JNCI J Natl Cancer Inst (2018) 0(0): pky002

doi: 10.1093/jncics/pky002
Article

ARTICLE

Mutation Signatures Including APOBEC in Cancer Cell Lines

Matthew C. Jarvis, Diako Ebrahimi, Nuri A. Temiz, Reuben S. Harris

Affiliations of Authors: Masonic Cancer Center (MCJ, DE, NAT, RSH), Center for Genome Engineering (MCJ, DE, RSH), Department of Biochemistry, Molecular Biology, and Biophysics (MCJ, DE, RSH), Institute for Molecular Virology (MCJ, DE, RSH), Institute for Health Informatics (NAT), Howard Hughes Medical Institute (RSH), University of Minnesota, Minneapolis, MN

Correspondence to: Reuben S. Harris, Ph.D., University of Minnesota, Masonic Cancer Center, 2231 6th St. S.E., CCRB Room 4-230, Minneapolis, MN 55455 (rsh@umn.edu); or Nuri A. Temiz, Ph.D., University of Minnesota, Masonic Cancer Center, 2231 6th St. S.E., CCRB Room 4-230, Minneapolis, MN 55455 (temizna@umn.edu).

Abstract

Background: Multiple endogenous and exogenous sources of DNA damage contribute to the overall mutation burden in cancer, with distinct and overlapping combinations contributing to each cancer type. Many mutation sources result in characteristic mutation signatures, which can be deduced from tumor genomic DNA sequences. Examples include spontaneous hydrolytic deamination of methyl-cytosine bases in CG motifs (AGEING signature) and C-to-T and C-to-G mutations in 5'-TC(A/T) motifs (APOBEC signature).
Methods: The deconstructSigs R package was used to analyze single-base substitution mutation signatures in more than 1000 cancer cell lines. Two additional approaches were used to analyze the APOBEC mutation signature.
Results: Most cell lines show evidence for multiple mutation signatures. For instance, the AGEING signature, which is the largest source of mutation in most primary tumors, predominates in the majority of cancer cell lines. The APOBEC mutation signature is enriched in cancer cell lines from breast, lung, head/neck, bladder, and cervical cancers, where this signature also comprises a large fraction of all mutations.
Conclusions: The single-base substitution mutation signatures of cancer cell lines often reflect those of the original tumors from which they are derived. Cancer cell lines with enrichments for distinct mutation signatures such as APOBEC have the potential to become model systems for fundamental research on the underlying mechanisms and for advancing clinical strategies to exploit these processes.

DNA damage and mutagenesis are enabling hallmarks of cancer broadly characterized as genome instability (1). Advances in DNA sequencing have identified both common and rare genomic alterations that define each type of cancer and, in some instances, identify therapeutic vulnerabilities. Sequencing has also provided deeper insights into mutation patterns in cancer and, together with bioinformatics approaches, has allowed deconvolution of these complex patterns into more than 25 mutation signatures (reviewed by [2–6]). Each distinct mutation signature suggests mechanistic linkage to a single source of DNA damage. For instance, the largest endogenous source of DNA damage and mutation affecting most cancer types is spontaneous deamination of methyl-cytosine bases in 5'-CG dinucleotide contexts yielding C-to-T mutations. This process is water-mediated and occurs spontaneously over time, and it therefore associates with the biological age of the patient (AGEING signature) (7). The second largest endogenous mutation source across cancer is attributed to APOBEC-catalyzed deamination of cytosine bases to uracil in 5'-TCW contexts (W = A or T), which, coupled with DNA repair processes, yields both C-to-T and C-to-G mutations (APOBEC signature) (7–11). APOBEC enzymes are normally involved in innate antiviral immunity, but at least two family members contribute to cancer mutagenesis (reviewed in [5,12]). In comparison, ultraviolet (UV) radiation is an exogenous source of DNA damage and mutation largely specific to skin cancer and generally characterized by the cross-linking of adjacent pyrimidine bases, which results in C-to-T mutations in 5'-YC dinucleotide contexts (Y = C or T; UV signature).

Received: September 19, 2017; **Revised:** December 12, 2017; **Accepted:** January 6, 2018

Cancer cell lines often maintain key properties of the tumor cells from which they are derived. This characteristic enables a wide range of mechanistic and preclinical studies. An excellent example is the use of *BRCA1* mutant cell lines to discover synthetic lethality upon chemical or genetic inhibition of the poly-ADP ribose polymerase 1 (PARP1) (13,14). BRCA1 deficiencies are now the prototypical example of a much broader category of homologous recombination deficiencies in cancer (HRD or BRCA signature) (7,15,16). Importantly, tumors with an HRD signature are often susceptible to PARP inhibitors, even when an underlying recombination defect cannot be defined (16). By analogy, it is likely that selecting the best cell lines to study other mutational processes, such as APOBEC, will be instrumental for delineating mechanisms and translating fundamental knowledge into clinical benefits.

Here we present comprehensive analyses of the single-base substitution mutation signatures in more than 1000 cancer cell lines through the COSMIC database. Overall, the single-base substitution mutation signatures of cancer cell lines reflect those that are also evident in the original tumor types from which they were derived. We are hopeful that these analyses will become a resource for the community as work advances to better understand the mutation mechanisms in cancer and develop strategies to leverage this knowledge for prognostic and diagnostic tests as well as for development of novel therapeutics.

Methods

Mutation Data Source, Reformatting, and Filtering

The file "CosmicCLP_MutantExport.tsv," version 81, was downloaded on July 17, 2017, from the COSMIC cell line project (CLP) online database using the Sanger Institute SSH file transfer protocol (http://cancer.sanger.ac.uk/cell_lines). The mutation data of 1020 cell lines were formatted and filtered as described in the following sections. The code for these operations is available from GitHub (https://github.com/mcjarvis/Mutation-Signatures-Including-APOBEC-in-Cancer-Cell-Lines-JNCI-CS-Supplementary-Scripts). To further facilitate the reproducibility of our analyses, the mutation data from three cell lines are used as examples in this section (BC-3, BT-474, and NALM-6), and all key raw and processed numbers are provided in Supplementary Tables 1–3 (available online).

Step 1: Download, organize, and filter raw mutation data: The fields cell line name (column 5), mutation (column 18), mutation type (column 20), version of the reference genome (column 23), chromosome position of the mutation (column 24), and DNA strand (column 25) were extracted from the "CosmicCLP_MutantExport.tsv" file using the following command:

```
awk 'BEGIN{FS="\t"; OFS="\t"}; 0 !~ 7#/ {print $5, $18, $20,
$23, $24, $25}'CosmicCLP_MutantExport.tsv > cosmic_mut.txt
```

Step 2: Removal of all non-single-base substitution mutations: All mutations that are not single-base substitutions (eg, insertions, deletions, and complex multibase substitutions) were filtered out of the table, leaving single-base substitution mutations annotated as nonsense, missense, or coding silent substitutions. This essential filtering step reduced the number of mutations in BT-474, BC-3, and NALM-6 from 1595 to 1407, 1537 to 1371, and 3291 to 2962, respectively.

Step 3: Additional filtering to remove nonunique chromosomal positions and file reformatting: All nonunique chromosome positions were filtered out of each cell line individually, which ensures that each mutation has only one associated chromosomal position within a cell line. A tab-separated file was created with chromosome number (eg, "chr1"), chromosomal position, reference allele, alternate (mutant) allele, strand of the substitution, and sample (cell line name) as columns. This table was reordered as follows for subsequent analyses: chr1-chr9, chrX, chrY, chr10-chr22, then by ascending chromosomal position, and it was then saved as a text file. This step reduced mutation numbers in BT-474, BC-3, and NALM-6 from 1407 to 1021, 1371 to 963, and 2962 to 2110, respectively.

Step 4: Filter out "nonmutations" and "nonmatching mutations": The hg38 reference genome (FastA file, GCA_000001405.2) was used to filter out nonmutations and nonmatching mutations (downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/ on July 16, 2017). Nonmutations are instances in which the alternate (mutant) allele matches the reference genome at that position. Nonmatching mutations are instances in which the reference allele does not match the reference genome at that position. These anomalies were filtered out of the single-base substitution mutation data set. This step caused a modest reduction in mutation numbers in a small number of cell lines. For instance, the numbers above for BT-474 and BC-3 were unchanged, but the number for NALM-6 reduced from 2110 to 2108. These single-base substation numbers were used to plot medians in Figure 1, and raw values are listed for each trinucleotide context in Supplementary Table 1 (available online). Following the filtering steps described above, the total single-base substitution mutation count across all cell lines was 663 075.

COSMIC Mutation Signature Determinations

The deconstructSigs R package (17) was used to determine the mutational signatures in each cancer cell line, including AGEING (signature 1) and APOBEC (signatures 2 plus 13). First, hg38 was used to place each single-base substitution mutation from step 4 above into an appropriate trinucleotide context. Second, the proportion of each distinct base substitution mutation occurring within a given trinucleotide context was determined. Finally, weights for each distinct mutation signature were determined using the signatures.cosmic reference file (https://github.com/raerose01/deconstructSigs). Signature visualization was done using the ggplot2 package in R (Figure 2) (18). Mutation loads were also analyzed using a linear model with a local polynomial regression fitting ($\alpha = .75$) and 95% highest posterior density confidence interval. Supplementary Table 2 (available online) shows the proportion of each COSMIC mutation signature in each cancer cell line.

Alternative Methods for Quantifying the APOBEC Mutation Signature

Two additional methods were used for quantification of the APOBEC mutation signature. The first is simply counting and calculating the proportions of C-to-T and C-to-G mutations in TCW motifs in the filtered single-base substitution mutation data sets described above. The second is an enrichment analysis (19). Briefly, C-to-A mutations are filtered out, as these are not relevant to APOBEC mutagenesis. Then a formatted text file containing single-base substitutions was used to generate a table with the following information: chromosome, position, trinucleotide context, base substitution (ie, ref > alt), strand sense, flanking the 20-base sequence on each side (ie, 20 nucleotides

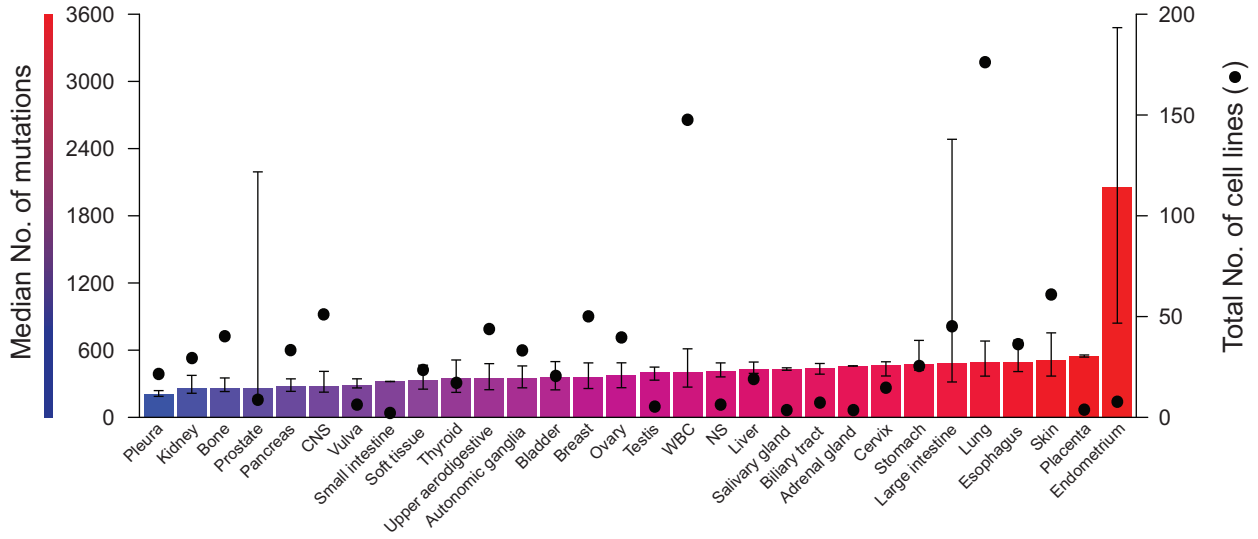

Figure 1. Single-base substitution mutation burdens in COSMIC cancer cell lines. Median numbers of base substitution mutations (**bars**) and absolute numbers of cancer cell lines (**filled circles**) for the indicated tumor types (central nervous system, nonspecified origin, white blood cell). **Whiskers** represent variation in terms of central quartiles below and above the median, respectively (ie, 25th–50th and 50th–75th percentiles). Large variations in mutation numbers are due to mismatch repair deficiencies in a subset of cell lines (large intestine) and to small numbers of cell lines (prostate and endometrium). CNS = central nervous system; NS = nonspecified origin; WBC = white blood cell.

on both the 5' and 3' ends of the substitution, totaling 41 base sequence), cytosine base counts within this 41-base sequence, and TCA and TCT trinucleotide contexts within the 41-base sequence. The formatted text file was used as input to calculate the above variables, using the following command (BT-474 is used as an example):

perl count_trinuc_muts_v8.pl pvcf hg38.fa bt474_mut_sort.txt

Finally, this table was imported into R to calculate the enrichment ratio and subsequent significance, defined as (19):

$$APOBEC\ Enrichment = \frac{Mut(TCW)/Con(TCW)}{Mut(C)/Con(C)}$$

In the above equation, Mut(TCW) is any substitution at a TCA or TCT trinucleotide context, and Con(TCW) is any TCT or TCA trinucleotide sequence that appears in the 41-base analysis window (20 bp on 5' and 3' sides of each single-base substitution). Mut(C) and Con(C) represent substitutions at cytosine and number of cytosine bases in the 41-base window, respectively. The values in the above equation were formatted into a contingency table, and a Fisher exact test was conducted to determine the statistically significant difference between the two ratios (Fisher *P* value). A Benjamini-Hochberg correction was applied to these *P* values to generate adjusted *q* values. Cell lines with a *q* value of .05 or less were considered enriched for APOBEC signature mutations (n = 129). Following the examples above, the *q* values for BT-474 and BC-3 are 1.2×10^{-98} and 7.3×10^{-90}, respectively, whereas the *q* value for NALM-6 is 1.0 (not statistically significant). Supplementary Table 3 (available online) provides a comparative listing of the proportions of APOBEC signature mutations determined using COSMIC, simple counting, and the enrichment approaches described above.

Results

The COSMIC Cell Line Project has generated whole-exome sequences and mutation data for 1020 cell lines (20). A total of 30 different tissue/tumor types are represented, including a category for cell lines of nonspecified (NS) origin (Figure 1). After filtering to remove insertions, deletions, multinucleotide mutations, and other complex mutations, median numbers of single-base substitution mutations ranged from 201 in cancer cell lines of the pleura to 2062 in those from the endometrium (Figure 1; raw numbers in Supplementary Table 1, available online). Representative examples include 1021 single-base substitution mutations in the breast cancer cell line BT-474, 963 single-base substitution mutations in the pleural effusion lymphoma cell line BC-3, and 2108 single-base substitution mutations in the B-cell line NALM-6. One lung cancer cell line, NCI-H1395, had fewer than 50 single-base substitution mutations and was excluded from subsequent analyses to minimize inaccurate signature assignments.

Most cancer cell lines show evidence for multiple mutational processes (see COSMIC signature depictions in Figure 2; raw numbers and signature proportions in Supplementary Tables 1 and 2, available online). The most common mutation signature in tumors is AGEING (COSMIC signature 1) (7). Similarly, for the six tumor/tissue types represented by more than 50 cancer cell lines, the proportion with an AGEING signature ranges from 89% (49/55, large intestine) to 100% (57/57, central nervous system [CNS]). Moreover, the percentage of AGEING signature mutations within each cancer cell line varies from as low as the 6% detection limit to as high as 80% in one of the CNS lines (SW1783). In comparison, the mutational contribution from APOBEC (COSMIC signatures 2 plus 13) is more variable, ranging from undetectable in cancer cell lines of the CNS (0/57) and large intestine (0/55), to intermediate levels in cancer cell lines of the white blood cells (WBCs; 5%, 9/178), skin (5%, 3/60), and lung (19%, 34/175), and high in breast cancer cell lines (48%, 25/52). The percentage of APOBEC signature mutations in each cancer cell line is similarly variable, ranging from a low of 6% (detection limit) to a high of 59% (BC-3). However, these analyses reveal a strong positive correlation in breast cancer cell lines between the proportions of APOBEC signature mutations and the overall base substitution mutation loads (Pearson's correlation $P \leq .005$)

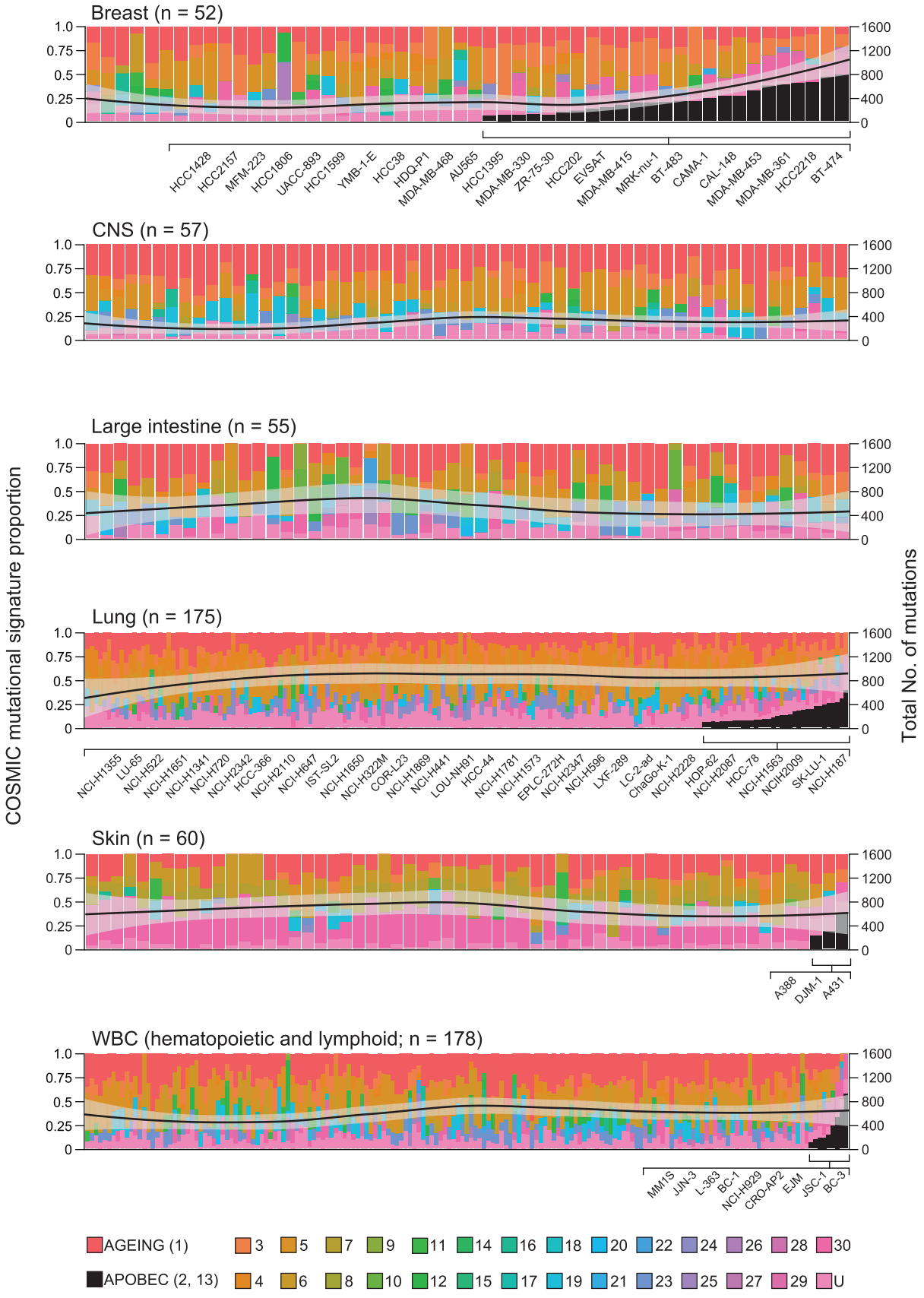

Figure 2. Mutation signatures in COSMIC cancer cell lines. Stacked **bar graphs** showing the proportion of each COSMIC mutational signature in cancer cell lines representing the indicated tumor/tissue types (n > 50 per condition) (see Supplementary Tables 1–3, available online, for exact values and for data for tumor/tissue types

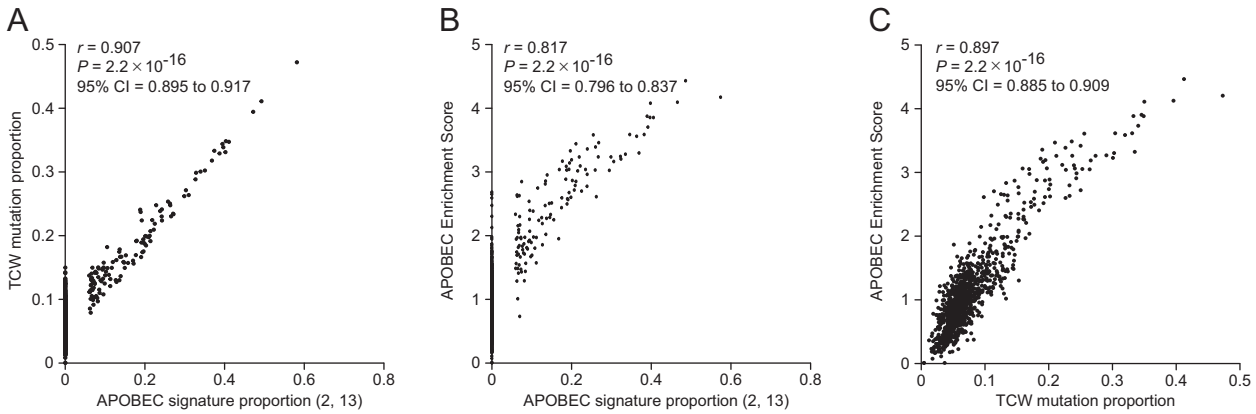

Figure 3. A comparison of methods used to quantify the APOBEC mutation signature. Scatterplots comparing the relationship between **(A)** the proportion of COSMIC signature 2 and 13 mutations and the proportion of C-to-T and C-to-G mutations in TCW motifs, **(B)** the proportion of COSMIC signature 2 and 13 mutations and the APOBEC enrichment score, and **(C)** the proportion of C-to-T and C-to-G mutations in TCW motifs and the APOBEC enrichment score. Pearson's product moment correlation coefficient, corresponding *P* value, and 95% confidence intervals are shown for each comparison. CI = confidence interval.

(black histogram bars and black line in the top panel of Figure 2).

Multiple computational approaches are being used to extract mutation signatures from complex data sets. For instance, the overall impact of the APOBEC mutation process has been estimated using the deconstructSigs R package (above) (17), enrichment score calculations (10), and simple counting of C-to-T and C-to-G mutations in APOBEC-preferred TCW trinucleotide motifs (9,21). To assess the utility of each approach, each method was used to independently analyze the filtered single-base substitution data described above. Interestingly, pairwise comparisons of the data from each analysis yielded highly significant positive correlations (Pearson's test, $P < 2.2 \times 10^{-16}$) (Figure 3; Supplementary Table 3, available online). These comparisons indicate that each approach is robust, although it should be noted that the deconstructSigs package requires a signature to exceed a threshold of 6% before a proportion can be assigned.

Another way to represent mutation signatures is by visualizing mutated trinucleotide motifs that define each source of mutation (Figure 4). For instance, cancer cell lines such as HCC1419 and CAL-33 harboring predominantly AGEING signature mutations are readily identifiable by large proportions of C-to-T mutations in NCG motifs (ACG, CCG, GCG, and TCG; denoted by asterisks in Figure 4). In most cases, the proportion of C-to-T mutations in each of these trinucleotide motifs is similar, reflecting the underlying mechanism of water-mediated deamination of methyl-CG motifs, which is not substantially influenced by the identity of the base pair immediately upstream of the mutated cytosine base. In contrast, cancer cell lines such as HCC2218 and NCI-H2009 show clear APOBEC mutation signatures, with large proportions of C-to-T and C-to-G mutations in both the TCA and TCT motifs (highlighted by boxes in Figure 4). As expected, many cancer lines such as HDQ-P1 and EJM show a mixture of these two mutation signatures (Figure 4). Finally, other cancer cell lines also show two or more combinations of mutation signatures (Figure 2; Supplementary Table 2, available online).

Discussion

Here we report a comprehensive analysis of mutation signatures in the COSMIC cancer cell line collection, currently comprising 1020 cell lines. As expected, most cell lines show evidence for multiple mutation signatures. The AGEING signature is the most common and, in many instances, also the most abundant in the majority of cancer cell lines. This likely reflects both the burden of water-mediated methyl-cytosine deamination events that accumulate during formation of the original tumors and additional spontaneous events occurring over time during, in many instances, hundreds of passages in cell culture. In comparison, several mutation signatures, such as the APOBEC signature, are less common and only enriched in a subset of cancer cell lines from a subset of tissue/tumor types. For instance, the APOBEC mutation signature is prevalent in more than 100 cancer cell lines, with contributions in less than 5% of cell lines for several cancer types to nearly 50% of cell lines from breast tumors. This overall distribution largely reflects the reported occurrence of APOBEC mutagenesis in many different cancer types, but to larger extents in tumors of the breast, lung, head/neck, bladder, and cervix (7–11).

One challenge with absolute quantification is that some mutational processes have overlapping signatures (for a recent review, see [6]). The AGEING and APOBEC signatures provide a clear example of this problem. The spontaneous deamination of methyl-C-to-T by water in CG motifs results in the AGEING signature. This occurs at similar frequencies in each of the four different NCG trinucleotide motifs, which is expected from the spontaneous nature of the underlying chemical process. In comparison, APOBEC enzymes have preferences for deamination of cytosine bases in a subset of trinucleotide motifs.

Figure 2. Continued
represented by less than 50 cell lines). COSMIC signature designations are color coded in the legend. COSMIC signatures 2 and 13 combine to make up the APOBEC signature, and cell lines with this signature are grouped to the right in order of ascending APOBEC signature proportion (**black bars**). The **black line** represents the results of a smoothed linear model ($\alpha = .75$) of the total number of mutations within each cell line (surrounding **lighter region** represents a 95% confidence interval). The positive relationship between increasing mutation numbers and APOBEC signature proportions are statistically significant only in the panel of breast cancer cell lines (Pearson's correlation, $P \leq .005$). CNS = central nervous system; U = unassigned; WBC = white blood cell.

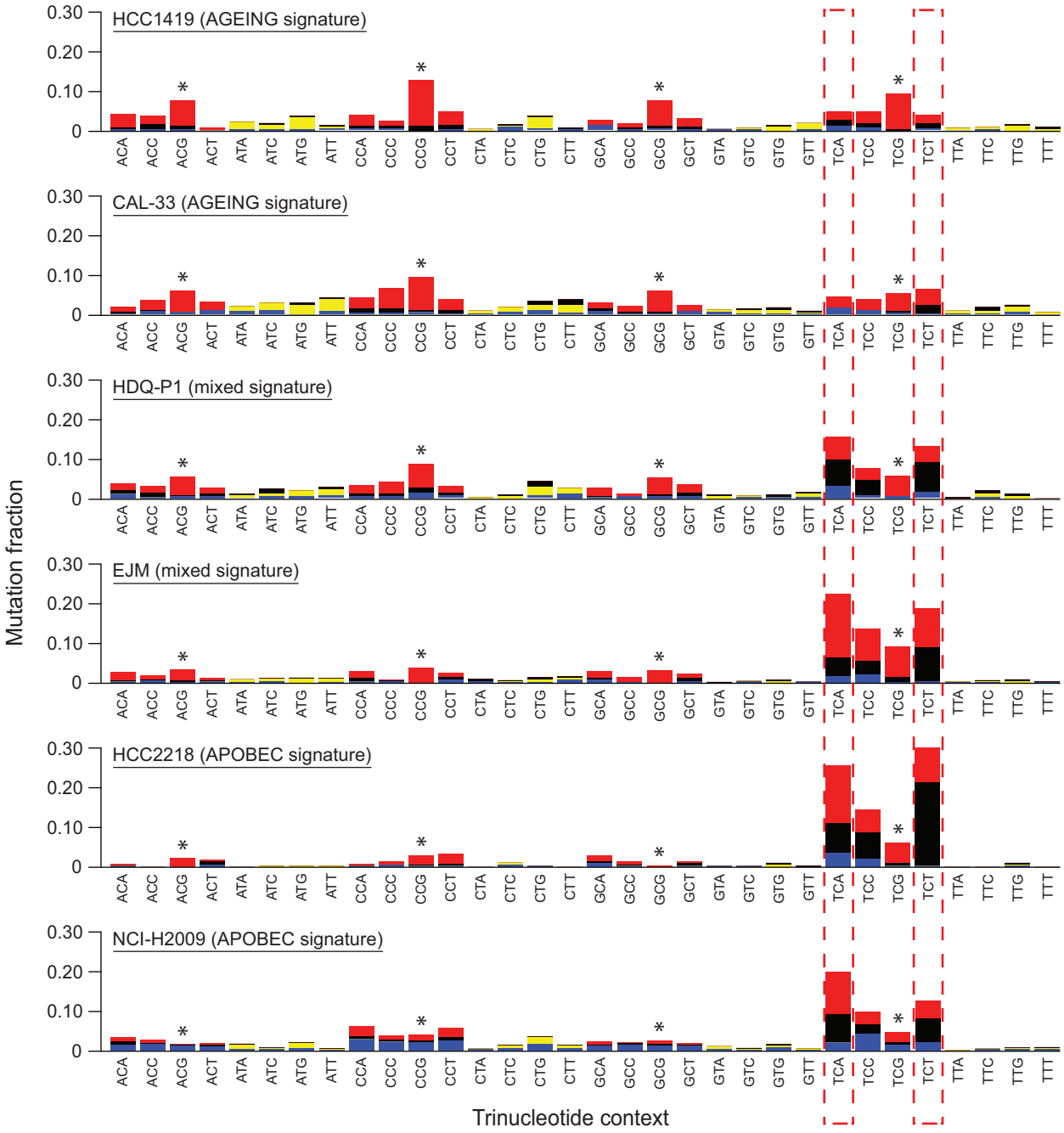

Figure 4. Single-base substitution profiles for cancer cell lines with AGING, APOBEC, and mixed mutation signatures. Stacked **bar plots** showing the proportion of each of the six different single-base substitutions in 32 possible trinucleotide motifs in the indicated cancer cell lines. Mutations from the central base to A, C, G, and T are represented by **blue, yellow, black, and red shaded bars**, respectively. **Asterisks and red dashed boxes** are used to highlight motifs characteristic of AGEING and APOBEC mutation processes, respectively. See the text for details and Supplementary Table 3 (available online) for raw counts.

For instance, APOBEC3B has a strong biochemical preference for TCA and TCG and a considerably weaker preference for TCC and TCT (8,21). Further complicating matters, APOBEC3B has the biochemical capacity to deaminate methyl-cytosine bases (22–24), and there is no way to know unambiguously whether the original cytosine in a CG context was methylated prior to deamination. Thus, both spontaneous and APOBEC-catalyzed processes lead to C-to-T mutations in TCG motifs (eg, EJM in Figure 4). This overlap can be ignored by restraining APOBEC signature events to TCA and TCT (TCW) motifs, as in the enrichment protocol (19), or included by estimating the contribution of each process by subtracting the average proportion of AGEING mutations in the other three trinucleotide motifs from the total proportion of C-to-T mutations in TCG motifs. Alternatively, an unambiguous signature "diagnosis" without absolute quantification may be possible by focusing on mutational events that are unique to each mechanism, such as detection of clustered C-to-G transversions in TCW motifs for APOBEC.

As described here for mutation signatures, cancer cell lines often maintain key properties of the tumor cells from which

they are derived. This in turn has the tremendous benefit of enabling a wide range of mechanistic and preclinical studies. An excellent example is the use of *BRCA1* mutant cell lines to discover synthetic lethality upon chemical or genetic inhibition of the poly-ADP ribose polymerase 1 (PARP1) (13,14). BRCA1 deficiencies are now the prototypical example of a much broader category of homologous recombination deficiencies in cancer (HRD or BRCA signature) (7). Importantly, tumors with an HRD signature are often susceptible to PARP inhibitors, even when an underlying recombination defect cannot be defined (16). By analogy, it is likely that selecting the best cell lines to study other mutational processes, such as APOBEC, will be instrumental for delineating mechanisms and translating fundamental knowledge into clinical benefits.

Funding

This work was supported in part by NCI R21 CA206309. Partial salary support for MCJ was provided by National Cancer Institute (NCI) T32 CA009138 and for NAT by NCI P30 CA77598 through the University of Minnesota Masonic Cancer Center. RSH is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished McKnight University Professor, and an Investigator of the Howard Hughes Medical Institute.

Notes

We thank Dr. Rena Levin-Klein for comments on the final draft of the manuscript. RSH is a co-founder, shareholder, and consultant of ApoGen Biotechnologies Inc. The other authors declare no competing financial interests.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646–674.
2. Alexandrov LB, Stratton MR. Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev*. 2014;24:52–60.
3. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*. 2014;15(9):585–598.
4. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: Footprints and mechanisms. *Nat Rev Cancer*. 2014;14(12):786–800.
5. Swanton C, McGranahan N, Starrett GJ, et al. APOBEC enzymes: Mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov*. 2015;5(7):704–712.
6. Nik-Zainal S, Morganella S. Mutational signatures in breast cancer: The problem at the DNA level. *Clin Cancer Res*. 2017;23(11):2617–2629.
7. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.
8. Burns MB, Lackey L, Carpenter MA, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013;494(7437):366–370.
9. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*. 2013;45(9):977–983.
10. Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013;45(9):970–976.
11. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979–993.
12. Caswell DR, Swanton C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med*. 2017;15(1):133.
13. Farmer H, McCabe N, Lord CJ, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. 2005;434(7035):917–21.
14. Bryant HE, Schultz N, Thomas HD, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*. 2005;434(7035):913–917.
15. Davies H, Glodzik D, Morganella S, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med*. 2017;23(4):517–525.
16. Lord CJ, Tutt AN, Ashworth A. Synthetic lethality and cancer therapy: Lessons learned from the development of PARP inhibitors. *Annu Rev Med*. 2015;66:455–470.
17. Rosenthal R, McGranahan N, Herrero J, et al. DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17:31.
18. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2009.
19. Chan K, Roberts SA, Klimczak LJ, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet*. 2015;47(9):1067–1072.
20. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777–D783.
21. Leonard B, Hart SN, Burns MB, et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res*. 2013;73(24):7222–7231.
22. Fu Y, Ito F, Zhang G, et al. DNA cytosine and methylcytosine deamination by APOBEC3B: Enhancing methylcytosine deamination by engineering APOBEC3B. *Biochem J*. 2015;471(1):25–35.
23. Ito F, Fu Y, Kao SA, et al. Family-wide comparative analysis of cytidine and methylcytidine deamination by eleven human APOBEC proteins. *J Mol Biol*. 2017;429(12):1787–1799.
24. Siriwardena SU, Guruge TA, Bhagwat AS. Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. *J Mol Biol*. 2015;427(19):3042–3055.