ARTICLE

Capturing the primordial Kras mutation initiating urethane carcinogenesis

Siqi Li1, David M. MacAlpine1 & Christopher M. Counter1✉

The environmental carcinogen urethane exhibits a profound specificity for pulmonary tumors driven by an oncogenic Q61L/R mutation in the gene Kras. Similarly, the frequency, isoform, position, and substitution of oncogenic RAS mutations are often unique to human cancers. To elucidate the principles underlying this RAS mutation tropism of urethane, we adapted an error-corrected, high-throughput sequencing approach to detect mutations in murine Ras genes at great sensitivity. This analysis not only captured the initiating Kras mutation days after urethane exposure, but revealed that the sequence specificity of urethane mutagenesis, coupled with transcription and isoform locus, to be major influences on the extreme tropism of this carcinogen.

1Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710, USA. ✉email: count004@mc.duke.edu

https://doi.org/10.1038/s41467-020-15660-8
Exposure of mice to the environmental carcinogen urethane primarily induces tumors in one organ (lung) with a single driver mutation in only one of the three Ras genes (Kras), at one position (Q61), with one substitution (L or R depending on the mouse strain)\(^4,5\). This extreme RAS mutation tropism is rather remarkable considering that oncogenic Ras is well known to cause a host of cancers in the mouse beyond the lungs\(^6\). Further, there are 57 possible point mutations generating 54 known or potentially oncogenic substitutions among the three Ras genes (Hras, Nras, and Kras) at the three major oncogenic positions (G12, G13, and Q61)\(^6\) from a total of six possible oncogenic amino acids (V, D, C, S, R, A at G12/G13, and L, R, K, H, E, P at Q61)\(^7\). Capturing the mutations arising immediately after urethane exposure in normal tissues in vivo would greatly inform the underlying mechanism by which this carcinogen induces one tumor type driven by an incredibly specific Ras mutation.

The challenge of detecting mutations early lies in the extremely low rate of urethane mutagenesis in vivo, which ranges from 5 to \(100 \times 10^{-6}\) or 1 mutant per 10 to \(2 \times 10^{5}\) templates\(^8\). This rate is significantly lower than the detection limit of conventional next generation sequencing (NGS)\(^9\), which varies from 1 to \(10 \times 10^{-3}\) or 1 mutant per 100 to 1000 templates. Various techniques have been developed to lower the error rate of NGS\(^10,11\), but improvements are still limited by mutations arising during early PCR steps and data recovery efficiency\(^12\). The recently developed error-corrected, high-throughput maximum-depth sequencing (MDS) method overcomes these limitations, identifying ultra-rare (1 \(\times 10^{-6}\), 1 mutant per \(10^6\) templates) antibiotic-resistance mutations arising in bacteria populations\(^13\). We thus sought to adopt this assay for the much larger mammalian genome to capture the initiating Kras mutation induced by urethane in vivo to elucidate the principles underpinning the extreme selectivity of this carcinogenic process.

By adapting MDS for the mammalian genome, we capture the dominant initiating Kras\(^{Q61L}\) mutation in the lungs of mice immediately following urethane exposure. Further, we show that the substitution and position tropism of urethane can be largely ascribed to the specificity of this carcinogen for C\(\rightarrow\)A and T\(\rightarrow\)C mutations, which generates the oncogenic Q61L mutation in Kras. The same mutations were also captured in Hras, arguing that these mutations are not sufficient to induce tumorigenesis, which speaks to the isoform tropism of urethane. Last, in terms of tissue tropism, oncogenic Ras mutations were generally undetectable in other organs tested, which we suggest is linked to the transcriptional status of genes. Collectively, these findings indicate that RAS mutation tropism is a multifactorial process, which may inform similar RAS mutation patterns observed in human cancers.

## Results

### Adapting MDS to the mammalian genome

A barrier to detecting initiating mutations in Kras at the time they occur in vivo after urethane exposure is that the mutation rate of this carcinogen is well below the detection limit of NGS. To overcome this limitation, we turned to the error-corrected, high-throughput sequencing approach of MDS, which recovered mutants in bacteria at a frequency as low as \(1 \times 10^{-6}\) or 1 mutant per \(10^6\) templates\(^13\). The key steps of MDS are first, synthesis of unique barcodes onto one strand of a genomic region-of-interest, second, linear amplification to obtain multiple direct copies of the barcoded genomic DNA, third, exponential amplification to obtain families of PCR products sharing the same barcode, and fourth, ultra-deep sequencing of millions of barcode families from the single region-of-interest\(^13\). Bona fide mutations are differentiated from PCR and sequencing errors by virtue of being detected in all members of one barcode family\(^13\). The challenge of adapting MDS to the mammalian genome is maintaining the recovery of a sufficient number of analyzable barcode families (with at least two or three members) in a genome that is three orders of magnitude larger in size and weight\(^14,15\). To this end, we optimized assay conditions (see Methods) for mammalian Kras (Supplementary Fig. 1a) and barcode recovery (Supplementary Table 1). To validate the sensitivity of this mammalian version of MDS, we generated a panel of Kras-mutant plasmids, each comprised of Kras cDNA with a unique set of co-occurring double or triple mutations in the region encoded by exon 1 and/or exon 2 (Supplementary Table 2). Each was spiked at specific concentrations into genomic DNA isolated from mouse embryonic fibroblasts (MEFs) or murine lungs to benchmark different levels of sensitivity. As the error rates of PCR and sequencing are unlikely to give the same two or three exact improper base calls, the actual frequency of mutants present in the sample was estimated by calculating the frequency of barcode families with the pre-engineered co-occurring mutations. The frequency of mutations determined by MDS was then compared against the aforementioned actual frequency. Using this approach, we demonstrated that MDS adapted for the transcribed strand of Kras exon 1 detected mutations at a sensitivity of \(5 \times 10^{-7}\) or 1 mutant per \(2 \times 10^6\) templates (Fig. 1a, Supplementary Fig. 1b, and Supplementary Table 1). We further validated the sensitivity of the MDS assay adapted for the non-transcribed strand of Kras exon 2 in the same fashion (Supplementary Fig. 1c and Supplementary Table 2). Thus, MDS optimized for mammalian genomic DNA detects mutations at a sensitivity potentially 20,000 times greater than conventional NGS.

### Capturing the initiating oncogenic mutation in Kras

Urethane induces pulmonary tumors driven by a Kras\(^{Q61L/R}\) oncogenic mutation\(^1,3,5\), exemplifying the selectivity of this carcinogen at the level of tissue, isoform, position, and substitution. To elucidate the processes behind this RAS mutation tropism, we exposed A/J mice to urethane or the vehicle PBS via three daily intraperitoneal injections. After 1, 2, 3, and 4 weeks, genomic DNA was isolated from the lungs of four to seven mice from each condition. The non-transcribed strand of exon 2 of the endogenous Kras gene was then sequenced by MDS. To ensure abundant depth for mutant recovery and the accuracy of detected mutation frequency, samples with \(<1.5 \times 10^6\) barcodes were excluded from analysis. For the remaining samples (barcode recovery listed in Supplementary Data 1), mutation frequencies were summed by either nucleotide position or substitution type, normalized to control PBS, log\(_{10}\) transformed, and then plotted as a heatmap (Fig. 1b, c). This analysis identified the well-established\(^1,3,5\) oncogenic L (and to a lesser extent R) mutation at codon Q61 preferentially in the urethane, but not PBS cohort of mice, as early as 1 week after exposure to this carcinogen. Consistent with being initiating events, these mutations expanded over time indicative of tumor growth (Fig. 1b and Supplementary Data 2), although a longer time course would formally confirm a tumorigenic identity. We also independently confirm by droplet digital PCR\(^16\) the presence of the Q61L mutation 4 week post urethane exposure at a frequency similar to that determined by MDS (Supplementary Fig. 1d). We thus capture and confirm the primordial initiating oncogenic mutation in Kras within days of exposure to urethane.

### Substitution tropism

Previous whole-exome sequencing of urethane-induced tumors revealed a strong bias toward A\(\rightarrow\)T/G substitutions\(^3\), consistent with ethenodeoxyadenosine adducts forming in vivo after urethane exposure\(^17,18\). These substitutions were also detected in Kras by MDS at a high frequency, although...
Fig. 1 MDS detects ultra-rare mutations induced by the carcinogen urethane. a Frequency of single (detected) versus co-occurring (present) mutations identified by MDS using a dilution series of Kras cDNAs with 2–3 different mutations engineered in exon 1 mixed with genomic DNA from mouse lung tissue. b–d Heatmap of the mutation frequency (MF) determined by MDS for the non-transcribed strand of exon 2 of Kras from the lungs of mice at the indicated time points after exposure to urethane (UR) or PBS (n = 7 mice for the UR and PBS cohorts at week 1, 5 mice for the PBS cohort at week 4, and 4 mice for all other cohorts from one experiment), plotted as (b, c) the log-transformed fold-change normalized to PBS-treated mice (FC over PBS) or (d) log-transformed versus each (b) nucleotide (annotated by amino acid at the top, Q61L and R mutations are highlighted in red, scaled by detected frequency), (c) type of substitution or (d) A>T transversions (nucleotide number as well as the 5′ and 3′ base of the substituted A are shown at the top). e, f Mean ± SEM mutation frequency of (e) each possible C>A to C>T transversion at the indicated time points after mice were exposed to urethane (UR) or PBS (n = 7 mice for the UR and PBS cohorts at week 1, 5 mice for the PBS cohort at week 4, and 4 mice for all other cohorts from one experiment) or (f) all possible missense mutations for Q61 codon in mice 1 week after urethane exposure (n = 7 mice from one experiment). p values calculated by e Dunn’s multiple comparison test following Kruskal–Wallis test or f Holm-Sidak multiple comparisons test following one-way ANOVA. ****p < 0.0001 and ***p < 0.001.
A>T transversions were far more common than A>G transitions (Fig. 1c). In agreement with this bias, CA_{182A→CTA} gives rise to the dominant Q61L oncogenic mutation in tumors of the A/J strain of mice exposed to urethane, while CA_{182A→CGA} gives rise to the rarer Q61R oncogenic mutation. Still, the overall A/T content of murine genome is about 58%, and A>T/G substitutions represent two-thirds of the possible base changes for this nucleotide. As such, this mutagenic signature is rather general compared with the extreme specificity of the initiating mutation. Further analysis of the mutation signature, using the log_{10}-transformed mutation frequencies of individual substitutions, revealed that the most prominent substitution detected by MDS in the lungs of mice after urethane (but not PBS) exposure at all time points was an A>T transversion within the context of a 5’T and 3’A nucleotide, namely a CAN trinucleotide (Fig. 1d, Supplementary Fig. 2a, b, and Supplementary Data 2, 3). In agreement, a 5’T was favored to some extent for A>T transversions in previous whole-exome sequencing of urethane-induced lung tumors. The frequency of CAN→CTN mutations recovered in the urethane-exposed cohort remained constant over time in all but one case; CAN→CTA encoding the oncogenic Q61L mutation expanded at subsequent time points ostensibly due to tumor growth (Fig. 1e). The same was true for the second most prominent urethane-specific substitution detected by MDS, an A>G transition preceded by 5’T (Supplementary Fig. 2b), where again CA_{182A→CGA} that gives rise to the rarer Q61R oncogenic mutation expanded over time (Supplementary Fig. 2c).

Substitutions other than CA_{182A→CTA} at codon 61 were rarely detected 1 week after urethane exposure (Fig. 1f), even though all the possible missense mutations at this codon generated by a single-nucleotide substitution (Q61L, R, K, E, P, and H) have been reported in human cancers in the COSMIC database. As such, an A>T/G substitution preceded by C greatly increases the specificity of urethane mutations for codon 61, reducing the number of potential non-synonymous changes in both strands of the murine Kras gene by fivefold, from 616 to 120. The selectivity of these two substitutions after urethane exposure thus appears to be a major contributing factor to the substitution bias toward Q61L/R mutations in Kras.

Position tropism. This bias of urethane for (C)A>T/G substitutions similarly argues against mutations arising at an appreciable level in codons 12 (G13,GT) or 13 (G13,GC) in exon 1, as neither fit the CAN pattern in either strand orientation. Related to this, despite the fact that oncogenic mutations at G12, and to a lesser extent G13, occur frequently in human cancers and when introduced into the lungs of mice are tumorigenic to varying degrees, they are rarely recovered from urethane-induced tumors. We therefore sequenced the transcribed strand of exon 1 of Kras by MDS from genomic DNA isolated from the lungs of mice 1, 2, 3, and 4 weeks after exposure to urethane or PBS. To overcome interference from strand-specific background (see Methods), we also sequenced the non-transcribed strand of exon 1 of Kras by MDS from the lungs of mice at the 1- and 4-week time points. While CAN→CTN transversions were again preferentially detected 1 week after urethane exposure (Fig. 2a, b, Supplementary Fig. 3a, b, and Supplementary Data 4, 5), indicating urethane mutagenesis occurred in this exon, oncogenic mutations were rarely recovered in either codon 12 or 13 (Fig. 2c). Interestingly, some G12 and G13 mutations were detected at a low frequency 4 weeks after urethane exposure (Fig. 2d). It is worth noting that oncogenic mutations at G12 have been reported in urethane-induced tumors, but are quite rare. This suggests that G12 and G13 mutations are induced by urethane exposure, but remain below the limit of detection of MDS unless a certain degree of clonal expansion occurs. Similarly, while a Q61H (A183>T) mutation was rarely detected 1 week after urethane exposure (Fig. 1f), it was more prevalent in later samples (Fig. 1d and Supplementary Data 2). Collectively, these findings argue that the mutational position tropism of urethane can be ascribed in large part to a mutational bias of this environmental carcinogen toward CAN→CT/GN mutations.

Isotropism. The other two Ras genes, Hras and Nras, encode the identical codon 61 (CAA). CAN→CT/GN substitutions at this codon generate the identical oncogenic Q61L/R mutations, which are well known to render Hras and Nras oncogenic. Despite this, oncogenic mutations in Hras or Nras are not recovered in urethane-induced lung tumors. This suggests that either these loci are resistant in some manner to urethane mutagenesis or oncogenic mutations in these two genes are unable to initiate tumorigenesis. To differentiate between these two possibilities, we optimized the MDS assay to detect mutations in the non-transcribed strand of exon 2 in Hras (see Methods). We then applied this approach to genomic DNA isolated from the lungs of mice 1 and 4 weeks after urethane or PBS. We found a high prevalence of A>T followed by A>G mutations in exon 2 of Hras (Fig. 3a and Supplementary Data 6), with again CAN→CTN transversions being the predominant mutation in the urethane cohort, including the oncogenic CA_{182A→CTA} mutation in codon 61 (Fig. 3b). CAN→CTN transversions in Hras were detected somewhat less often than in Kras 1 week (Fig. 3c), but similarly 4 weeks after urethane exposure (Fig. 3d). Unlike in the case of Kras, however, oncogenic mutations in Hras did not expand appreciably over time (Fig. 3d). Hras therefore appears to acquire oncogenic mutations at a detectable frequency, but such mutations do not support tumorigenesis. This suggests that the isoform tropism of urethane is a product of the Hras locus and not an inability to induce oncogenic mutations at this site.

Organ tropism. Pulmonary lesions are the primary tumors arising in mice after intraperitoneal injections of urethane. However, activating an oncogenic Kras allele in a broad spectrum of murine organs has been documented to be tumorigenic. The lungs, liver, and pancreas were removed and subjected to LC/MS/MS to measure the levels of urethane and its active metabolite vinyl carbamate. Similar levels of both compounds were detected in the lungs and liver, but less in the pancreas over the three time points, with the terminal time point showing the highest concentration in the liver, followed by the lung, and then the
Mutational strand asymmetry has been observed for other mutational processes and correlated with the transcriptional status of mutated genes. Kras mRNA levels determined by quantitative RT-PCR (RT-qPCR) or RNA-seq have been reported to be higher in the murine lung compared with the liver. In agreement, we validated the higher expression of Kras mRNA in lung compared with liver and pancreas by RT-qPCR (Fig. 5c). In addition, strand bias is not significant in the liver, consistent with a general lack of mutations detected in this organ after urethane exposure (Fig. 5d). Prompted by this, we examined the relationship between mutation frequency and gene expression using mutations detected in a previous published whole-exome sequencing of urethane-induced lung adenomas and adenocarcinomas and a published RNA-seq dataset generated from the adult mouse lung. Genes were partitioned into quartiles based on expression level and the number of C > T transversions in the non-transcribed or transcribed strand in each quartile were compared with the previous determined mutation frequency at codon Q61 (Fig. 1, n = 7 mice at week 1 and 4 mice at week 4 from one experiment) compared with the previous determined mutation frequency at codon Q61 (Fig. 1, n = 7 mice at week 1 and 4 mice at week 4 from one experiment) compared with the previous determined mutation frequency at codon Q61 (Fig. 1, n = 7 mice at week 1 and 4 mice at week 4 from one experiment).
**Fig. 3** MDS detects the isoform tropism of urethane.  
**a**, **b** Heatmap of the mutation frequency (MF) determined by MDS for the non-transcribed strand of exon 2 of Hras from the lungs of mice at the indicated time points after exposure to urethane (UR) or PBS (n = 4 mice at each time point from one experiment), plotted as the (a) log-transformed fold-change normalized to PBS-treated mice (FC over PBS) or (b) log-transformed versus each (a) type of substitution or (b) A>T transversions (nucleotide number as well as the 5' and 3' base of the substituted A are shown at the top).  
**c**, **d** Mean ± SEM mutation frequency of each possible CAN to CTN transversion in exon 2 of Hras (n = 4 mice from one experiment) compared with the previous determined mutation frequency in exon 2 of Kras (Fig. 1, n = 7 mice at week 1 and 4 mice at week 4 from one experiment) at (c) 1 week or (d) 4 weeks after exposure to urethane (UR) or PBS.  
*p* values calculated by (c) Holm-Sidak multiple comparisons test following one-way ANOVA, or (d) Dunn’s multiple comparison test following Kruskal-Wallis test. ****p < 0.0001 and ns: not significant.

**Fig. 4** MDS detects the organ tropism of urethane.  
**a**, **b** Mean ± SEM mutation frequency each possible CAN to CTN transversion determined by MDS for the non-transcribed strand of exon 2 of Kras in the pancreas or liver of mice (n = 3 mice for pancreas samples from the UR cohort at weeks 1 and 4, 5 mice for liver samples from the UR and PBS cohorts at week 1, and 4 mice for all other cohorts from one experiment) compared with the previous determined mutation frequency in the lung (Fig. 1, n = 7 mice at week 1 and 4 mice at week 4 from one experiment) either (a) 1 week or (b) 4 weeks after exposure to urethane (UR) or PBS.  
*p* values calculated by (a) Holm-Sidak multiple comparisons test following one-way ANOVA or (b) Dunn’s multiple comparison test following Kruskal-Wallis test. ****p < 0.0001, **p < 0.01, and ns: not significant.
revealed by MDS sequencing, CAN→CIN transversions increased with gene expression on the non-transcribed strand but decreased on the transcribed strand (Fig. 5c). The same trends were observed when the RNA-seq dataset for adult mouse lung from the mouse ENCODE project was analyzed (Supplementary Fig. 5c). Collectively, these findings point toward the organ tropism of urethane being related to the high transcription of Kras in the lung.

Discussion
Here we adapted MDS, an error-corrected, high-throughput sequencing approach originally developed for use in microbiology, to now detect extremely rare mutations in the mammalian genome at a sensitivity of up to $5 \times 10^{-7}$ (1 mutant per $2 \times 10^{6}$ templates). While we developed this assay to study RAS mutation tropism, MDS could find value in other applications, such as early detection. Nevertheless, by leveraging MDS to study the mutagenesis process at the earliest stage of tumorigenesis, we detected the initiating Q61L/R mutations in Kras in the lungs of mice only days after exposure to urethane, capturing the very birth of cancer. We note that mutant allele-specific amplification and droplet digital PCR have documented Kras mutations after carcinogen exposure. However, we chose to develop MDS for the mammalian settings as these assays are either not as quantitative and sensitive, or are designed to examine pre-selected mutations. Indeed, capitalizing on the

Fig. 5 Urethane strand bias. a, b, d Mean ± SEM mutation frequency of the indicated CAN to CIN transversions and reverse-complementary substitutions averaged by nucleotide positions determined by MDS sequencing of Kras exon 2 (a) non-transcribed strand in lungs (Fig. 1, n = 7 mice from one experiment) or (b) transcribed strand in lungs in mice 3 weeks (n = 4 mice from one experiment) after exposure to urethane (UR) or PBS. c Mean ± SEM relative expression of Kras mRNA in the lung, liver, and pancreas (normalized to lung) of mice determined by RT-qPCR (n = 4 mice from one experiment). e Mean ± SEM frequency of CAN to CIN transversions of the non-transcribed versus transcribed strand in urethane-induced tumors from whole-exome sequencing data in genes binned by their mRNA levels from the mouse lung (n = 66 tumors). p values calculated by Holm-Sidak multiple comparisons test following a, b, d, e two-way ANOVA or c one-way ANOVA. ****p < 0.0001, *p < 0.05 and ns: not significant.
ability of MDS to detect potentially any sequence variation in targeted regions of Ras genes at great sensitivity, we show at least three features underpinning the extreme mutational tropism of urethane—the mutational bias of this environmental carcinogen, transcription, and the gene locus.

With regards to the substitution and position bias of urethane, we demonstrate that the prevalence of Q61L/R mutations arises in large part due to the known preference of urethane for A>T/G substitutions, especially as we show here in the context of a 5’ C. This mutational bias, coupled with codon 61 containing a C ANC trinucleotide that when the A is mutated to either T or G gives rise to an oncogenic L (C/G12R) or R (C/G12S) amino acid, favors the KrasQ61L/R driver mutation characteristic of this carcinogen. Other oncogenic mutations at Q61, G12, or G13 codons do not result from CA→CT/G substitutions, and in agreement, were rarely detected following urethane exposure. The implication being that a mutagenic preference may influence the type of initiating mutations in cancer. Similarly in humans, a CCT→CTC mutation characteristic of C>T transitions induced by UV encodes an activating PS3 mutation in RAC1 in sun-exposed melanoma.

While Q61H, G12, and G13 oncogenic mutations in Kras, which are not favored by urethane mutagenesis, were rare or absent 1 week after urethane exposure, they were detected 4 weeks later. This implies that extremely rare mutations induced by urethane, provided they have a favorable oncogenic outcome, may initiate tumorigenesis (although we cannot formally rule out that these were pre-existing mutations unveiled by a cooperating mutation induced by urethane). In agreement, while the Q61L mutation is more frequent than Q61R in urethane-induced lung tumors of the A/J mouse strain, the reverse is true in the B6 strain. Similarly, the mutation spectrum of urethane is also shifted in a variety of mutant Ras backgrounds. If the mutagenesis preference of urethane is independent of strain background, the prevalence of the Q61R mutation suggests that this less common mutation is more conducive to tumor initiation in the B6 strain. As such, the most dominant mutation of a mutagen may not always dictate the initiating event, echoing reported discordances between the mutagenic signatures and the putative initiating mutation in certain human cancers.

Another fascinating feature of urethane mutagenesis revealed by MDS sequencing relates to isoform tropism. We found that codon 61 was readily mutated in Hras in lung tissue, yet the oncogenic Hras allele was not expanded appreciably over time. This suggests that either HrasQ61L is not as oncogenic as KrasQ61L or the encoded protein is expressed too low (or high) to be tumorigenic. In support of the first, RAS isoforms differ in their residency at different membranes and the composition of proteins within the immediate vicinity differs between RAS isoforms with proteins like PIP5K1A2, calmodulin, galectin-3, and so forth documented to specifically associate with KRAS. In support of the second, a Kras allele whereby the 3’ end was replaced with Hras exons to encode Hras protein was found mutated in urethane-induced tumors, indicating that under a Kras promoter HrasQ61L is indeed oncogenic in the lung. Whether the inability of oncogenic mutations in Hras to promote lung tumorigenesis is because the protein is less oncogenic, expressed too low, too high, combinations thereof, or for other reasons remains to be elucidated. Nevertheless, the finding that Hras is mutated yet such mutations are not recovered in lung tumors after urethane exposure is in itself an important finding, and perhaps related, of the three RAS genes, HRAS is mutated the least often in human cancers.

With regards to organ tropism, a very different mechanism appears to be at play. In this case, we found that Kras is rarely mutated in the liver and pancreas, despite the presence of the carcinogen. While a number of factors could contribute to this variation in mutagenesis, one notable difference is that Kras mRNA levels are higher in the lung compared with these other tissues, suggestive of increased transcription. In fact, the lung was found to have the second highest levels of Kras mRNA of 15 adult murine tissues analyzed, second only to the brain. Related, we discovered that the non-transcribed strand of Kras is preferentially mutated, which for other mutagens has been linked to transcription-coupled repair of the transcribed strand or transcription-coupled damage of the displaced, non-transcribed strand. Indeed, we found a global correlation between mRNA levels and the mutation frequency of urethane. This is not to say that there is a universal concordance between high gene transcription and an elevated mutation frequency of the non-transcribed strand. Indeed, high transcription has been associated with a lower mutation frequency in chromatin-dense genomic regions in cutaneous squamous cell carcinomas. Thus, the type of cancer, mutational process, specific genes, and so forth may influence the bias of a mutagenic process. In the case of urethane, however, we suggest that the tissue tropism is related to the high transcription of Kras in the lung, increasing the susceptibility of this gene to urethane mutagenesis.

In humans, there are also very distinct patterns to RAS mutations at the level of the organ (e.g., RAS is commonly mutated in pancreatic but rarely in breast cancer), isoform (e.g., KRAS is mutated in lung cancer while NRAS is mutated in melanoma), position (e.g., G12 is mutated in CMMML while Q61 is mutated in thyroid carcinoma), and substitution (e.g., G12V is the primary mutation in bladder carcinoma while it is G12S in mouth carcinoma). There is no definitive mechanism to explain this phenomenon, although the pattern itself has been widely reported for decades. In this regard, the extreme specificity of urethane carcinogenesis for KrasQ61L/R-mutant pulmonary tumors may inform the basic principles of the RAS mutaton patterns observed in these clinical samples. Admittedly, urethane is not a major environmental carcinogen in humans compared with, for example, tobacco smoke. KrasQ61L/R mutations are also rare in human lung cancers. With these two provisos, we speculate that the RAS mutation tropism of human cancers may similarly be a product of mutagenesis selectivity factors, for example the specificity of the mutagenic process or susceptibility of a specific locus to mutations, and selection factors, for example differences in the oncogenic activity of one isoform over another. Moreover, it is entirely possible, if not likely, that different combinations of these or even other factors such as cooperating mutations, as elegantly demonstrated in MNU carcinogenesis, cell type, signaling intensity, and so forth underlie the RAS mutation tropism human cancers. As such, each cancer initiating event may be molded by a unique set of factors, each with varying influence.

Methods

Cell culture. Mouse embryonic fibroblasts (MEFs) derived from E13.5 mouse embryos were stably infected with an ecotropic retrovirus derived from pBabe-Hygro encoding the early region of SV40 and selected with 100 μg·ml⁻¹ hygromycin to establish immortalized cultures using standard procedures and then cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin.

Construction of Kras-mutant plasmids. A region upstream of Kras start codon was amplified from murine genomic DNA (termed PCR1). PCR reactions were comprised of 100 ng of genomic DNA, 2.5 μl of 10 μM forward (5’-AATTCGGC CGCGCAGCAGGGAATAGCTAGCTATGCAGAAAT-3’) and reverse (5’-CATT TCACGCGGGCFTTACAACTGAGAAGT-3’) primers, 4 μl of 2.5 mM DNTP, 10 μl of 5X buffer (NEB), and 0.5 μl Q5® Hot Start High-Fidelity DNA Polymerase (NEB) in a total volume of 50 μl. PCR cycles were as follows: one cycle at 98 °C for 30 s, 28 cycles at 68 °C, 28 cycles at 72 °C, and one cycle at 72 °C for 5 min.
98 °C for 8 s, 64 °C for 15 s, 72 °C for 2 min, and one cycle at 72 °C for 2 min. PCR products were gel purified as described above.

Products from PCR1 and PCR2 were fused through overlap PCR (termed PCR3). Twenty nanograms product from PCR1 and 40 ng product from PCR2 were mixed with 4 μl of 2.5 μM dNTP, 10 μl of 5X buffer (NEB), and 0.5 μl Q5® Hot Start High-Fidelity DNA Polymerase (NEB) in a total volume of 50 μl reaction. PCR conditions were 98 °C for 30 s and 10 cycles 98 °C for 8 s, 63 °C for 15 s, and 72 °C for 15 s. 2.5 μl of forward primer from PCR1 and 2.5 μl of reverse primer from PCR2 were then added and the reaction was continued in the following conditions: 98 °C for 30 s, 25 cycles at 98 °C for 8 s, 72 °C for 40 s, and one cycle at 72 °C for 2 min. PCR products were gel purified as described above.

Plasmid backbone was amplified from the pUC19 (Addgene 50005) plasmid (termed PCR4). PCR reactions were comprised of 1 ng of pUC19 DNA, 2.5 μl of 10 μM forward (5'-AAATTTGACATCAAACACCATGAGGAC-3') and reverse (5'-TAAAGGCGGGGTTTGGTATTGGGGG-3') primers, 4 μl of 2.5 μM dNTP, 10 μl of 5X buffer (NEB), and 0.5 μl Q5® Hot Start High-Fidelity DNA Polymerase (NEB) in a total volume of 50 μl reaction. PCR conditions were 98 °C for 30 s, 28 cycles at 98 °C for 8 s, 65 °C for 15 s, and one cycle at 72 °C for 2 min. PCR products were gel purified as described above.

Products from PCR3 and PCR4 were digested with SalI and NotI according to the manufacturer's protocol (NEB). Digested products were column purified using QIAquick PCR Purification Kit following the manufacturer's protocol (Qiagen) and validated by Sanger sequencing. Ten clones with different sets of co-occurring mutations in Kras exon 1 or/and 2 were selected to be spiked into wild-type mouse genomic DNA at different ratios to test the detection limit of maximum-depth sequencing (see below).

Urethane treatment. Six- to eight-week-old male and female A/J mice (JAX Stock #000646) were intraperitoneally injected daily for 3 days with either urethane (Sigma U2500) and vinyl carbamate (VC) (Santa Cruz Biotechnology sc-213157) concentrations in plasma and tissues. The LC-MS/MS analysis was performed using an Agilent 1200 series UPLC instrument (Agilent Technologies) coupled to an AB SCIEX QTRAP 5500 mass spectrometer (AB SCIEX, Foster City, CA). The UPLC separation was performed on a Shimadzu 20A series LC column (#AJ0-4287) and Agilent ZORBAX Eclipse Plus C18 150 × 4.6 mm 1.8-μm column (Agilent Technologies). The mobile phases were A: 0.1% formic acid in water (pH 2.0) and B: 100% methanol. Elution gradient: isocratic, 100% A from 0.0 to 1.0 min, increased to 75% A at 1.0 min, and continued to 100% A at 1.5 min. The separation was performed at a flow rate of 0.8 ml/min. Flow rate: 0.8 ml/min.

Maximum-depth sequencing (MDS). The MDS assay was adapted for mammalian Kras genes as follows. Twenty to fifty micrograms of genomic DNA was incubated with Stul (NEB) for analysis of the transcribed strand of Kras exon 1, EcoRV (NEB) and EcoRI (NEB) for analysis of the non-transcribed strand of Kras exon 1, and HphI (NEB) for analysis of the non-transcribed strand of Kras exon 2. Reaction conditions were 5 units of the indicated restriction enzyme and per 1 μg DNA per 20 μl reaction (e.g., 20 μg genomic DNA, 5 μl enzyme (20 units/μl)), and 40 μl 10X buffer in 400-μl reaction). Digested genomic DNA was column purified using QIAquick PCR Purification Kit following the manufacturer's protocol (Qiagen) and resuspended in ddH2O (35 μl H2O per 10 μg DNA). The barcode and adaptor were added to the target DNA by incubating purified DNA with the appropriate barcode primer (see below) for one cycle of PCR. PCR reactions were comprised of 10 μg DNA, 2.5 μl of 10 μM barcode primer, 4 μl of 2.5 μM dNTP, 10 μl of 5X buffer (NEB), and 0.5 μl Q5® Hot Start High-Fidelity DNA Polymerase (NEB) in a total volume of 50 μl. The number of PCR reactions was scaled according to the amount of DNA. PCR conditions were 98 °C for 1 min, barcode primer annealing temperature (see below) for 15 s, and 72 °C for 1 min. One microliter of 20,000 U/ml 1 exonuclease I (NEB) and 5 μl of 10X exonuclease I buffer (NEB) was then added to each PCR reaction to remove unincorporated dNTPs. PCR conditions were 98 °C for 1 h and then 80 °C for 20 min. Processed DNA was column purified using QIAquick PCR Purification Kit as above and resuspended in ddH2O (35 μl H2O per column). The concentration of purified product was measured with a NanoDrop spectrophotometer (NEB). Samples were linear amplified with forward adaptor primer (see below) and 2.5 μl of 10 μM forward-adaptor primer, 4 μl of 2.5 μM dNTP, 10 μl of 5X buffer (NEB), and 0.5 μl Q5® High-Fidelity DNA Polymerase (NEB) in a total volume of 50 μl. The number of PCR reactions was scaled according to the amount of DNA. PCR reactions were as follows: 12 cycles of 98 °C for 8 s, 16 cycles of 98 °C for 15 s, 72 °C for 15 s, and 72 °C for 8 s. The final library was size selected and purified with Ampure XP beads according to the manufacturer’s protocol (Beckman Coulter). Sequencing was performed using HiSeq 2500 100 bp PE rapid run, HiSeq 4000 150 bp PE or NovaSeq 6000 S Prime 150 bp PE at Duke Center for Genomic and Computational Biology. For the optimization of barcode recovery, the same amount of genomic DNA was processed in parallel by MDS assay targeting Kras exon 1 transcribed strand and the PCR products were pooled together at different concentrations in one library to obtain different sequencing depths.

Primers for maximum-depth sequencing. Barcode primer: [Forward adaptor] [Index][Barcode][Primer]

Where

[Forward adaptor] = 5'-TACGGGCAACCGGGGATCGTACCTACCTTCCCTACAGGGCTGTCTCTGACTCTCT-3'

[Index] = variable length of known sequences from 0 to 7 nucleotides

[Supplementary Data 1]

[Barcode] = NNNNNNNNNNNNNNNNN

Kras exon 1 StuI [Primer] = 5'-GGGCTGTGGAATAGCTGAG-3' (annealing temperature: 63 °C)

Kras exon 2 PflI [Primer] = 5'-TACAGACTCTCTCAGAAAC-3' (annealing temperature: 63 °C)

Kras exon 2 HphI [Primer] = 5'-GTAGTTGCTCAGTGTACCTG-3' (annealing temperature: 66 °C)

Kras exon 3 XmnI [Primer] = 5'-AATGATTACCGGCGACACCGGAG-3' (annealing temperature: 70 °C)

Exon-specific reverse primer: [Reverse adaptor][Index][Primer]

9
Droplet digital PCR (ddPCR). ddPCR was performed with the QX200 AutoDD Droplet Digital PCR System (Bio-Rad) following the manufacturer’s protocol in a 22 μl ddPCR reaction containing 11 μl of 2X ddPCR SuperMix for probes (no dUTP) (Bio-Rad), 66 ng template DNA, 450 nM forward and reverse primers, and 250 nM FAM- and HEX-labeled probes. The primer and probe oligonucleotides were synthesized (IDT) based on sequences previously described with minor modifications. The sequences for the primers are Kras_3f: 5'-GAGGAGAAGAGGATAGTCTTTCG-3' and Kras_3r: 5'-GATGATGATGACTTCTTCCGAGTTGATTTG-3'. The sequences for the probes are Kras_3f probe: 5'-5FAM-CAAGCTCTGTGATGGGCTTC-3' and Kras_3r probe: 5'-5HEX-GCTAGCTGCTCTTCACT-3' where "5" denotes the following base is a locked nucleic acid. Following droplet generation on the AutoDD, the plate was sealed with pierced foil having a 0.2 μl hole in the center (Bio-Rad) and PCR amplification was performed on a C1000 Touch™ thermal cycler (Bio-Rad). Thermal cycling conditions were as follows: once cycle at 95 °C for 10 min, 40 cycles at 94 °C for 30 s and 60 °C for 60 s, once cycle at 98 °C for 10 min, and 4 °C until the sample was removed. Every ddPCR run included no template control, wild type control with DNA from PBS-treated mice, and mutation-positive control. The assay detection sensitivity of 1 in 10,000, each sample contained at least two wells. Plates were read on a QX200 droplet reader (Bio-Rad) and analyzed with QuantaSoft® Analysis Pro software (version 1.0.596) (Bio-Rad) to assess the number of droplets positive for mutant DNA, wild-type DNA, both, or neither. The mutant allele fraction was estimated as follows: The concentration of mutant DNA (copies of mutant DNA per droplet) was estimated from the Poisson distribution using the formula number of mutant copies per droplet Mmu = ln (1−(mut/nmu)), where nmu = number of droplets positive for mutant FAM probe and n = total number of droplets. The DNA concentration in the reaction was estimated using the formula MdnAconc = ln(1−([dAconc/n]), where dAconc = number of droplets positive for wild-type HEX probe and n = total number of droplets. The mutant allele fraction was Mmu/ MdnAconc.

RNA isolation and quantitative PCR. RNA was extracted from the lung, liver and pancreas of 6-week-old A/J mice using TRIzol® (Thermo Fisher Scientific) and converted to cDNA using iScript™ cDNA Synthesis Kit (Bio-Rad) following the manufacturer’s instructions. Quantitative PCR reactions were performed using iTaq Universal SYBR Green Supermix (Bio-Rad) and CFX98 touch real-time PCR detection system (Bio-Rad) using the forward (5'-CCAGGCTCTGTGATTAGCGA-3') and reverse (5'-CCAGGCTGTCCGAGCAAAGAC-3') primers (Bio-Rad) to co-denature the Hprt mRNA and the forward (5'-GCAAGAGGGCTTGACGAGC-3') and reverse (5'-CATGTACTGCTCCATCATGGAC-3') primers (IDT) to detect Kras mRNA. Gene expression values were calculated using the comparative Ct (ΔΔCt) method, using Hprt housekeeping gene as internal control.

Whole-exome analysis of mutation frequency versus gene expression. Mutation counts were obtained from published datasets [https://www.nature.com/articles/nature18898#sec19]. Single-nucleotide variations (SNVs) identified by the whole-exome sequencing of uracil-induced adenomas and adenoscarcinomas were examined. The expression level of the genes containing these SNVs were determined from published datasets [https://www.nature.com/articles/s41598-017-04520-2#sec18]. FPKM values of genes expressed in the lung of 6-week-old C57BL/6J/cd mice were used. The second set of gene expression data were obtained from mouse ENCODE project [http://chromosome.sdsc.edu/mouse/download.html]. FPKM values of genes expressed in the lung of 8-week-old male C57BL/6j mice were used. To bin the genes into different expression groups, the genes were sorted by the mean FPKM value across biological replicates and split into 4 quartiles. The sum of C>T substitutions at an individual nucleotide was obtained, then the fold change of each type of substitutions at an individual nucleotide was obtained, then the fold change of each type of substitutions was calculated and log10 transformed and plotted.

Generation of heatmaps. All heatmaps were generated using Morpheus (https://software.broadinstitute.org/morpheus). All mutation frequencies used in heatmap were corrected by the addition of the detection limit at a barcode recovery of 1.5 × 10−6. The heatmap showing the mutation frequency per nucleotide (Fig. 1b), the sum of the corrected mutation frequencies for all substitutions at an individual nucleotide was obtained, then the fold change of each type of substitution was calculated and log10 transformed for plotting. For the heatmaps showing the summary or frequency of each type of substitution (Figs. 1c, 2a, and 3a), the sum of the corrected mutation frequencies for each type of substitutions was obtained, then the fold change of each type of substitution was calculated and log10 transformed for plotting. For the heatmaps showing the frequency of A>T transversions (Figs. 1d, 2b, and 3b; Supplementary Figs. 1a and 3a), the corrected mutation frequency for each A>T transversion was log10 transformed and plotted.

Statistics. The number of independent experiments and the statistical analysis used are indicated in the legends of each figure. Data are represented as mean ± SEM. p values were determined by Holm-Sidak multiple comparisons test following one-way or two-way ANOVA, non-parametric Dunn’s multiple comparison test following Kruskal–Wallis test, and Tukey’s multiple comparison test following Graphpad® software. The Dunn’s multiple comparisons test following Kruskal–Wallis test was executed using Graphpad® software. The Mann–Whitney U test was executed using excel supplemented with Real Statistics Resource Pack (www.real-statistics.com).
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All raw Illumina® sequencing data has been deposited to NCBI Sequence Read Archive (SRA) under accession number PRJNA561927. The remaining data can be found within the Article, Supplementary Information or available from the authors upon reasonable request.

Received: 6 September 2019; Accepted: 23 March 2020; Published online: 14 April 2020

References
1. You, M., Candrian, U., Maronpot, R. R., Stoner, G. D. & Anderson, M. W. Activation of the K-ras protooncogene in spontaneously occurring and chemically induced lung tumors of the strain A mouse. Proc. Natl. Acad. Sci. USA 86, 3070–3074 (1989).
2. Forkert, P. G. Mechanisms of lung tumorigenesis by ethyl carbamate and vinyl carbamate. Drug Metab. Rev. 42, 355–378 (2010).
3. Westcott, P. M. et al. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. Nature 517, 489–492 (2015).
4. Nuzum, E. O., Malkinson, A. M. & Beer, D. G. Speciﬁcity of rare mutations with massively parallel sequencing.
5. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 72, 2457–2467 (2012).
6. Hernandez, L. G. & Forkert, P. G. In vivo mutagenicity of vinyl carbamate and its role as an intermediate in the formation of 1,6-ethenoadenosine. Chem. Res. Toxicol. 4, 413–421 (1991).
7. Buir, C. E. et al. Mutation-scpictural patterns of lung carcinogenesis in mice. Nature 400, 141–145 (2002).
8. Hernandez, L. G. & Forkert, P. G. Oxidation of vinyl carbamate and formation of 1,6-ethenoadenosine. Mol. Carcinog. 17, 1711–1718 (1996).
9. Buir, C. E. et al. The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462 (1997).
10. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 592–600 (2002).
11. Smittle, W. M. et al. Detection of ultra-rare mutations by next-generation sequencing. Proc. Natl. Acad. Sci. USA 109, 14508–14513 (2012).
12. Lou, D. E. et al. The comprehensive mouse transcriptomic bodymap across 17 tissues by RNA-seq. Sci. Transl. Med. 9, eaan2415 (2017).
13. Feng, C. et al. Identification of recurrent somatic RAC1 mutations in melanoma. Nat. Genet. 44, 1006–1014 (2012).
14. Reuter, V. E. et al. Es PIP5K1A as a KRAS-speciﬁc genotyping of KRAS mutant non-small cell lung cancer using a multiplexed droplet digital PCR approach. PLoS ONE 11, e0159074 (2015).
15. Krauthammer, M. et al. Exome sequencing identiﬁes recurrent somatic RAC1 mutations in melanoma. Nat. Genet. 44, 1006–1014 (2012).
16. Persing, N. L. et al. Rare codons capable Kras-driven de novo tumorigenesis. J. Clin. Invest. 125, 222–233 (2015).
17. Huang, L., Carney, J., Cardona, D. M. & Counter, C. M. Decreased lung tumorigenesis in mice with a Kras point mutation at C118. Cancer Res. 55, 8604–8611 (2015).
18. Temple, D., Tomlinson, I. P. M., Severini, S., Schuster-Bockler, B. & Hagiwara, K. The detection of chemically initiated cells having the mutation of K-ras gene at an early stage of lung carcinogenesis in mice. Anal. Biochem. 244, 187–189 (1997).
19. Blattner, F. R. et al. The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462 (1997).
20. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 592–600 (2002).
21. Hindson, B. J. et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal. Chem. 83, 8604–8610 (2011).
22. Hernandez, R. C., Nair, J., Barbin, A., Miller, J. A. & Bartsh, H. Detection of 1,6-ethenoadenosine and 3,6-ethenoectydine by immunoaﬃnity/32P-postlabelling in liver and lung DNA of mice treated with ethyl carbamate (urethane) or its metabolites. Carcinogenesis 17, 1711–1718 (1996).
23. Forkert, P. G. et al. Oxidation of vinyl carbamate and formation of 1,6-ethenoadenosine in murine lung. Drug Metab. Dispos. 35, 713–720 (2007).
24. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 47, D941–D947 (2019).
25. Winters, L. P. et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantitation of Kras variant oncogenicity. Nat. Commun. 8, 2053 (2017).
26. Buir, C. E. et al. The germine control of somatic Kras mutations in mouse lung tumors. Mol. Carcinog. 57, 754–751 (2018).
27. Butler, C. E. et al. Mutation-speciﬁc RAS oncogenicity explains NRAS codon 61 selection in melanoma. Cancer Discov. 4, 1418–1429 (2014).
28. Kiesling, M. K. et al. Mutant HRAS as novel target for MEK and mTOR inhibitors. Oncotarget 6, 42183–42196 (2015).
29. Li, S., Balmain, A. & Counter, C. M. A model for RAS mutation patterns in cancers: ﬁnding the sweet spot. Nat. Rev. Cancer 18, 767–777 (2018).
58. Stephen, A. G., Esposito, D., Bagri, R. K. & McCormick, F. Dragging ras back in the ring. Cancer Cell 25, 272–281 (2014).
59. Adar, S., Ha, J., Lieb, J. D. & Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proc. Natl Acad. Sci. USA 113, E1214–E1213 (2016).
60. Hoﬄer, U. & Ghanayem, B. I. Increased bioaccumulation of urethane in CYP2E1−/− versus CYP2E1+/+ mice. Drug Metab. Dispos. 33, 1144–1150 (2005).
61. Supek, F. & Lehner, B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. Cell 170, 534–547 (2017).
62. To, M. D. et al. A functional switch from lung cancer resistance to tumorigenic web. Cancer Cell 25, 272–281 (2014).
63. Dassano, A. et al. Mouse pulmonary adenoma susceptibility 1 locus is an expression QTL modulating Kras-4A. PLoS Genet. 10, e1004307 (2014).
64. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat. Rev. Mol. Cell Biol. 9, 958–970 (2008).
65. Zheng, C. L. et al. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. Cell Rep. 9, 1228–1234 (2014).
66. Pylaeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. Nat. Rev. Cancer 11, 761–774 (2011).
67. Haigis, K. M. KRAS alleles: the devil is in the detail. Cancer Res. 77, 547–549 (2017).
68. Bos, J. L. ras oncogenes in human cancer: a review. Cancer Res. 49, 4682–4689 (1989).
69. Xu, X. et al. Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. Proc. Natl Acad. Sci. USA 109, 4910–4915 (2012).
70. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res. 18, 3587–3596 (1990).
71. O’Hayer, K. M. & Counter, C. M. A genetically deﬁned normal human somatic cell system to study ras oncogenes in vivo and in vitro. Methods Enzymol. 402, 637–647 (2006).
72. Norrander, J., Kempe, T. & Messing, J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26, 101–106 (1983).
73. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544 (2018).
74. Zhang, J., Kobot, K., Fouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd merger. Bioinformatics 30, 614–620 (2014).
75. Arbeithuber, B., Makova, K. D. & Tiemann-Boege, I. Artifactual mutations performing the quality check and sequencing of NGS libraries.
76. Rowlands, V. et al. Optimisation of robust singleplex and multiplex droplet digital PCR assays for high conﬁdence mutation detection in circulating tumour DNA. Sci. Rep. 9, 12620 (2019).
77. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Method Cell 25, 402–408 (2001).