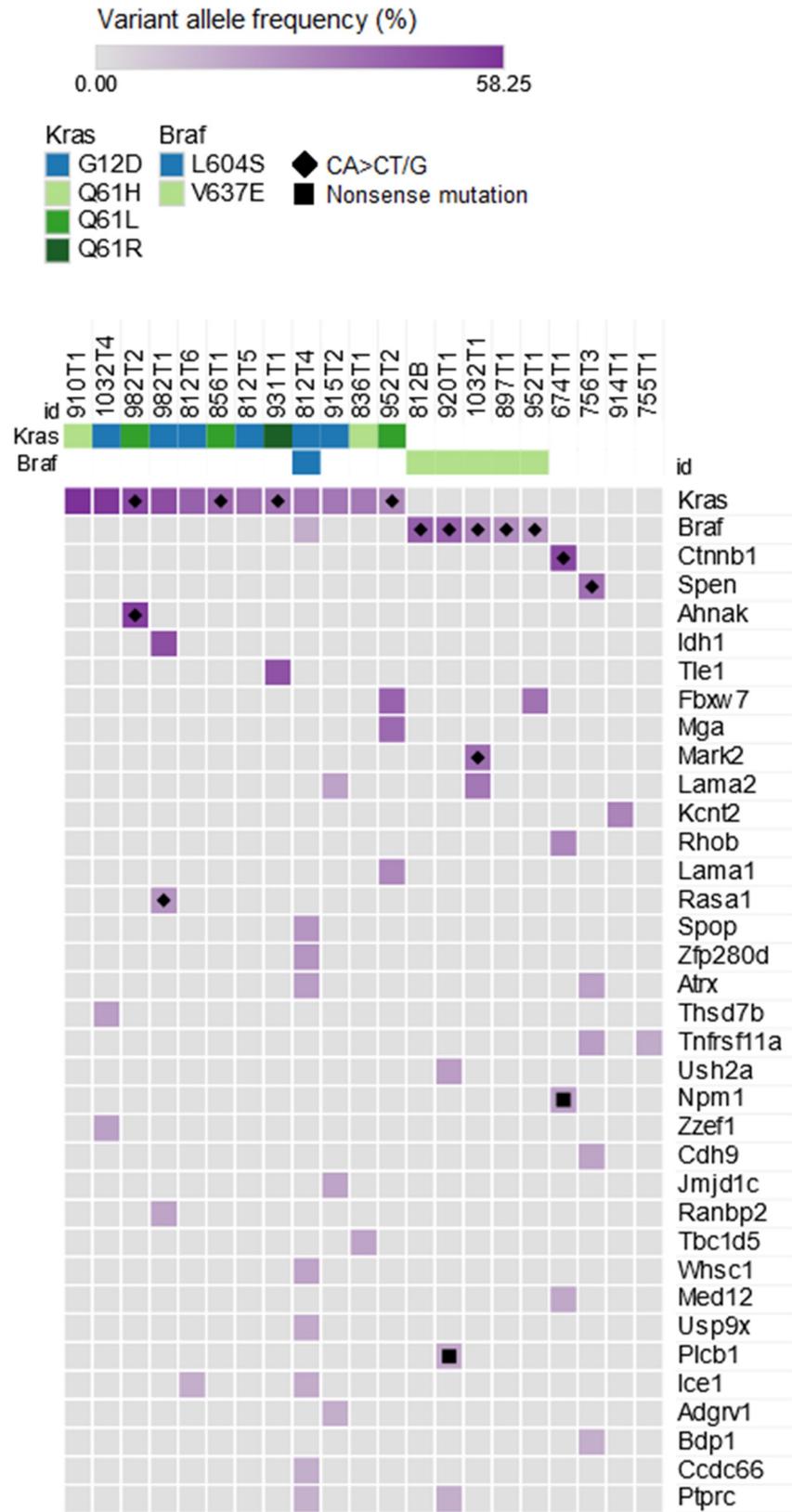


Fig 2. Identification of non-canonical genomic driver mutations. Putative genomic driver mutations (missense and nonsense mutations) in a list of 460 cancer genes identified in urethane-induced mouse lung tumors with variant allele frequency higher than 15.7% (calculated as the mean-3SD of the VAFs of *Kras* mutations). Two tumors with 0 putative genomic driver mutations identified by these criteria were not shown.

<https://doi.org/10.1371/journal.pone.0267147.g002>

arising spontaneously in mice [29], we confirmed that this mutation was indeed somatic by Sanger sequencing exon 18 of *Braf* from the tumor and matched lung (normal) tissue (S2 Table), and further, not a product of p53 loss, as the same mutation was detected in urethane-induced tumors in a p53 wild-type background (Fig 4A). As noted above, the $\text{Braf}^{\text{V637E}}$ mutation has a greater overlap with the urethane-PBS mutation signature, which is even more evident upon plotting the ratio of the optimality of the urethane-PBS to PBS mutation signatures (Figs 1A and 3B). Indeed, the mutation giving rise to $\text{Braf}^{\text{V637E}}$ is $\text{G}_{1909}\text{TG}\rightarrow\text{GAG}$, which on the reverse strand matches the $\text{CAN}\rightarrow\text{CT/GN}$ consensus sequence of urethane mutations determined by MDS after urethane exposure [7], although admittedly having a C at the 3' position makes it one of the less common mutations of this consensus (Fig 1A). Importantly, we did not detect V_{637}M , K_{638}E , G_{506}A , D_{631}N , or N_{618}S mutations, the murine equivalent to the human V_{600}M , K_{601}E , G_{469}A , D_{594}N , and N_{581}S oncogenic *BRAF* mutations [30, 31]. In agreement, none of these mutations contained the urethane mutation $\text{CAN}\rightarrow\text{CT/GN}$ consensus, and related, these mutations were largely discordant with the urethane-PBS signature, and instead mostly tracked with the PBS signature (Fig 4B). Thus, akin to $\text{Kras}^{\text{Q61L/R}}$ mutations, the mutation specificity of the carcinogen appears to underlie the type of oncogenic *Braf* mutations promoting urethane carcinogenesis.

The non-canonical $\text{Kras}^{\text{G12D}}$ and $\text{Kras}^{\text{Q61H}}$ mutations match the mutation signature of PBS

Interestingly, we find that the urethane mutation signature is more complex than the urethane-PBS mutation signature, with the former sharing more similarity with the WES mutation signature. Comparing the urethane and PBS mutation signatures revealed extensive overlap, suggesting that the origin of this increased complexity may arise from a mutation signature in normal cells (Fig 3A). One important caveat to this interpretation is a technical explanation for the SNVs detected in the PBS control tissue [32]. However, MDS was originally designed to detect the *de novo* mutations arising in bacteria populations [6], it is tempting to speculate that the incredible sensitivity of MDS may similarly be reporting on an endogenous mutagenic process in normal cells in mammals. Following, these SNVs would then become clonal in the tumors, adding to the complexity of the WES mutation signature and its similarity to the PBS mutation signature. As such, the discordance between the G_{12}D and Q_{61}H mutations in *Kras* with the urethane-PBS mutation signature, and concordance with the urethane, PBS, and WES mutation signatures, points to these mutations arising independent of the mutagenic activity of urethane, as suggested for other carcinogens [29]. Although admittedly rather speculative, one possibility is that these mutations were perhaps derived from the same process that gave rise to the PBS mutation signature.

Mutagenic specificity and oncogenic selection influence the RAS mutation patterning in human lung adenocarcinoma

The above findings are consistent with the mutational bias of urethane generating $\text{Kras}^{\text{Q61L/R}}$ and $\text{Braf}^{\text{V637E}}$ mutations in derived tumors, but not the G_{12}D and Q_{61}H mutations. As urethane is still required to induce tumors with the latter mutations, the carcinogen may be acting in a promotional manner to foster the expansion of cells with non-canonical oncogenic

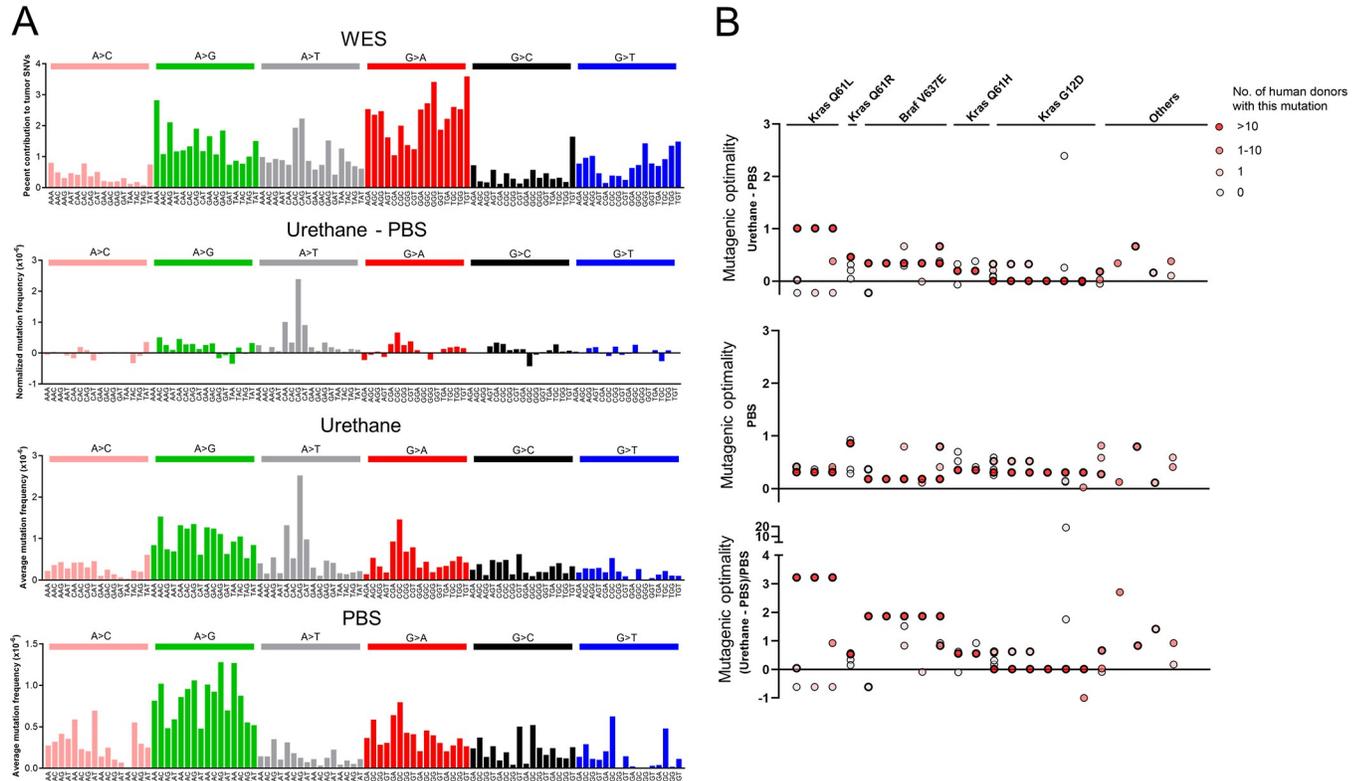


Fig 3. Multiple mutational processes contribute to putative genomic driver mutations. (A) Mutation spectrum measured by MDS for PBS, urethane, urethane normalized to PBS-treated mouse lung tissues and mutation spectrum measured by WES for urethane-induced mouse lung tumors. (B) Mutagenic optimality of putative genomic driver mutations identified by MDS for urethane normalized to PBS-treated and PBS-treated mouse lung tissues as well as the ratio of the former divided by the latter. Data are plotted for individual tumors and the genomic driver mutations occurring in more than 10 human cancer patients are indicated at the top. Others refers to tumors without a genomic driver mutation identified. Thick circles denote mutants validated by Sanger sequencing.

<https://doi.org/10.1371/journal.pone.0267147.g003>

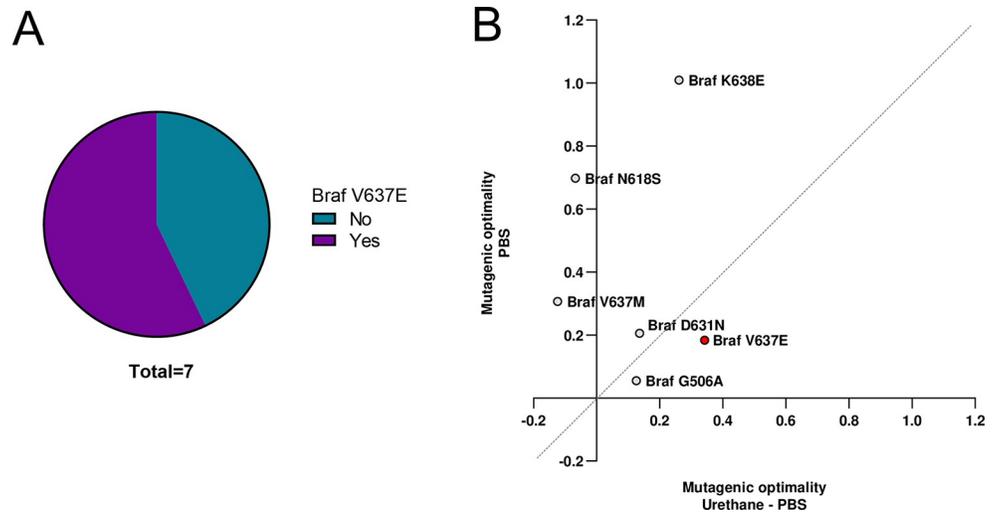


Fig 4. Multiple mutational processes contribute to putative genomic driver mutations. (A) Detection of *Braf*^{V637E} mutation in urethane-induced p53 wildtype lung tumors using Sanger sequencing. (B) Mutagenic optimality of the mouse equivalent of human oncogenic BRAF mutations identified by MDS for urethane-treated normalized to PBS-treated mouse lung tissues versus for PBS-treated mouse lung tissues. Red highlight *Braf*^{V637E} mutation detected in urethane-induced mouse lung tumors.

<https://doi.org/10.1371/journal.pone.0267147.g004>

mutations. In support, recent Whole-Genome Sequencing (WGS) analysis of numerous different carcinogen-induced tumors found few tumors whereby the mutation signature of the carcinogen matched the genomic driver mutation [29]. Recent analysis of human cancer genomes also showed that in addition to mutational processes, functional selection based on tissue-of-origin, signaling property of *KRAS* allele, as well as cooperating genetic events, also shape the mutational bias of *KRAS* gene across cancer types [33]. Thus, these previous studies collectively argue that in some cases, carcinogens may function as non-mutagenic tumor promoting agent contributing to the selection of specific *Kras* mutations in cancer. To determine if there is evidence for such a promotional activity in the selection of oncogenic *KRAS* mutations in humans, we extracted the mutation signatures in the COSMIC database [27, 34] for human *KRAS*-mutant lung adenocarcinomas. We then compared the concordance of all missense *KRAS* mutations extracted from ICGC data portal for lung adenocarcinomas [26] to the aforementioned mutation signatures (Fig 5A). Consistent with the mouse data, *KRAS* mutations in human lung adenocarcinomas displayed different levels of concordance with the mutation spectrum predicted from the mutation signatures. In more detail, the G₁₂C mutation in *KRAS*^{G12C}-mutant tumors had high concordance with the mutation spectrum of these tumors, consistent with the contribution of SBS4, associated with tobacco smoking, to G₁₂C mutations [35]. *KRAS*^{G12V}-mutant tumors have a similar mutation spectrum, and the mutation spectrum of both *KRAS*^{G12C}- and *KRAS*^{G12V}-mutant tumors favors these mutations, suggesting that either mutation arise from a potentially common mutational event. However, the mutations spectrum of *Kras*^{G12C/V}-mutant human lung adenocarcinomas also strongly favor G₁₃C, G₁₃V, Q₆₁K, and L₁₉F mutations, which are not recovered in this cancer. In fact, G₁₂A, and to a lesser degree G₁₂D and other *KRAS* mutations, are discordant with the mutation spectrum of the corresponding tumors. As in urethane carcinogenesis, we thus find evidence for secondary mutational process underlying some of these mutations. Namely, comparing the known *KRAS* mutations to the different mutation signatures present in each tumor revealed that while most oncogenic *KRAS* mutations tracked to the dominant mutation signature of each tumor, nearly a quarter of the *KRAS*^{G12/13C} mutations tracked with a minor mutation signature (Fig 5B). Thus, the concordance of the mutation signature with the genomic driver mutation supports a mutational event leading to a tumor, while discordance between these two suggests another possible mode of tumor initiation.

Discussion

Urethane displays a bias towards CAN→CT/GN mutations that perfectly match with the *Kras*^{Q61L/R} genomic driver mutations detected in developing tumors. We also show that the same holds true for the *Braf*^{V637E} oncogenic mutation, which again may explain its origins. Why the latter occurs less often remains to be resolved, but the two oncoproteins are not fungible, as *Kras* acts upstream of *Braf* and engages different proteins. Additionally, the site for a *Braf*^{V637E} mutation diverges from the most idea mutation consensus sequence more than the *Kras*^{Q61L/R} mutation. Nevertheless, as mentioned, these mutations appear to be a product of the mutational bias of the carcinogen.

The same cannot be said for the non-canonical mutations detected in *Kras*, which raises the question as to how they arose in the first place. One hint to their origins comes from the finding that the G→A substitutions responsible for these mutations were detected by MDS in the PBS-treated cohort. While we cannot rule out that these mutations are induced by urethane at a level beyond the detection limit of MDS, a very important consideration, these data do suggest the intriguing possibility that the point mutations encoding G₁₂D and Q₆₁H in *Kras* may instead be urethane-independent. Perhaps, and again rather speculatively, these mutations and

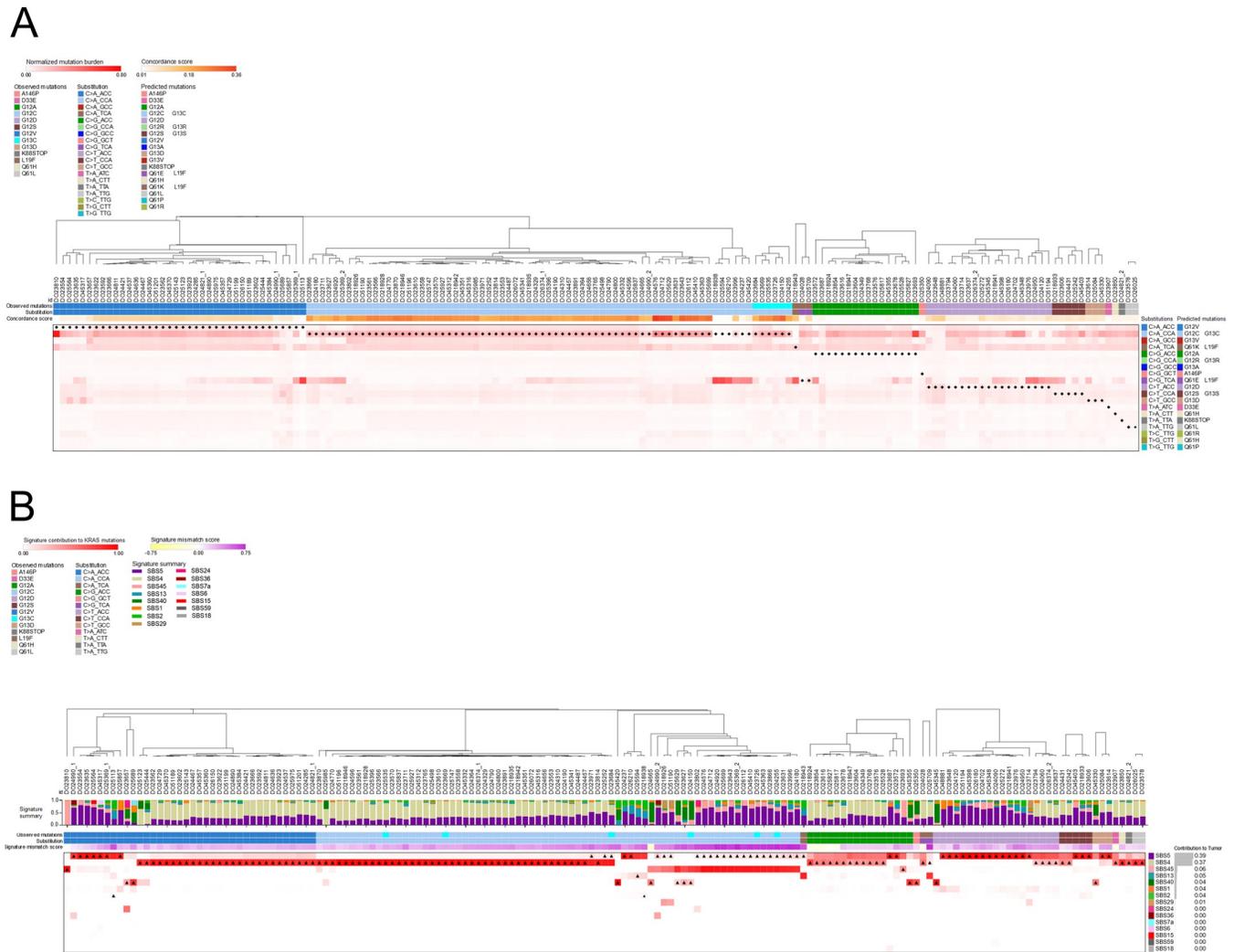


Fig 5. Multifactorial processes determine the type of *KRAS* mutations observed in human lung adenocarcinomas. (A) Heatmap of the mutation burden for substitutions matching *KRAS* oncogenic mutations in human lung adenocarcinomas estimated from the mutation signatures active in these tumors and normalized to per tumor proportions. Black diamond indicates the substitution type matching the *KRAS* mutation in that tumor. The *KRAS* mutation and the substitution type for the *KRAS* mutation for each tumor were shown below the donors. Concordance score reflects the degree of concordance between the most likely *KRAS* mutation predicted by the estimated mutation burden and the actual *KRAS* mutation observed in that tumor. (B) Heatmap of the contribution of mutation signatures to *KRAS* mutations detected in human lung adenocarcinomas. Black triangles indicate the dominant mutation signature in that tumor measured by contribution to total mutation burden. Signature summary displays the contribution of mutation signatures to total mutation burden in each tumor. The *KRAS* mutation and the substitution type for the *KRAS* mutation for each tumor were shown below the signature summary. Signature mismatch score reflects the degree of the discordance between the contribution to mutations overall and the contribution to *KRAS* mutation alone for the mutation signature most responsible the *KRAS* mutation observed in that tumor.

<https://doi.org/10.1371/journal.pone.0267147.g005>

SNVs detected by ultra-sensitive MDS sequencing of the lungs of control mice injected with the vehicle PBS are generated by the same process. As urethane is still required for tumorigenesis, perhaps the carcinogen causes another mutation that favors the expansion of a pre-existing or subsequently induced oncogenic *Kras*^{G12D/Q61H} mutation due to an endogenous mutagenic process. In support, oncogenic mutations have been detected in normal human tissues without evidence of cancer [36]. Moreover, *Kras*^{G12D}-mutant urethane-induced tumors had an average of 14 other truncal (high VAF) *CA*→*CT/G* mutations, including the recurrent mutations *Zfp991*^{H211R}, *Nccrp1*^{M197T}, and *Tnn*^{V863A}. These tumors had even more mutant genes when lowering the threshold to the more general urethane consensus A→T/G, which identified the

cancer-related mutations Rab3gap2^{Y467C}, Zfp280d^{N593S}, Cps1^{F565I}, and Evc2^{Q510L} as well as the recurrent mutations Gucyl1a3^{E232V}, Icel^{E1919V}, Tcp10c^{I194T}, Vmn1r82^{I47V}, and Flna^{S2276T} (S2 Table). Alternatively, urethane may simply have non-mutagenic activities [37, 38] that promote the expansion of pre-existing *Kras*^{G12D/Q61H}-mutant cells, consistent with recent finding of a common discordance between genomic driver mutations and carcinogen-specific mutation signatures in a wide spectrum of carcinogen-induced tumors [29]. Collectively, these findings suggest that selection, rather than mutational specificity, may be the dominating factor underlying which RAS mutation drives tumor initiation.

We also identified four tumors that lacked an oncogenic mutation in either *Kras* or *Braf*. Two of these tumors had a *CAN*→*CT/GN* mutation with a high VAF in either *Ctnnb1* or *Spn*, and neither was accompanied by a well-characterized oncogenic mutation, suggestive of a genomic driver mutation. *Ctnnb1* encodes β-catenin, which is well known to play a role in cancer [39], and this gene is mutated in other carcinogen as well as spontaneous murine tumors [29]. Furthermore, the detected mutation, S₂₃C, occurs at a site of phosphorylation [40] and O-GlcNAcylation [41]. However, while an S₂₃R mutation has been reported in human cancers [42], this mutant protein does not transform cells [40]. Furthermore, the specific mutation we detected was S23C, not S23R, and an S23C mutation has not been reported in human cancer, and hence failed our last criteria for being classified as a genomic driver mutation. *Spn* encodes a transcriptional repressor of Notch signaling, again a pathway implicated in cancer [43]. While this gene is found mutated at a low frequency in a variety of human cancers [44–46], the identified mutation Q₁₇₉L has not been reported, and hence also failed the last criteria for being classified as a genomic driver mutation in urethane-induced tumors. Thus, the nature of the genomic driver mutation in these two tumors remains unclear. Similarly, we were unable to assign an obvious genomic driver mutation to the other two tumors. However, activation of MAPK pathway may yet still play a role in the progression of tumors negative for *Kras/Braf* genomic driver mutations. Specifically, copy number analysis revealed a gain of chromosome 6, which contains both the *Kras* and the *Braf* gene, in some of these tumors (S1 Fig). Furthermore, copy number analysis also detected the loss of chromosome 9 in these tumors (S1 Fig), which the *Keap1* and *Setd2* genes, both of which have been shown to function as tumor suppressors in *Kras*^{G12D}-driven mouse model of lung adenocarcinoma [47, 48]. Thus, there are provocative candidate genomic driver mutations to explore in these other tumors.

As noted, we uncovered two types of genomic driver mutations promoting urethane carcinogenesis, those that matched the mutation signature of the carcinogen, and those that did not. It is worth noting that discordance, rather than concordance of a genomic driver mutation with the tumor mutation signature is indeed observed in human cancers. Case in point, the mutation giving rise to oncogenic *BRAF*^{V600E} fails to match the mutation signature of UV light exposure [35]. The same can be said for *BRAF*^{V600E} in colorectal cancer, *PTEN*^{R130G} in uterine carcinoma [35], *TP53* mutations in APOBEC+ human cancers [49], and a number of other genomic driver mutations in melanoma [25]. In agreement with a previous, and far more exhaustive analysis of human tumor genomic data [33], we find that the majority of the human lung adenocarcinomas analyzed had a mutation spectrum favoring a subset of mutations that included the known oncogenic *KRAS* mutation of the tumor. Nevertheless, a minority of these tumors exhibited a discordance between the genomic driver mutation and the mutation spectrum. In some cases this could be attributed to a mutagenic spectrum that was either extremely diverse (e.g. tracking with many RAS mutations) or specific (e.g. tracking with no RAS mutations), but there were definitely cases of discordance with a mutagenic signature normally tracking with a specific *KRAS* mutation. Such a discordance points to biological selection as a factor in the establishment of the genomic driver mutation in these cancers.

Interestingly, there was no mutagenic signature specific for just one type of oncogenic *KRAS* mutation. Whether this reflects mutagenic processes that give rise to the mutation signatures in human lung adenocarcinomas typically being less specific than urethane in mice, the mutagenic signature at tumor initiation being degraded over time due to additional mutagenic processes, or a frank inability of a highly specific mutagenic process to initiate lung tumorigenesis remains to be determined. Nevertheless, we hypothesize that the RAS mutation patterns of human lung adenocarcinomas could be ascribed, at least in part, to varying degrees of mutagenic specificity, akin to how we envision urethane inducing tumors with a *Kras*^{Q61L/R} or *Braf*^{V637E} genomic driver mutation, and biological selection, akin to how we speculate that urethane induces tumors with a *Kras*^{G12D/Q61H} genomic driver mutation.

Supporting information

S1 Fig. Copy number analysis reveals copy number changes in chromosome 6 and 9 in multiple tumors. (A) Heatmap of \log_2 -transformed copy number ratios showing areas of genomic gain or loss within each tumor. Columns correspond to genomic bins and rows correspond to individual tumors. Tumors are annotated by the type of *Kras* or *Braf* mutations present in the tumor. (B) Cancer-associated genes with copy number gain (\log_2 ratio > 0.25) or loss (\log_2 ratio < -0.25) in chromosome 6 and 9 in each tumor. Included genes are from the list of 460 cancer genes examined in Fig 2.

(TIF)

S1 Table. Information of sequenced tumors.

(XLSX)

S2 Table. SNVs and putative genomic driver mutations detected in sequenced tumors.

(XLSX)

S3 Table. Mutation spectrum of MDS.

(XLSX)

S4 Table. Sanger sequencing primers.

(XLSX)

S1 File. Guideline checklist.

(PDF)

Acknowledgments

We thank Dr. David MacAlpine (Duke University) for thoughtful discussions and colleagues from Duke Cancer Institute Sequencing and Genomic Technologies Shared Resource for preparing the sequencing library for whole exome sequencing and performing the sequencing of next generation sequencing libraries.

Author Contributions

Conceptualization: Siqu Li, Christopher M. Counter.

Data curation: Siqu Li.

Formal analysis: Siqu Li.

Funding acquisition: Christopher M. Counter.

Investigation: Siqu Li.

Methodology: Siqi Li.

Project administration: Christopher M. Counter.

Resources: Christopher M. Counter.

Supervision: Christopher M. Counter.

Validation: Siqi Li.

Visualization: Siqi Li.

Writing – original draft: Siqi Li, Christopher M. Counter.

Writing – review & editing: Siqi Li, Christopher M. Counter.

References

1. Gowd V, Su H, Karlovsky P, Chen W. Ethyl carbamate: An emerging food and environmental toxicant. *Food Chem.* 2018; 248:312–21. <https://doi.org/10.1016/j.foodchem.2017.12.072> PMID: 29329860
2. Dwyer-Nield LD, McQuillan J, Hill-Baskin A, Radcliffe RA, You M, Nadeau JH, et al. Epistatic interactions govern chemically-induced lung tumor susceptibility and Kras mutation site in murine C57BL/6J-ChrA/J chromosome substitution strains. *Int J Cancer.* 2010; 126(1):125–32. <https://doi.org/10.1002/ijc.24743> PMID: 19609923
3. Fernando RC, Nair J, Barbin A, Miller JA, Bartsch H. Detection of 1,*N*⁶-ethenodeoxyadenosine and 3, *N*⁴-ethenodeoxycytidine by immunoaffinity/³²P-postlabelling in liver and lung DNA of mice treated with ethyl carbamate (urethane) or its metabolites. *Carcinogenesis.* 1996; 17(8):1711–8. <https://doi.org/10.1093/carcin/17.8.1711> PMID: 8761431
4. Forkert PG. Mechanisms of lung tumorigenesis by ethyl carbamate and vinyl carbamate. *Drug Metab Rev.* 2010; 42(2):355–78. <https://doi.org/10.3109/03602531003611915> PMID: 20205516
5. Westcott PM, Halliwill KD, To MD, Rashid M, Rust AG, Keane TM, et al. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature.* 2015; 517(7535):489–92. <https://doi.org/10.1038/nature13898> PMID: 25363767
6. Jee J, Rasouly A, Shamovsky I, Akivis Y, Steinman SR, Mishra B, et al. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature.* 2016; 534(7609):693–6. <https://doi.org/10.1038/nature18313> PMID: 27338792
7. Li S, MacAlpine DM, Counter CM. Capturing the primordial Kras mutation initiating urethane carcinogenesis. *Nat Commun.* 2020; 11(1):1800. <https://doi.org/10.1038/s41467-020-15660-8> PMID: 32286309
8. Valentine CC 3rd, Young RR, Fielden MR, Kulkarni R, Williams LN, Li T, et al. Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. *Proc Natl Acad Sci USA.* 2020; 117(52):33414–25. <https://doi.org/10.1073/pnas.2013724117> PMID: 33318186
9. Lampson BL, Pershing NL, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta CV, et al. Rare codons regulate Kras oncogenesis. *Curr Biol.* 2013; 23(1):70–5. <https://doi.org/10.1016/j.cub.2012.11.031> PMID: 23246410
10. Pershing NL, Lampson BL, Belsky JA, Kaltenbrun E, MacAlpine DM, Counter CM. Rare codons capacitate Kras-driven *de novo* tumorigenesis. *J Clin Invest.* 2015; 125(1):222–33. <https://doi.org/10.1172/JCI77627> PMID: 25437878
11. Li S, Counter CM. Signaling amplitude molds the Ras mutation tropism of urethane. *Elife.* 2021; 10:e67172. <https://doi.org/10.7554/eLife.67172> PMID: 33998997
12. Munoz-Espin D, Serrano M. Cellular senescence: from physiology to pathology. *Nat Rev Mol Cell Biol.* 2014; 15(7):482–96. <https://doi.org/10.1038/nrm3823> PMID: 24954210
13. Burd CE, Liu W, Huynh MV, Waqas MA, Gillahan JE, Clark KS, et al. Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov.* 2014; 4(12):1418–29. <https://doi.org/10.1158/2159-8290.CD-14-0729> PMID: 25252692
14. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018; 46(W1):W537–44. <https://doi.org/10.1093/nar/gky379> PMID: 29790989
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404

16. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26(5):589–95. <https://doi.org/10.1093/bioinformatics/btp698> PMID: 20080505
17. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*. 2011; 27(12):1691–2. <https://doi.org/10.1093/bioinformatics/btr174> PMID: 21493652
18. Li H. Improving SNP discovery by base alignment quality. *Bioinformatics*. 2011; 27(8):1157–8. <https://doi.org/10.1093/bioinformatics/btr076> PMID: 21320865
19. Acosta YY, Zafra MP, Ojeda G, Bernardone IS, Dianzani U, Portoles P, et al. Biased binding of class IA phosphatidylinositol 3-kinase subunits to inducible costimulator (CD278). *Cell Mol Life Sci*. 2011; 68(18):3065–79. <https://doi.org/10.1007/s00018-010-0606-1> PMID: 21188463
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
21. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016; 12(4):e1004873. <https://doi.org/10.1371/journal.pcbi.1004873> PMID: 27100738
22. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490(7418):61–70. <https://doi.org/10.1038/nature11412> PMID: 23000897
23. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672
24. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011; 477(7364):289–94. <https://doi.org/10.1038/nature10413> PMID: 21921910
25. Dietlein F, Weghorn D, Taylor-Weiner A, Richters A, Reardon B, Liu D, et al. Identification of cancer driver genes based on nucleotide context. *Nat Genet*. 2020; 52(2):208–18. <https://doi.org/10.1038/s41588-019-0572-y> PMID: 32015527
26. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal. *Nat Biotechnol*. 2019; 37(4):367–9. <https://doi.org/10.1038/s41587-019-0055-9> PMID: 30877282
27. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019; 47(D1):D941–7. <https://doi.org/10.1093/nar/gky1015> PMID: 30371878
28. Pakneshan S, Salajegheh A, Smith RA, Lam AK. Clinicopathological relevance of BRAF mutations in human cancer. *Pathology*. 2013; 45(4):346–56. <https://doi.org/10.1097/PAT.0b013e328360b61d> PMID: 23594689
29. Riva L, Pandiri AR, Li YR, Droop A, Hewinson J, Quail MA, et al. The mutational signature profile of known and suspected human carcinogens in mice. *Nat Genet*. 2020; 52(11):1189–97. <https://doi.org/10.1038/s41588-020-0692-4> PMID: 32989322
30. Bracht JWP, Karachaliou N, Bivona T, Lanman RB, Faull I, Nagy RJ, et al. BRAF mutations classes I, II, and III in NSCLC patients included in the SLLIP trial: The need for a new pre-clinical treatment rationale. *Cancers*. 2019; 11(9):1381.
31. Nieto P, Ambrogio C, Esteban-Burgos L, Gomez-Lopez G, Blasco MT, Yao Z, et al. A Braf kinase-inactive mutant induces lung adenocarcinoma. *Nature*. 2017; 548(7666):239–43. <https://doi.org/10.1038/nature23297> PMID: 28783725
32. Arbeithuber B, Makova KD, Tiemann-Boege I. Artfactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications. *DNA Res*. 2016; 23(6):547–59. <https://doi.org/10.1093/dnares/dsw038> PMID: 27477585
33. Cook JH, Melloni GEM, Gulhan DC, Park PJ, Haigis KM. The origins and genetic interactions of KRAS mutations are allele- and tissue-specific. *Nat Commun*. 2021; 12(1):1808. <https://doi.org/10.1038/s41467-021-22125-z> PMID: 33753749
34. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020; 578(7793):94–101. <https://doi.org/10.1038/s41586-020-1943-3> PMID: 32025018
35. Temko D, Tomlinson IPM, Severini S, Schuster-Bockler B, Graham TA. The effects of mutational processes and selection on driver mutations across cancer types. *Nat Commun*. 2018; 9(1):1857. <https://doi.org/10.1038/s41467-018-04208-6> PMID: 29748584

36. Yizhak K, Aguet F, Kim J, Hess JM, Kubler K, Grimsby J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science*. 2019; 364(6444):eaaw0726. <https://doi.org/10.1126/science.aaw0726> PMID: 31171663
37. Stathopoulos GT, Sherrill TP, Cheng DS, Scoggins RM, Han W, Polosukhin VV, et al. Epithelial NF- κ B activation promotes urethane-induced lung carcinogenesis. *Proc Natl Acad Sci USA*. 2007; 104(47):18514–9. <https://doi.org/10.1073/pnas.0705316104> PMID: 18000061
38. Zaynagetdinov R, Sherrill TP, Polosukhin VV, Han W, Ausborn JA, McLoed AG, et al. A critical role for macrophages in promotion of urethane-induced lung carcinogenesis. *J Immunol*. 2011; 187(11):5703–11. <https://doi.org/10.4049/jimmunol.1100558> PMID: 22048774
39. Kim S, Jeong S. Mutation hotspots in the beta-catenin gene: lessons from the human cancer genome databases. *Mol Cells*. 2019; 42(1):8–16. <https://doi.org/10.14348/molcells.2018.0436> PMID: 30699286
40. van Noort M, van de Wetering M, Clevers H. Identification of two novel regulated serines in the N terminus of beta-catenin. *Exp Cell Res*. 2002; 276(2):264–72. <https://doi.org/10.1006/excr.2002.5520> PMID: 12027456
41. Olivier-Van Stichelen S, Dehennaut V, Buzy A, Zacharyus JL, Guinez C, Mir AM, et al. O-GlcNAcylation stabilizes beta-catenin through direct competition with phosphorylation at threonine 41. *Faseb J*. 2014; 28(8):3325–38. <https://doi.org/10.1096/fj.13-243535> PMID: 24744147
42. Legoix P, Bluteau O, Bayer J, Perret C, Balabaud C, Belghiti J, et al. Beta-catenin mutations in hepatocellular carcinoma correlate with a low rate of loss of heterozygosity. *Oncogene*. 1999; 18(27):4044–6. <https://doi.org/10.1038/sj.onc.1202800> PMID: 10435629
43. Aster JC, Pear WS, Blacklow SC. The varied roles of Notch in cancer. *Annu Rev Pathol*. 2017; 12:245–75. <https://doi.org/10.1146/annurev-pathol-052016-100127> PMID: 27959635
44. Zhao X, Little P, Hoyle AP, Pegna GJ, Hayward MC, Ivanova A, et al. The prognostic significance of low-frequency somatic mutations in metastatic cutaneous melanoma. *Front Oncol*. 2018; 8:584. <https://doi.org/10.3389/fonc.2018.00584> PMID: 30662871
45. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet*. 2018; 50(5):645–51. <https://doi.org/10.1038/s41588-018-0078-z> PMID: 29610475
46. Stephens PJ, Davies HR, Mitani Y, Van Loo P, Shlien A, Tarpey PS, et al. Whole exome sequencing of adenoid cystic carcinoma. *J Clin Invest*. 2013; 123(7):2965–8. <https://doi.org/10.1172/JCI67201> PMID: 23778141
47. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat Med*. 2017; 23(11):1362–8. <https://doi.org/10.1038/nm.4407> PMID: 28967920
48. Walter DM, Venancio OS, Buza EL, Tobias JW, Deshpande C, Gudiel AA, et al. Systematic *in vivo* inactivation of chromatin-regulating enzymes identifies Setd2 as a potent tumor suppressor in lung adenocarcinoma. *Cancer Res*. 2017; 77(7):1719–29. <https://doi.org/10.1158/0008-5472.CAN-16-2159> PMID: 28202515
49. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science*. 2019; 364(6447):eaaw2872. <https://doi.org/10.1126/science.aaw2872> PMID: 31249028