KMT2C-deficient tumors have elevated APOBEC mutagenesis and genomic instability in multiple cancers

Xiaoju Hu[†], Antara Biswas[†] and Subhajyoti De [ID]*

Rutgers Cancer Institute of New Jersey, Rutgers the State University of New Jersey, New Brunswick, NJ 08901, USA

ABSTRACT

The histone methyltransferase *KMT2C* is among the most frequently mutated epigenetic modifier genes in cancer and plays an essential role in MRE11-dependent DNA replication fork restart. However, the effects of *KMT2C* deficiency on genomic instability during tumorigenesis are unclear. Analyzing 9,663 tumors from 30 cancer cohorts, we report that *KMT2C* mutant tumors have a significant excess of APOBEC mutational signatures in several cancer types. We show that *KMT2C* deficiency promotes APOBEC expression and deaminase activity, and compromises DNA replication speed and delays fork restart, facilitating APOBEC mutagenesis targeting single stranded DNA near stalled forks. APOBEC-mediated mutations primarily accumulate during early replication and tend to cluster along the genome and also in 3D nuclear domains. Excessive APOBEC mutational signatures in *KMT2C* mutant tumors correlate with elevated genome maintenance defects and signatures of homologous recombination deficiency. We propose that *KMT2C* deficiency is a likely promoter of APOBEC mutagenesis, which fosters further genomic instability during tumor progression in multiple cancer types.

GRAPHICAL ABSTRACT
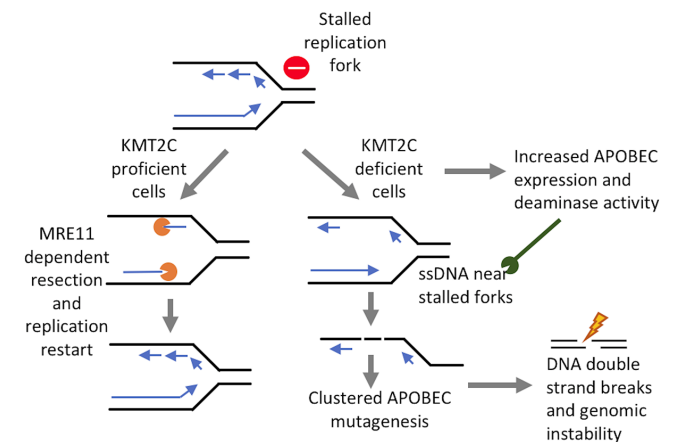
INTRODUCTION

Histone Lysine Methyltransferase (KMT2) family genes are frequently mutated in multiple cancer types ([1]). Histone Lysine Methyltransferase 2C (*KMT2C*), also known as myeloid/lymphoid or mixed-lineage leukemia protein 3 (MLL3) is among the most frequently mutated cancer genes in major cancer types ([2–5]). *KMT2C* is a member of the ASC-2/NCOA6 complex (ASCOM), that is involved in transcriptional coactivation and has histone methylation activity ([6–8]). Although type 2 lysine methyltransferases have substantial homology in sequences and closely related functions, *KMT2C* appears to also have an additional role in DNA replication and genome maintenance ([4,9,10]). Timely and accurate inheritance of genetic information depends on the fidelity of DNA replication, which gets challenged by DNA damage and difficult to replicate regions, leading to replication fork stalling ([11]). It has been shown that TP53 recruits MRE11 to sites of stalled replication forks in a KMT2C-dependent manner, and failed restart of replication fork can impair genome maintenance ([12]). Furthermore, *KMT2C* downregulation leads to extensive changes in the activity of DNA dam-

age response and DNA repair genes, and compromises homologous recombination-mediated double-strand break DNA repair, such that treatment with the PARP1/2 inhibitor olaparib leads to synthetic lethality (9). However, the impact of *KMT2C* deficiency on mutational landscape and genomic instability at a genome-wide scale in tumor genomes is unclear. Here, using analysis of genomic data from multiple cancer cohorts and biochemical assays, we examine how *KMT2C* deficiency compromises replication-dependent genome maintenance processes in cancer.

Mutagenesis by apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) family genes is frequent in many cancer types. In humans, the APOBEC family (13) includes *AID* (activation-induced cytidine deaminase), *APOBEC1*, *APOBEC2*, *APOBEC3A-D*, *APOBEC3F-H* (there is no *APOBEC3E*) and *APOBEC4*, conferring innate immune response against viral infections by deaminating cytidine bases in viral genomes. In tumor cells, however, aberrant DNA hypermutation and promiscuous RNA editing activities of APOBEC enzymes is widespread (14). Among the eleven members of APOBEC superfamily, members of APOBEC3 subfamily *APOBEC3A* and *APOBEC3B* are thought to contribute to the COSMIC mutational signatures SBS2 and SBS13 (15–18). Error prone DNA polymerases replicating across abasic sites, which are generated by uracil removal via base excision repair pathway, likely result in the SBS2 or SBS13 mutations, while SBS2 mutations may also arise directly by DNA replication across uracil bases. SBS2 and SBS13 APOBEC mutational signatures usually co-occur in the same tumor samples (17). What triggers frequent APOBEC mutagenesis in human cancers remains poorly understood. In this study we examine whether *KMT2C* deficiency could be a potential driver of APOBEC mutagenesis and genomic instability in cancer genomes.

MATERIALS AND METHODS

Cancer genomic datasets

We analyzed genomic data including *KMT2C* mutation status in 7574 samples from 17 major cancer types from the TCGA (19) and 2089 samples from 13 cancer types from ICGC (20) cohorts (Supplementary Table S1). Pan-Cancer Analysis of Whole Genomes (PCAWG) datasets were obtained from the ICGC data Portal. Different classes of somatic mutations in *KMT2C* locus were annotated using SnpEff v3.3c (21). The maftools package (22) and UCSC Genome Browser (23) were used for visualization of mutation spectrum and mutation hotspots in the gene locus. The tumor samples that had at least one non-silent (missense, nonsense, frameshift) mutation that potentially alters the *KMT2C* gene product were designated as *KMT2C*-mutant. We also obtained the catalog of somatic structural variations from these ICGC cancer cohorts as well as breast cancer cohort (BRCA-onco) derived from Nones *et al.* (24).

Analysis of point mutation and structural variations

We used deconstructSigs (25) with default parameters to determine the proportion of COSMIC v2 mutational signatures (17) in the tumor genomes. COSMIC mutational signatures SBS2 and SBS13 are attributed to mutagenesis by APOBEC3 family genes (14,16–18), and usually referred to as APOBEC mutational signatures. We also categorically analyzed the somatic mutations at T$\underline{C}$W trinucleotide contexts (W = A or T as per AUPAC nomenclature; including TCA to TTA or TGA, TCT to TTT or TGT), which are attributed to APOBEC3A or APOBEC3B-mediated mutagenesis (26). We used mut.to.sigs.input function from the deconstructSigs (19) package to assign each single-base substitution (SBS) mutation with an appropriate trinucleotide context using hg38 reference genome. Then a published method (27) was used to analyze the mutations at T$\underline{C}$W trinucleotide contexts. Since APOBEC3A and APOBEC3B show relative preferences for YT$\underline{C}$A and RT$\underline{C}$A (Y = pyrimidine, R = purine as per AUPAC nomenclature) contexts, respectively (28), we further estimated the prevalence of somatic mutations at these contexts. For the cancer cohort level analysis, the samples from each of the TCGA or ICGC cohorts were grouped as *KMT2C* mutant and others, and then mutational patterns were compared between the two groups using caseControlSig.R function that implemented in the mutSigTools (29). For the pan-cancer analysis, all samples from the TCGA or ICGC cohorts were pooled and grouped based on the *KMT2C* status as above. For the cohorts with whole genome sequencing data, we analyzed genomic rearrangements as deletions, duplications and inversions, and further grouped the intra-chromosomal rearrangements within 10 kb. We further used the burden of somatic mutations attributed to mutational signature SBS3 (30) as a proxy for homologous recombination (HR) mediated DNA repair defects. Kataegis was also termed as the localized hypermutation marked by at least five consecutive mutations observed within 10 kb (31).

Context-guided analysis of mutational signatures

We annotated the genomic regions based on (i) genomic features—whole genes, exons, repeats, and telomeres, (ii) chromatin features - strong euchromatin, weak euchromatin, intermediate chromatin, weak heterochromatin, strong heterochromatin, (iii) nuclear localization—inter-lamina regions at the nuclear interior, and lamina-associated regions at the nuclear periphery, and compared enrichment of different mutational signatures between *KMT2C* mutant and other tumors at each context using a published approach (29). We retrieved ATAC-seq (32) and somatic mutation data from the same TCGA cancer cohorts, and analyzed the cohorts with at least 10 samples with both types of data. We also obtained data on sites of histone modifications in *KMT2C* deficient bladder HTB9 (H3K4me1 and H3K27ac) (9) and breast MCF7 (H3K27ac and H3K9ac) (6) cell lines, which may serve as proxy for different chromatin contexts. We analyzed somatic mutations from BRCA-EU cohort by overlaying long-range interactions inferred from HiC data for T47D breast cancer cell line (33), in the original paper, libraries were generated using HindIII restriction enzymes, and Illumina HiSeq paired-end reads were aligned using BWA to the human reference genome hg19, duplicate pairs were counted only once, and processed intra- and inter-chromosomal contact matrices were presented at 100kb resolution, as detailed

elsewhere (GEO: GSM1294039). We considered normalized contact score ≥3 between pairs of genomic regions for analyzing patterns of long-range interactions. We compared the observed frequency of the number of long-range interactions between APOBEC-mutations, with that expected by chance, when we randomly shuffle the mutations within respective samples.

Replication timing analysis

Replication timing data for representative human cell lines (wgEncodeEH002247) was obtained from the ENCODE project ([34]). The wavelet-smoothed replication time signal data were divided into deciles (D1, D2, ..., D10) according to their decreasing ordered values, and equal signals were assigned to each decile. Genomic segments were annotated according to their replication timing decile, such that the earliest and latest replicating regions were mapped to first and last deciles, respectively. Somatic mutations in the cancer samples were analyzed by overlaying replication timing data of developmentally related cell lines ([34]). For instance, BRCA-EU mutation data was analyzed with replication timing data from MCF7 breast cancer cell line available from the ENCODE project. Somatic mutations within genomic regions corresponding to different replication decile were identified and mutations enrichment of corresponding COSMIC mutational signatures ([17]) were calculated using published approaches ([35]). Several mutagenic processes (including APOBEC mutagenesis) result in clusters of mutations on the same DNA strand, which is quantified by processivity. Processive groups were determined as maximal stretches of adjacent mutations containing the same reference alleles and were generated by the same signature in each sample. The mutational signature of each substitution was assigned to the highest a posteriori probability, which was identified using the quadratic programming (QP) approach ([36]). Kataegis events were filtered to decrease bias signals, which did not alter the key conclusions.

Regulation of APOBEC expression

We obtained data on transcriptional regulatory regions and transcription factors (TFs) binding sites located around *APOBEC3B* locus in the U2OS cell line from a published report ([37]). Primarily TFs from *NF-κB* gene family (*NFKB1*, *NFKB2*, *RELA*, *RELB*) and from the *AP-1* gene family (*FOS*, *FOSL1*, *FOSL2*, *JUN*, *JUNB*, *JUND*) were reported to co-regulate *APOBEC3B* expression. We also obtained the processed ATAC-seq peak calls ([32]) and *APOBEC3B* expression data for the TCGA cancer cohorts ([19]) with sufficient number of *KMT2C* proficient and deficient samples (TCGA-BRCA and TCGA-COAD; Supplementary Table S1).

Cell culture and knockdown of *KMT2C* expression

HEK293T, RPE1 and MDA-MB-231 cells were cultured in Dulbecco's modified eagle's medium (D6429, Sigma-Aldrich), and SkBr3 in Dulbecco's modified eagle medium/Nutrient mixture F-12 (11320033, GIBCO) supplemented with 10% fetal bovine serum (97068085,

VWR) and 1% Penicillin–Streptomycin solution (97063708, VWR), at 37°C in a humidified incubator with 5% CO_2, and harvested using trypsin-EDTA solution (25200056, GIBCO) or rubber-tipped cell scraper. SMARTpool siRNAs were used for knockdown of *KMT2C* (siKMT2C) (L-007039, Horizon Discovery) and Scrambled (siSCR) was used as a negative control (D-001810, Horizon Discovery). For protein knockdown, cells were transfected with the indicated siRNAs at a final concentration of 50 nM using Lipofectamine (13778075, Thermo Fisher Scientific) following standard protocols for cell lines. After incubating the cells for 48 h, 5 mM HU (400046, Sigma-Aldrich) was added where indicated and the cells were incubated for an additional 1 h.

Measurement of gene expression by real-time quantitative PCR (RT-qPCR)

Total RNA from cells was isolated using Trizol (15596026, Thermo Fisher Scientific) as per manufacturer's instructions. 1μg of total RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (4368814, Thermo Fisher Scientific). One-tenth of this cDNA was used per real-time PCR assay and each sample was assayed in triplicates. Power SYBR green master mix (4367659, Thermo Fisher Scientific) was used for RT-qPCR and samples were quantified on QuantStudio™ 12K Flex Real-Time PCR System (Thermo Fisher Scientific). In brief, CT values using gene specific primers were normalized to the CT value of beta-actin. Fold change in gene expression across samples was calculated by the formula $2^{-\Delta\Delta CT}$. Gene expression values from control experiments were set to unity and data from all treatments were calculated as fold change over 1. All assays were done using RNA from three independent experiments unless otherwise stated. Primers used were - *KMT2C* (F) : TTA CAC ACA GTG CGC TCC TT, *KMT2C* (R) : AGG GTC TGC ACA TGC TAC AA; *MRE11* (F) : CAG TGT TTA GTA TTC ATG GCA ATC ATG, *MRE11* (R) : AAT GTC CAA GGC ACA AAG TGC; *APOBEC3B* (F) : GAC CCT TTG GTC CTT CGA C, *APOBEC3B* (R) : GCA CAG CCC CAG GAG AAG, *ACTIN* (F) : GAG CAC AGA GCC TCG CCT TT, *ACTIN* (R) : TCA TCA TCC ATG GTG AGC TGG. The experiment was performed at least thrice and data analyzed using two-tailed t-test. Values in the plot are means ± SEM.

DNA fiber assay

DNA fiber assay on Scrambled siRNA and *KMT2C* siRNA treated HEK293T cells was performed. Experiments performed with or without hydroxyurea (HU) treatment under conditions indicated in the schematic. Briefly, asynchronous HEK293T cells were labelled with 25 μM CldU for 20 min, washed with PBS, treated or not with 2 mM HU for 2 h, washed again in warm PBS and exposed to 250 μM IdU for 20 min before collection. Cells were then lysed and DNA fibers stretched onto glass slides. The fibers were denatured with 2.5 M HCl for 1 h, washed with PBS and blocked with 2% BSA in phosphate buffered saline Tween-20 for 30 min. The newly replicated CldU and IdU tracts were incubated with anti-BrdU antibodies recognizing CldU and IdU respectively (ab6326, Abcam; 347580, BD Biosciences) for 2 h

at room temperature in a humidified chamber. Fibers were then washed with PBS and incubated with secondary antibody against CldU (A-11007, Thermo Fisher Scientific) and IdU (A-11029, Thermo Fisher Scientific) for 1 h at room temperature in a humidified chamber. After washing in PBS, the slides were air-dried and mounted using VEC-TASHIELD Mounting Medium (H-1200, Vector Laboratories). Images were taken at 40× with oil immersion objective using a Nikon Eclipse TE2000-U microscope (Nikon). Images of the same group were captured with identical exposure time using NIS-Elements software and analyzed using ImageJ software (National Institutes of Health), with custom-made modifications. Only replication signals from high-quality ssDNA and not DNA fiber bundles, were selected for analyses. Fibers were marked for evaluation by 'blind' measurers (not knowing which samples were being measured), fiber length was measured using the ImageJ software followed by compilation into a worksheet. The length of minimum 150 tracts from each condition were measured. IdU/CldU values were analyzed using Mann–Whitney U-test with median and interquartile range shown. Fork stalling/ new origins firing was estimated by analyzing red tracts only for stalled replication forks and green only and green–red–green for newly fired origins, using two-tailed t-test with >250 fibers measured for each condition. The experiment was performed at least twice using independent biological isolations of DNA fibers.

Quantification of 53BP1 and RAD52 double immunostained foci

HEK293T and RPE1 cells were washed with PBS following treatments, fixed with 3% paraformaldehyde and permeabilized with 0.5% Triton X-100. Cells were blocked with 1% BSA for 1 h, then incubated sequentially with primary antibodies (*53BP1*- ab21083, Abcam; *RAD52*- MA5-31888, Thermo Fisher Scientific) and secondary antibodies (A11008, Thermo Fisher Scientific; A32742, Thermo Fisher Scientific respectively) for 1 h each at 37°C, with three PBS washes in between. Coverslips were mounted onto glass slides with VECTASHIELD Mounting Medium with DAPI (H-1200, Vector Labs). Images were captured at 20× objective using EVOS M500 microscope (Thermo Fisher Scientific). Images of the same group were captured with identical exposure time. Images were processed using ImageJ software, and cells were scored as displaying either diffused or punctuated staining. Cells with punctuated staining were further analyzed for calculation of the number of foci. The experiment was performed at least thrice and data analyzed using two-tailed t-test. Values in the plot are means ± SEM.

Preparation of protein lysate and cytidine deamination assays

Cells were lysed in HNET buffer (25 mM HEPES-NaOH, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 1 mM DTT, 1× protease inhibitor cocktail (P2714, Sigma-Aldrich). Protein lysates were quantified by BCA method (PI23227, Thermo Fisher Scientific) and equal amounts were used for subsequent assays. Deaminase assay was performed as described by (38). Briefly, equal amounts of protein lysate (20

µg) were incubated with 5 pmol probe in 1× deaminase buffer (10 mM Tris–HCl pH 8.0, 50 mM NaCl, 1 mM DTT) at 37°C for 2 h followed by incubation with 0.75 U UDG (EN0361, Thermo Fisher Scientific) in 1× UDG buffer at 37°C for 45 min. The reaction mixture was treated with 0.15 M NaOH at 37°C for 20 min prior to its termination by heating at 95°C for 3 min with equal volume of 2× RNA loading dye and prompt chilling on ice. Samples were resolved by electrophoresis at 150 V for 3 h at room temperature in Novex TBE–urea gels (EC6885BOX, Thermo Fisher Scientific) in 1× TBE buffer and cleaved and uncleaved products were imaged on ChemiDoc (Bio-Rad) using Alexa Fluor 488 module. Percent substrate cleaved was determined by quantification of the intensity of the substrate and cytidine deamination cleavage product using ImageJ (NIH). Fluorescently labeled DNA probes used for the assay were: test probe: 5Alex488N/AT AAT AAT AAT AAT AAT AAT AAT ATC CAT AAT AAT AAT AAT AAT AAT A; positive probe: 5Alex488N/AT AAT AAT AAT AAT AAT AAT AAT ATU UAT AAT AAT AAT AAT AAT AAT A. The experiment was performed at least thrice, and data analyzed using two-tailed t-test. Values in the plot are means ± SEM.

Statistical analyses

Statistical analyses of genomic data were performed using R version 3.6.1 and Prism 8 (GraphPad). Asterisks indicate statistical significance wherein **$P < 0.01$; *$P < 0.05$; ns, non-significant. For DNA fiber assay we used more stringent test with additional cutoffs with ****$P < 0.0001$, ***$P < 0.001$. Statistical tests and corresponding P-values are listed for respective analyses. P-values were adjusted for multiple testing using Fisher method from the metap (v1.8) R package (Dewey, 2022).

RESULTS

KMT2C mutations frequently occur in multiple cancer types

KMT2C is a relatively large gene with the primary transcript coding a protein of 4911 amino acids long and spanning 59 exons. We analyzed mutation status of histone methyltransferases in 7574 samples from 17 major cancer types from the TCGA (19) and 2089 samples from 13 ICGC cohorts (20) (Figure 1A and Supplementary Table S1; see Materials and Methods). Proportion of samples with *KMT2C* mutations varied between 3.23% (11/364) in liver cancer (TCGA-LIHC) and 40% in colorectal cancer (12/30, ICGC-COCA-CN) and lung squamous cell carcinoma (12/30, ICGC-LUSC-KR). A majority of these mutations are missense, nonsense and frameshift mutations (Figure 1B) with potentially loss-of-function consequences, as reported elsewhere (6,9). In general, somatic mutations were distributed throughout the *KMT2C* gene locus in all cancer types, and there were very few recurrent mutations at the base-pair level. Nonetheless, in some cancer types, mutations occurred at a higher frequency around exon 36–38, and exon 43–52, closer to the 3′ end of *KMT2C* gene (Figure 1B); exons 36–38 and 43–52 both overlap with the extended PHD zinc finger domain annotation, while 3′ side of exon 52 also overlaps with the FY-rich domain.
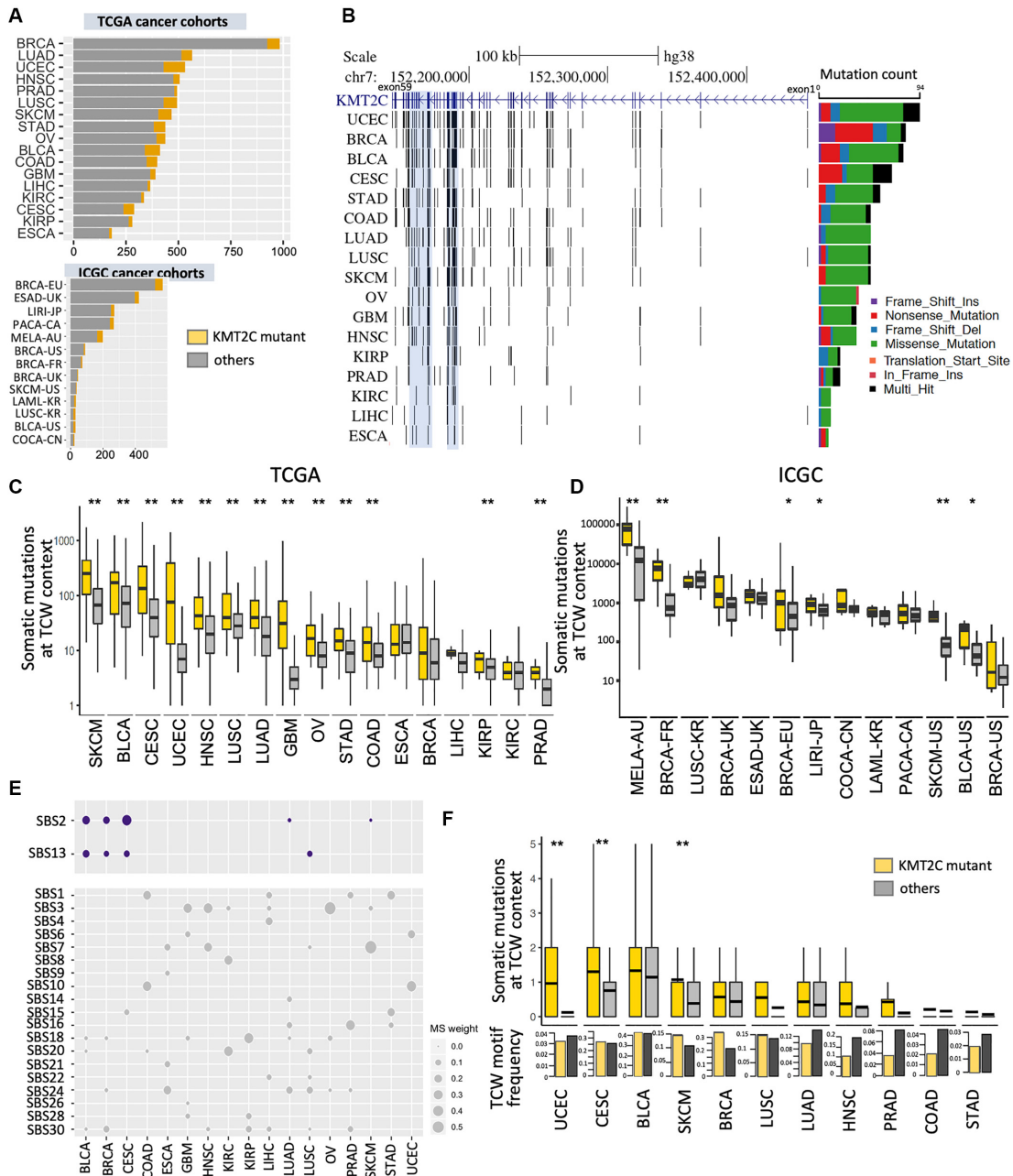
Figure 1. *KMT2C* mutant tumors have a significant excess of APOBEC mutational signatures in multiple cancer types. (**A**) The number of tumor samples in 17 TCGA and 13 ICGC cancer cohorts. The samples with *KMT2C* mutations are indicated in yellow. (**B**) Genome graph shows distribution of non-silent, coding mutations in *KMT2C* locus in the TCGA cancer cohorts. The shadowed blue regions indicate two mutational hotspots spanning exons 36–38 and 43–52. The bar graph on the right depicts total frequency of different classes of mutations in respective cohorts, as indicated by the keys. (C, D) Boxplot showing the somatic mutation count at TCW contexts (including TCA to TTA or TGA, and TCT to TTT or TGT mutations) in *KMT2C* mutant (yellow) and other samples (grey) for (**C**) TCGA and (**D**) ICGC cohorts. *P*-value was calculated by Mann–Whitney *U*-test, *P-value < 0.05, **P-value < 0.01. (**E**) Bubble plot showing the weights of different COSMIC v2 mutational signatures in non-silent *KMT2C* mutations in respective cancer cohorts. (**F**) Boxplot showing the somatic mutation count at TCW contexts within ATAC-seq peak regions in *KMT2C* mutant and other samples in the TCGA cohorts. The bar showing the mean value, and the top and bottom of the boxes indicated first and third quartiles respectively. Barplots in the panel below show the frequency of TCW motifs in these ATAC-seq regions in the two groups in these cancer cohorts. *P*-value was calculated by Mann–Whitney U-test, *P-value < 0.05, **P-value < 0.01. For description of the cohorts, see (A) and Supplementary Table S1. BRCA: breast invasive carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma; HNSC: head and neck squamous cell carcinoma; UCEC: uterine corpus endometrial carcinoma; LUSC: lung squamous cell carcinoma; SKCM: skin cutaneous melanoma; OV: ovarian serous cystadenocarcinoma; STAD: stomach adenocarcinoma; GBM: glioblastoma multiforme; LIHC: liver hepatocellular carcinoma; COAD: colon adenocarcinoma; BLCA: bladder urothelial carcinoma; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; CESC: cervical and endocervical cancers; ESCA: esophageal carcinoma; BRCA-EU: breast ER+ and HER2– cancer – EU/UK; ESAD-UK: esophageal adenocarcinoma – UK; LIRI-JP: liver cancer – RIKEN, JP; PACA-CA: pancreatic cancer – CA; LAML-KR: acute myeloid leukemia – KR; COCA-CN: colorectal cancer – CN; LUSC-KR: lung cancer – squamous cell carcinoma – KR; MELA-AU: skin cancer – AU; BRCA-FR: breast cancer – FR; BRCA-UK: breast triple negative/lobular cancer – UK; BRCA-US: breast cancer – TCGA, US; SKCM-US: skin cutaneous melanoma – TCGA, US; BLCA-US: bladder urothelial cancer – TCGA, US.

KMT2C mutant tumors have an excess of APOBEC mutational signatures

Given the dual role of *KMT2C* in both epigenetic regulation and replication, we first examined whether somatic mutations in *KMT2C* mutant tumors have an excess of signatures of certain mutational processes that could in turn provide mechanistic insights into the effects of loss of *KMT2C* on genome maintenance. In each cancer cohort, we identified the samples with non-silent mutations in *KMT2C* (see Methods) and designated them as *KMT2C*-mutant tumors. We then examined the single base substitution patterns characteristic of different mutational processes in somatic mutations in the tumor samples in the cancer cohorts.

Mutagenesis by APOBEC3 gene family is a common source of somatic mutations that arise during replication stress in cancer genomes and preferentially occurs at TCW trinucleotide context (W = A or T; specifically, TCA to TTA or TGA, TCT to TTT or TGT) [26]. Comparing frequencies of these somatic mutation-classes between the *KMT2C* mutant and other samples in different TCGA [19,20] and ICGC cancer cohorts, we observed that the former had a significant excess of corresponding mutational signatures (referred to as the APOBEC mutational signatures here on) in most of the cohorts (Figure 1C and D). This trend was especially prominent in the cohorts with high baseline APOBEC mutational signatures such as bladder cancer (TCGA-BLCA and ICGC-BLCA-US), cervical cancer (TCGA-CESC), head and neck cancer (TCGA-HNSC), lung cancer (TCGA-LUSC and TCGA-LUAD), and breast cancer (ICGC-BRCA-FR and ICGC-BRCA-EU). Next, using an unbiased approach we examined all COSMIC (v2) mutational signatures [17] in all the samples from the TCGA and ICGC projects. Among the mutational signatures prominently present (>5% mutational weight) in both datasets, APOBEC mutational signatures SBS2 and SBS13 showed systematic increase in mutational weight in *KMT2C* mutant samples (Supplementary Figure S1A). Subsequently, in a cohort-wise analysis we calculated the proportional weights of each mutational signature in each cancer cohort, and observed that APOBEC mutational signature (combined mutational weight of SBS2 and SBS13) was substantially higher in *KMT2C* mutant tumors, especially in the cohorts that had reasonable footprints of APOBEC mutagenesis, including cervical cancer (TCGA-CESC), bladder cancer (TCGA-BLCA), head and neck cancer (TCGA-HNSC), lung cancer (TCGA-LUSC) and breast cancer (ICGC-BRCA-EU) (Supplementary Figure S1B and S1C). Taken together, enrichment for APOBEC mutational signature was consistently observed in both the TCW context- and mutational signature-guided analyses. Besides the APOBEC mutational signatures, several other mutational signatures were also detected in the pan-cancer and cohort-wise analyses, but their effects were moderate and/or cohort-specific; in the pan-cancer analysis SBS6, SBS7, SBS10 and SBS12 showed over-representation in *KMT2C* mutant tumors (Supplementary Figure S1A), while in the cohort-wise analysis SBS20, which is attributed to defective DNA mismatch repair (MMR) was also enriched in the *KMT2C* mutant samples (Supplementary Figure S1C). *KMT2C* deficiency contributes to replication stress [39], which promote microsatellite instability (MSI) and hypermutation.

APOBEC family enzymes APOBEC3A and APOBEC3B are potent mutator deaminases but APOBEC3A and APOBEC3B have differences in the tetranucleotide context preferences for deamination: APOBEC3A prefers YTCA, while APOBEC3B prefers RTCA (Y = pyrimidine, R = purine) motifs [28]. Furthermore, APOBEC3B has a much higher expression level than APOBEC3A, although the later may have higher mutagenic activity in cancer genomes [40]. In our analysis, APOBEC-mediated mutations at both the RTCA and YTCA contexts showed increased incidence in *KMT2C* mutant tumors, suggesting that the associations are probably not specific to APOBEC3A or APOBEC3B-mediated mutagenesis (Supplementary Figure S1D).

To examine whether APOBEC-mediated or other mutagenic processes led to *KMT2C* mutations in the cancer genomes, we determined the burden of different mutational signatures on the non-silent mutations in *KMT2C* in each cohort. There were no single, universal mutagenic processes driving *KMT2C* mutations across different cancer types, albeit SBS1, SBS3, SBS7 and SBS10 were observed in multiple cancer types. SBS2 and SBS13—the signatures of APOBEC mutagenesis were observed in some TCGA cancer cohorts (e.g. BLCA, BRCA, CESC, SKCM, LUSC and LUAD; Figure 1E), which had high prevalence of APOBEC-mediated mutations. In some cases, these mutations are causal events, while in other once a loss-of-function mutation is acquired, the non-functional allelic copy might continue to accumulate additional mutations under the prevailing mutagenic processes in the affected samples, but we could not ascertain the order of mutational events.

Patterns of mutational signatures in epigenomic contexts

Given the role of KMT2C as an epigenetic modifier, we first assessed whether APOBEC mutagenesis predominantly occurs within the altered chromatin domains in the *KMT2C* mutant tumors. We jointly analyzed somatic mutations and ATAC-seq data from the same samples from the TCGA cancer cohorts, and observed that *KMT2C* mutant tumors had higher burden of APOBEC mediated mutations in ATAC-seq peak regions, when compared to that in the ATAC-seq peak regions in other samples in the same cohorts. This was also applicable to a majority of the cohorts with high prevalence of APOBEC-mediated mutations such as breast cancer (TCGA-BRCA), bladder cancer (TCGA-BLCA), cervical cancer (TCGA-CESC), lung squamous cell carcinoma (TCGA-LUSC), which also had high frequency of TCW motifs in the ATACseq peak regions (Figure 1F). In a complementary analysis, we overlaid open chromatin data from *KMT2C* null and proficient HTB9 bladder and MCF7 breast cancer cell lines with somatic mutations from TCGA tumors, and found that altered chromatin regions gained in *KMT2C* null conditions had excess of somatic mutations at TCW trinucleotide context in multiple cancer types (Supplementary Figure S2A). However, a vast majority of the APOBEC mediated mutations occurred outside the peak regions (Supplementary Figure S2A), and in both analyses the increased

burden of APOBEC-mediated mutations was not restricted only to the altered chromatin domains but observed genome-wide - led us to investigate mutagenesis in different genomic, epigenomic, nuclear, and replication contexts systematically.

Therefore, next we broadened the analysis, segmenting the genome into a number of genomic segments that are relevant for key cellular processes such as transcription, regulation of gene expression and DNA replication, and calculating enrichment for the mutational signatures in these genomic segments in the *KMT2C* mutant tumors (Figure 2A and Supplementary Figure S2B). The repeat and gene regions showed enrichment for APOBEC-mediated mutations (SBS2 and SBS13), but within gene regions, exons did not show such preferences. There were other, signature-specific differences as well. For instance, SBS13, which is known to occur more often in early replicating regions (35) showed higher burden in euchromatin in *KMT2C* mutant tumors in BRCA-EU cohort. Interestingly, SBS2, which preferentially occurs in heterochromatic, late replicating regions, also showed relative enrichment in all chromatin contexts in *KMT2C* mutant tumors in this cancer type (Figure 2A). Some other differences were cancer-type specific (Supplementary Figure S2B). Taken together, in *KMT2C* mutant tumors, APOBEC mediated mutations occurred not just in altered chromatin domains but genome-wide and showed preferences for broader (epi)genomic contexts. These observations, as well as the role of KMT2C in MRE11-dependent fork start during DNA replication motivated us to examine the mutational signatures in light of DNA replication-associated defects in *KMT2C*-deficient cells.

KMT2C loss causes replication stress

KMT2C plays a key role in recruitment of MRE11 to restart stalled DNA replication fork (10) facilitating replication and genome maintenance (Figure 2B). Stalled replication forks and unscheduled origin firings produce ssDNA, which can be potential targets of APOBEC mutagenesis (40). Thus, we examined whether *KMT2C* deficiency affects DNA replication dynamics. We first evaluated replication fork dynamics in HEK293T cells using DNA fiber assay, which allows visualization and analysis of replication at DNA-strands level. We knocked down the expression level of *KMT2C* (siKMT2C) in these cell lines, labelled them with thymidine analogs CldU (red tracks) and IdU (green tracks) consecutively and measured frequency and lengths of fiber tracks. Cells transfected with Scrambled siRNA (siSCR) served as a control. Under unperturbed growth conditions, *KMT2C* deficiency caused DNA replication stress in HEK293T cells, as indicated by shorter nascent tract length (Supplementary Figure S2C) and decreased replication fork speed (Figure 2C). Earlier studies have reported implication of KMT2C-mediated chromatin opening in MRE11 nuclease recruitment to stalled forks. We reasoned that apart from contribution to downstream replication fork maintenance KMT2C also functions upstream to prevent DNA replication stress. To further dissect the role of KMT2C in DNA replication, we treated cells with hydroxyurea (HU), a genotoxic agent which de-

pletes nucleotide pools, as a prototype of replication stress. We prelabeled cells with CldU prior to HU treatment, followed by IdU and determined the length and frequency of ongoing replication forks stalled upon hydroxyurea treatment, as well as licencing of new/dormant origins, if any. We observed an increase in frequency of stalled forks in *KMT2C* silenced cells challenged by HU (Figure 2D). We also observed an increase in aberrant origin firing, which could be compensation for the fork stalling, further impeding progression of replication process (Figure 2E). For the stalled forks that could resume DNA synthesis, *KMT2C* knocked down cells exhibited defective replication resumption compared to wild-type cells as assessed by the ratio of IdU to CldU tract lengths (Figure 2F, G). These observations are consistent with reports that recruitment of MRE11 on stalled replication forks to initiate DNA repair is dependent upon KMT2C mediated chromatin opening (10,41) This suggests that *KMT2C* deficiency invokes replication stress and disables the cells for efficient replication restart and fork progression upon genotoxic stress. We further observed *MRE11* and *KMT2C* expression significantly associated with mutational signatures of APOBEC mutagenesis (DBS11) (17) and replication stress (SBS40) (29)in the PCAWG pan cancer dataset (Supplementary Figure S3A), which led us to examine APOBEC expression and activity in *KMT2C* deficient conditions.

Replication stress triggers APOBEC activity

Among APOBEC family genes, *APOBEC3A* or *APOBEC3G* have almost undetectable base-line expression whereas *APOBEC3B* remained detectable in most human cell lines (28). We performed RT–qPCR analysis in *KMT2C* knockdown HEK293T, RPE1 and MDA-MB-231 cells with low, moderate and high *APOBEC3B* expression levels, respectively (Supplementary Figure S3B). Cells transfected with non-targeting scrambled siRNA (siSCR) served as a control. Figure 3A shows the levels of *KMT2C*, *MRE11* and *APOBEC3B* in HEK293T cells, with small but non-significant reduction in *MRE11* and unchanged *APOBEC3B* expression levels in *KMT2C* deficient cells as compared to control cells. However, when *KMT2C* depleted cells were challenged with DNA replication inhibitor HU, significant reduction in *MRE11* expression and induction in *APOBEC3B* expression compared to that in the control cells were observed. Surprisingly, in RPE1 and MDA-MB-231 cells, *KMT2C* silencing resulted in *MRE11* depletion and *APOBEC3B* induction, without additional replication stress. Overall, we observed a trend of elevating *APOBEC3B* levels upon *KMT2C* loss in all three cell lines, HEK293T, RPE1 and MDA-MB-231 cells with different baseline levels of *APOBEC3B* expression. Integrating transcription factor binding sites and ATAC-seq data for the TCGA cancer cohorts we observed that (i) *KMT2C* mutant tumors have relatively more open chromatin near well-characterized regulatory regions (37) around *APOBEC3B* promoter, (ii) these regulatory regions overlap with the binding sites of the NF-κB and AP-1 family of transcription factors, which are known regulators of *APOBEC3B* and (iii) such chromatin-level alterations are also associated with increased *APOBEC3B* mRNA
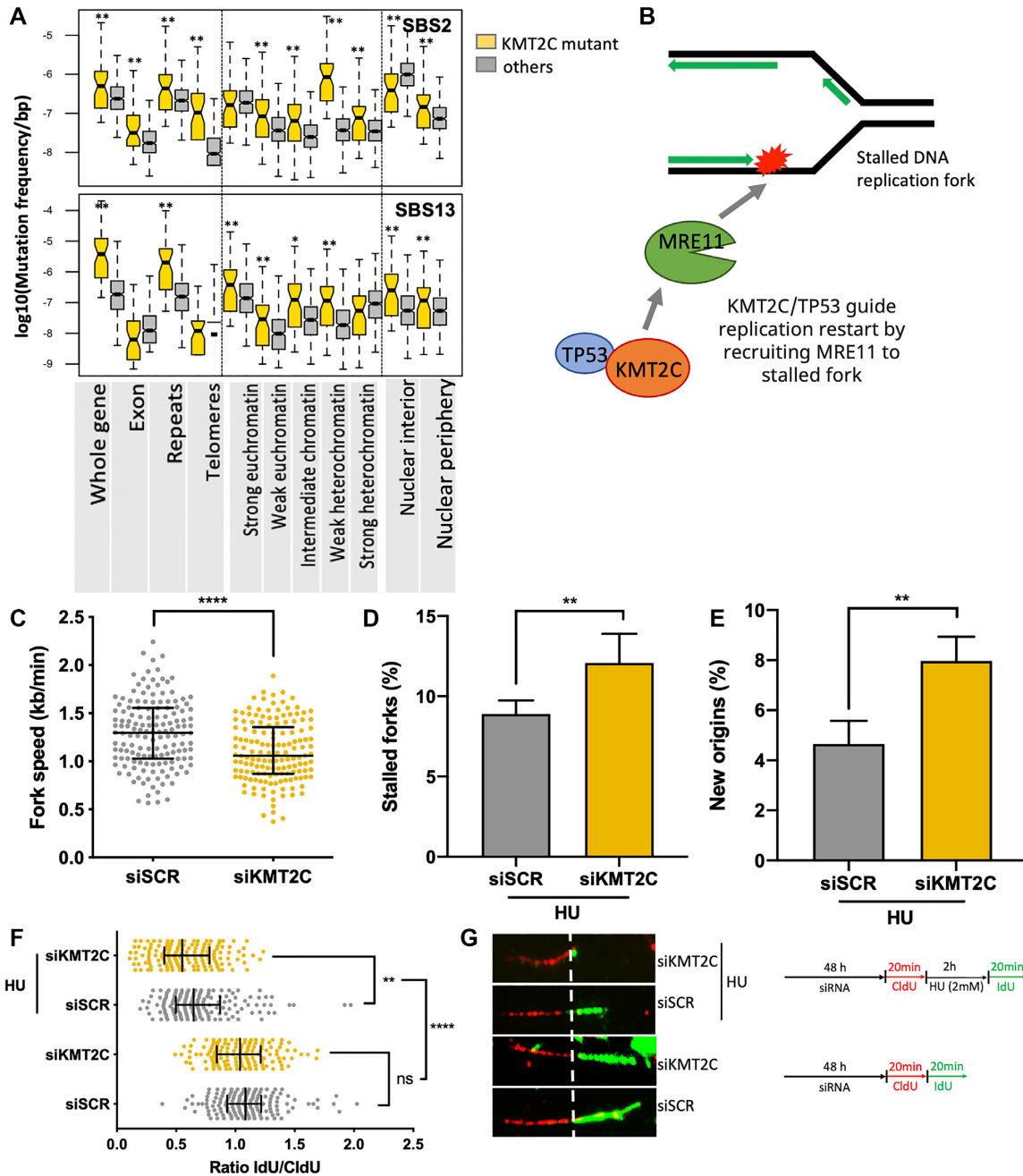
Figure 2. *KMT2C* deficiency is associated with defects in replication. (**A**) Mutational burden of SBS2 and SBS13 mutational signatures in different genomic, epigenomic, and nuclear localization contexts in *KMT2C* mutant (yellow) and other (grey) samples in the breast cancer cohort (BRCA-EU). Similar results for other cohorts are shown in Supplementary Figure S2B. *P*-value was calculated by Mann–Whitney U-test, *P-value < 0.05, **P-value < 0.01. (**B**) Schematic representation showing the paradigm that KMT2C/TP53 guide replication restart by recruiting MRE11 to stalled DNA replication fork. (**C-G**) DNA fiber assay on control (siSCR) and *KMT2C* knockdown (siKMT2C) HEK293T cells showing fork speed with line and bars representing the median and interquartile range (**C**), % of stalled forks with error bars representing SEM (**D**), % of newly initiated origins with error bars representing SEM (**E**), ratio of IdU/CldU with line and bars representing the median and interquartile range (**F**) and representative images of DNA fibers with a schematic representation of the DNA fiber assay conditions (**G**). Experiments performed with or without hydroxyurea (HU) treatment are indicated in the schematic. *P*-value was calculated using Mann–Whitney *U*-test wherein *P-value < 0.05, **P-value < 0.01, ***P-value < 0.001, ****P-value < 0.0001.
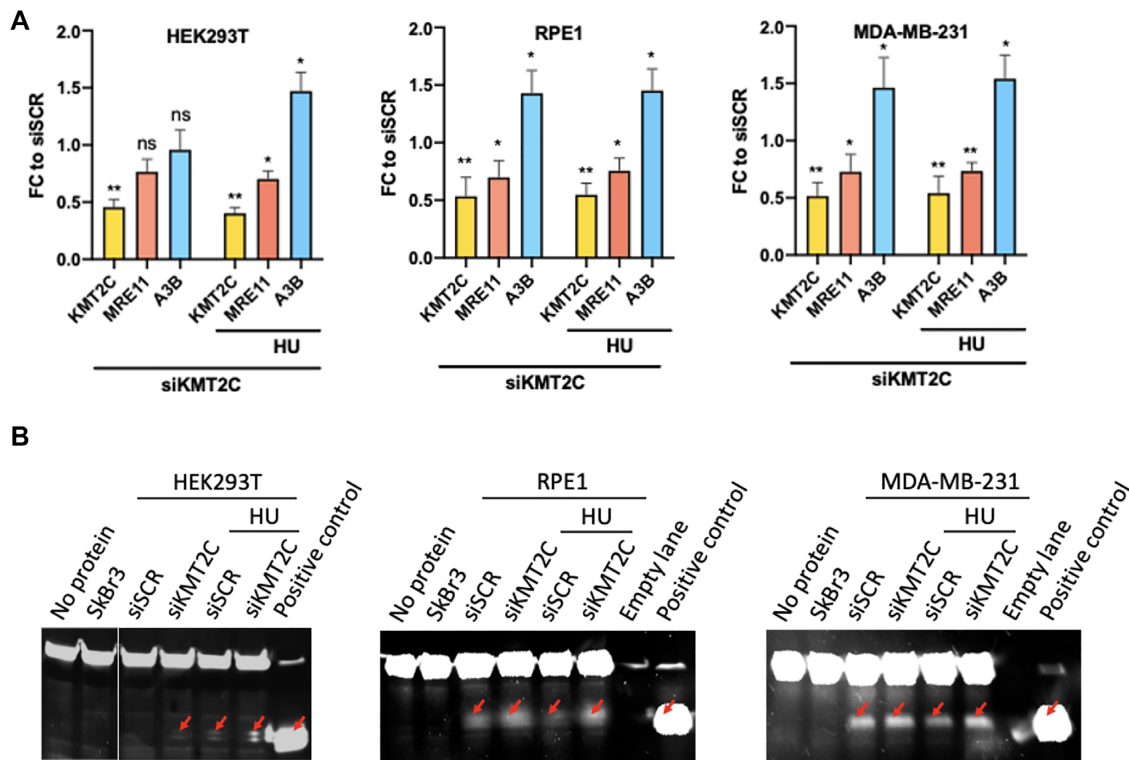
Figure 3. Replication stress due to *KMT2C* deficiency promotes APOBEC expression and deaminase activity (**A**) Fold change (FC) in expression of indicated genes in HEK293T, RPE1 and MDA-MB-231 cells in response to *KMT2C* loss (siKMT2C) and/or HU treatment as compared to control cells (siSCR) with error bars representing SEM. Asterisks indicate statistical significance. *P*-value was calculated using *t*-test wherein *P-value < 0.05, **P-value < 0.01. (**B**) Protein lysates were assessed in deaminase assay. Test probe without protein lysate and lysate from SkBr3 cell line were used as negative controls. Probe containing uracil rather than cytosine and not incubated protein lysate, showing the mobility of the cleaved product was used as positive control (last lane). Red arrows indicate cleaved probe.

expression in the same samples, as observed in multiple cancer cohorts (Supplementary Figure S3D).

Although the base-line expression of *APOBEC3B* in human cells is generally low, and elevated expression is associated with increased APOBEC cytidine deaminase activity ([14]), we further examined whether *KMT2C* loss triggers elevated levels of APOBEC activity as well. We examined the deamination activity present in *KMT2C* deficient cell lysates using an oligonucleotide-based cytidine deamination assay where C to U conversion in labeled oligonucleotide allows fluorescence detection following cleavage by uracil-DNA glycosylase activity. In lysate from SkBr3 cell line, no deaminase activity was detected, an *APOBEC3B*-null control for this assay (Figure 3B). Scrambled siRNA transfected HEK293T cells showed no detectable deaminase activity, owing to low baseline APOBEC expression, but in corresponding *KMT2C* deficient cells deaminase activity was elevated as evidenced by cleaved probe, deaminated by increased APOBEC expression in cell extracts, which further increased several folds when subjected to HU treatment. Similar to RT-qPCR results, in RPE1 and MDA-MB-231 cells, upon *KMT2C* depletion APOBEC deaminase activity was elevated, even without genotoxic insults. Quantified deamination results indicating the percentage of the probe cleaved with cell lysate under indicated conditions is shown in bar graphs (Supplementary Figure S3C). This evidence is consistent with a model that *KMT2C* loss and

replication stress increase APOBEC expression and activity, contributing to APOBEC mutagenesis.

APOBEC mutational signatures in replication context

APOBEC mutagenesis during replication results in the mutational signatures SBS2 and SBS13, which preferentially occur in late and early replication contexts, respectively (Figure 4A). Analyzing somatic mutations from the breast cancer cohort (BRCA-EU) with repliseq data from MCF7 cell line, we observed that in the tumors with no *KMT2C* mutations indeed SBS2 mutation density increased from early to late replication, while SBS13 had comparable mutation density through all replication timing domains. In contrast, *KMT2C* mutant tumors had an excess of both SBS2 and SBS13 signatures in early replication contexts, although the proportional increase was higher for SBS2 (Figure 4B). APOBEC-mediated mutations tend to occur sporadically and in bursts, and in case of replication stress there are clusters of APOBEC-associated mutations which have extended processivity, i.e. groups of similar substitutions attributed to the same mutational signature on the same replication strand ([35],[40]). We calculated processivity of the APOBEC mutational signatures SBS2 and SBS13, and compared that between *KMT2C* proficient and deficient tumors. The processivity groups where consecutive mutations are within 10 kb were considered, since more distant ones may not nec-
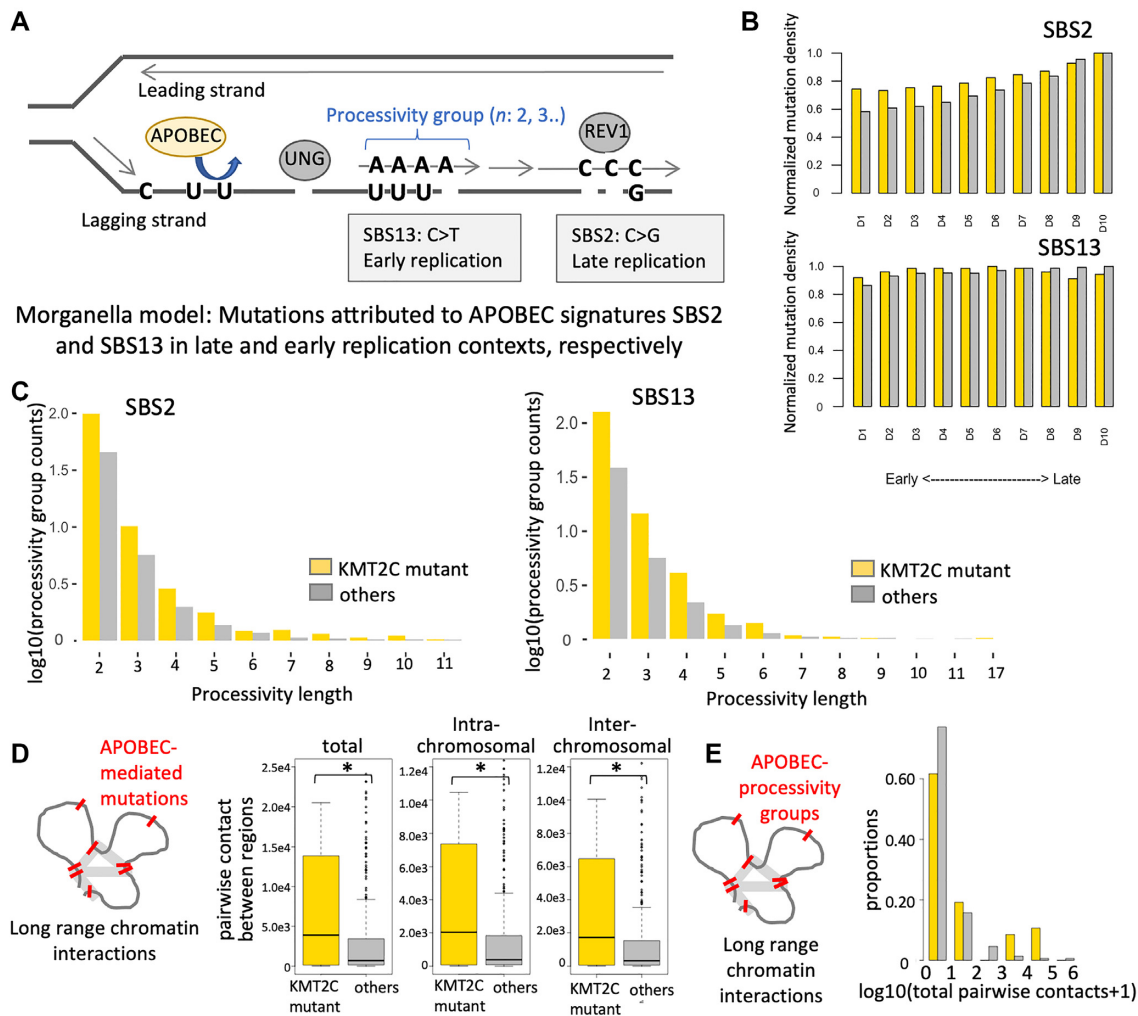
Figure 4. APOBEC mutational signatures in replication contexts (**A**) Schematic representation showing that mutations attributed to APOBEC signatures SBS2 and SBS13 preferentially occur in late and early replication contexts, respectively. Such mutations may occur in clusters on the same strand, as indicated by the length of processivity groups. (**B**) Boxplots showing normalized mutation densities of APOBEC mutational signatures SBS2 and SBS13 in different DNA replication contexts in *KMT2C* mutant (yellow, $n = 47$) and other (grey, $n = 522$) samples in the breast cancer cohort (see Methods for details). The frequencies of acquired mutations in respective replication contexts were normalized by the total length of such regions ('A's,'C's,'G's,'T's, and excluded 'N's). D1–D10 were ordered from earliest to latest replicating regions. *P*-value was calculated by fisher's exact test for each replicating region, and no significance found between *KMT2C* mutant and other samples. (**C**) Frequencies of APOBEC processivity group events corresponding to different processive group lengths in *KMT2C* mutant (yellow) and other (grey) samples in the breast cancer (BRCA-EU) cohort. The frequencies of acquired mutations in each group were normalized by the number of samples in corresponding groups, and nominalized counts were $\log_{10}$ transformed. *P*-value was calculated by fisher's exact test for each replicating region, and no significance found between *KMT2C* mutant and other samples. (**D**) Boxplots showing the total, intra- and inter-chromosomal pairwise contacts between genomic regions with APOBEC-mediated mutations in *KMT2C* mutant and other tumors in the BRCA-EU cohort. *P-value < 0.05. (**E**) Histogram showing the proportions of tumors with different frequencies of pairwise contacts between genomic regions with APOBEC-mediated mutation processive groups in *KMT2C* mutant and other tumors in the BRCA-EU cohort. *P*-value = 9.53e–6; Mann–Whitney *U*-test.

essarily occur in the same replication stress-related event. SBS13 was generally associated longer processivity group sizes than SBS2, which is consistent with previous studies (35). Compared to other tumors, *KMT2C* mutant tumors had higher frequencies across all processivity groups for SBS2 and SBS13, and had preference for longer processive group events (e.g. $n > 6$), although the numbers were low, suggesting preference for clustered APOBEC mutagenesis (Figure 4C). Clustered mutagenesis can be associated with Kataegis, but we found no significant enrichment for such events in *KMT2C* mutant tumors.

Clusters of APOBEC-mediated mutations localized in 3D nuclear contexts

Since replication of multiple DNA segments concurrently progress within 3D nuclear organizations called replication factories (42), we examined whether there is APOBEC-mediated mutations cluster within 3D nuclear contexts, even when those might be distal on the linear DNA, especially in the *KMT2C* mutant tumors. We overlaid HiC-based long-range chromatin interaction data (33) with somatic mutations attributed to APOBEC-mediated mutagenesis for the breast cancer cohort (Methods) and observed

that *KMT2C* mutant tumors had significantly higher number of pairwise long-range interactions between genomic regions harboring APOBEC-associated mutations compared to other samples in the cohort, and the overall trends were similar for both intra- and inter-chromosomal interactions (Figure 4D; total, P-value = 1.35e–2; intra-chromosomal, P-value = 1.40e–2; inter-chromosomal, P-value = 1.18e–2). We then repeated the analyses only for a subset of the genomic regions that harbor the APOBEC processivity groups ($n \geq 2$) as above, and found that 3D clusters of mutations, marked by excessive number of between-region contacts was significantly more common in *KMT2C* mutant tumors (Figure 4E; P-value = 9.5e–06). This was not due to long processivity groups within the same DNA segments; long-range interaction had a resolution of 100 kb, each pair of regions was counted only once per sample, and we observed similar results for within- and between-chromosome interactions. Our results suggest that *KMT2C* mutant tumors accumulate frequent clusters of APOBEC-mediated mutations within 3D nuclear contexts. It is plausible that many such mutation clusters may arise concurrently during replication within the same replication factory, but it is challenging to establish that conclusively.

Replication stress skews DNA repair away from canonical HR pathway

Replication stress and delay in fork restart in *KMT2C* deficient conditions can promote DNA double strand breaks (43,44), and it was shown that *KMT2C* loss contributes to defects in homologous recombination (HR), causing sensitivity to PARP inhibitors in cell line models (9). We quantified the extent of genomic DNA damage in *KMT2C* depleted HEK293T and RPE1 cells by immunostaining 53BP1 which identifies DNA breaks and directs repair pathway away from homologous recombination. We observed that DNA damage repair in *KMT2C* deficient cells, as evidenced by the number of 53BP1 foci positive cells, number of cells with more than five foci and number of foci per cell, was higher as compared to those measured in the control cells (Figure 5A, B and Supplementary Figure S4A). Similar to 53BP1, RAD52 negatively regulates resection of double strand break ends and plays an important role in backup DNA repair pathways in HR-deficient cells (45–47). Therefore, we also immunostained cells for RAD52 and observed an increase in RAD52 foci in *KMT2C* silenced cells along with partial colocalization of RAD52 and 53BP1 foci, although RAD52 signals were not as prominent as 53BP1 signals (Figure 5A, B, Supplementary Figure S4A and B). Using a complementary approach, we analyzed tumor genomic data from the ICGC cancer cohorts (Supplementary Table S1, (20)) to compare patterns of genomic rearrangements in *KMT2C* mutant tumors. We found that small deletions (<10 kb) were consistently more frequent in *KMT2C* mutant tumors compared to other tumors in all the breast cancer cohorts analyzed (Supplementary Figure S4C). We observed consistent results for other cancer cohorts as well (Supplementary Figure S4C). Since length of homology in structural variation (SV) junction regions was not always available to infer the DSB repair mechanisms for the SVs, we used the single base substitution signature SBS3 as a proxy for HR deficiency-associated genomic instability.

In breast cancer cohorts, *KMT2C* mutant tumors had significantly higher burden of SBS3 (Figure 5C; BRCA-EU: P-value = 2.8e–3; BRCA-FR: P-value = 1.2e–3; Mann–Whitney U-test). We observed similar results for other cancers including skin (MELA-AU), colorectal (COCA-CN), liver (LIRI-JP) and esophageal (ESAD-UK) cohorts as well (Supplementary Figure S4D). Taken together, *KMT2C* deficiency contributes to genomic instability and HR defects in human cancers.

APOBEC mutagenesis promotes genomic instability in *KMT2C* mutant tumors

To determine whether APOBEC mutagenesis in *KMT2C* deficient tumors promotes further genomic instability and favors specific classes of genomic rearrangements, we jointly examined *KMT2C* mutation status, APOBEC (SBS2 and SBS13) and HR deficiency (SBS3) signatures, as well as the frequency of different classes of genomic rearrangements for ICGC cohorts with available data (Methods, Supplementary Table S1). We divided *KMT2C* mutant and other tumors into groups that have high or low weight of APOBEC/SBS3 mutational signatures. Small deletion frequencies varied across the groups (Figure 5D; BRCA-EU, P-value = 2.2e–6; BRCA-FR, P-value = 3.5e–2); both *KMT2C* mutations and APOBEC activity were associated with increased frequency of small deletions, and the samples with high APOBEC mutational signature and *KMT2C* mutations had higher burden of deletions than any other category. Extending the analyses to other classes of genomic rearrangements in these cohorts, we observed similar results for small (<10 kb) inversions (Supplementary Figure S4E).

DISCUSSION

Mutagenesis due to cytidine deaminase functions of APOBEC family genes, particularly *APOBEC3A* and *APOBEC3B* is common in many cancer types, leading to distinct APOBEC mutational signatures, which are characterized by mutations predominantly at RTCA/YTCA contexts—that are designated as SBS2 and SBS13 in the COSMIC database (15–18). The mutational signatures appear to occur in sporadic, burst-like manner in somatic cells, as observed in lineage-tracing and biochemical experiments (40,48), and often arise late during cancer progression (49). But its triggers and impact during carcinogenesis are poorly understood. We propose a model that *KMT2C* deficiency is one of the major factors leads to replication stress and APOBEC activity, which synergistically contribute towards mutagenesis and genomic instability in some cancers (Figure 5E).

KMT2C catalyzes H3 lysine 4 methylation and establishes H3K4me1 and H3K4me3 enabling MRE11 recruitment at stalled replication forks (9,10,50,51). *KMT2C* deficiency slows down replication and delays replication fork restart by compromising TP53 dependent MRE11 recruitment to distressed replication forks. Replication stress in *KMT2C* compromised condition likely promotes APOBEC3B expression potentially via *AP-1* and *NF-κB* signaling (37) leading to elevated level of deaminase activity. Alongside, deficiency in KMT2C and TP53 mediated recruitment of MRE11 also causes delays in replication restart at stalled forks (12), where ssDNA can act as
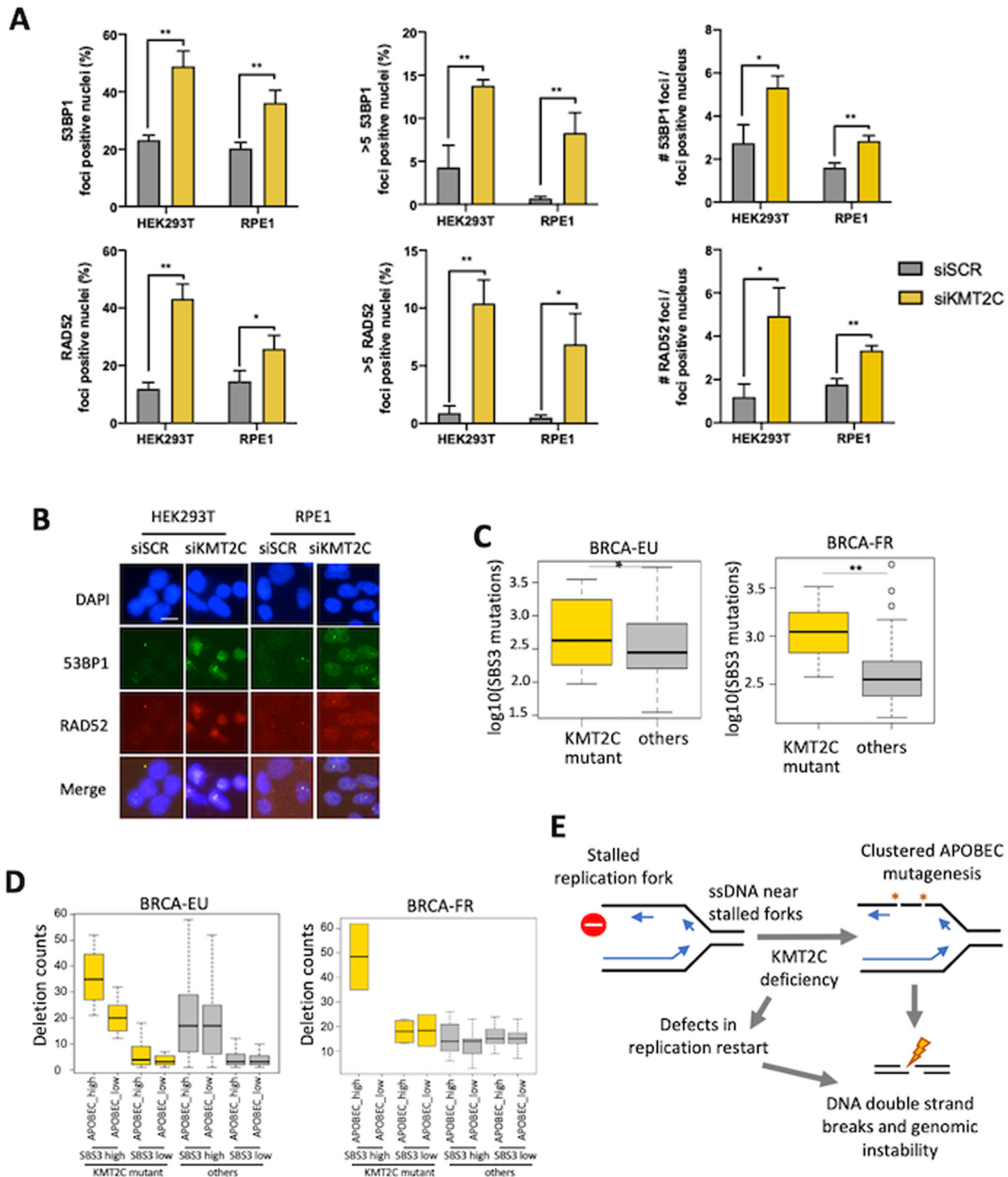
Figure 5. *KMT2C* mutant tumors have elevated genomic rearrangements and signatures of HR defects. (**A**, **B**) Immunofluorescence analysis and representative images show 53BP1 and RAD52 foci upon knockdown of indicated genes in HEK293T and RPE1 cells with error bars representing SEM. *P*-value was calculated using *t*-test; *P-value < 0.05, **P-value < 0.01. Scale bars indicate 20 μm. DAPI was used as a nuclear counterstain. (**C**) Barplots showing the burden of somatic mutations attributed to SBS3 in *KMT2C* mutant and other samples in the breast cancer cohorts BRCA-EU and BRCA-FR. Plots for other cohorts are shown in Supplementary Figure S4C. *P*-value was calculated by Mann–Whitney *U*-test, *P-value < 0.05, **P-value < 0.01. (**D**) Barplot showing the frequencies of small deletions (<10 kb) in tumor samples grouped according to *KMT2C* mutation status, and the burden of somatic mutations attributed to APOBEC and SBS3 mutational signatures in the breast cancer (BRCA-EU and BRCA-FR) cohorts. The samples with above mean burden of somatic mutations attributed to SBS3 were considered SBS3 high, and otherwise SBS3 low. Likewise, the samples with above mean burden of somatic mutations attributed to SBS2 or SBS13 were considered APOBEC_high, and the samples with below mean burden of SBS2 and SBS13 were designated as APOBEC_low. *P* value was calculated using one-way ANOVA. *P*-value: BRCA-EU: 2.2e–6; BRCA-FR: 3.5e–02. (**E**) Schematic representation of a model that defects in replication restart and APOBEC mutagenesis promote DNA double strand breaks and genomic instability in *KMT2C* deficient cells.

substrate for APOBEC-mediated mutagenesis (17,52). Our pan-cancer analysis shows that *KMT2C* mutant tumors have a significant excess of APOBEC mutational signatures. We further observed excess of somatic mutations at both the RTCA and YTCA contexts, which suggests that perhaps both APOBEC3B and APOBEC3A contribute towards elevated somatic mutation burden in *KMT2C* mutant tumors.

KMT2C deficiency also promotes HR deficiency and may lead to enhanced reliance on alternative DNA repair pathways, consistent with other reports (9). We indeed observed an increase in 53BP1 and RAD52 foci in *KMT2C* deficient cells and mutational signatures of HR deficiency in tumor genomes. Although RAD52 is implicated in a secondary role in presence of RAD51, when HR pathway is compromised, it takes up a key role in repair of collapsed DNA replication forks and delimits homology-dependent gene conversion process (12,45,52,53), but further work is required to establish DNA repair mechanisms in *KMT2C* deficient cells. Alongside, increased APOBEC deaminase activity likely confers further genomic instability by acting on long stretches of ssDNA generated on accumulated stalled forks in *KMT2C* deficient cells, leading to both point mutations and DNA double strand breaks. Abasic sites generated by APOBEC deaminase activity may further slow replication fork progression (54), which in turn may facilitate additional mutagenesis, complex processivity patterns, and double strand breaks leading to rearrangements. Indeed, APOBEC mutational signatures co-occur with complex patterns of genomic instability (55). Taken together, these mutagenic processes likely promote progressive genomic instability in *KMT2C*-compromised tumor genomes, but further mechanistic studies would be needed to establish that. While HR deficiency can be targeted by PARP inhibitors (9), APOBEC mutational signatures are considered as potential predictive markers for immunotherapy response (51). *KMT2C* is one of the most frequently mutated cancer genes, and occurs in older patient populations, often without effective targeted therapies. In such patients, a composite mutational signature of *KMT2C* deficiency may help explore opportunities for effective treatment strategies.

DATA AVAILABILITY

All publicly available genomic data was obtained from the ICGC/TCGA data portal (https://dcc.icgc.org/), and the details are in source data files. Additional data could be obtained from the group's Github page https://github.com/sjdlabgroup/KMT2C_APOBEC_resources/ or from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

ACKNOWLEDGEMENTS

The funders had no role in the design and conduct of the study, nor the decision to prepare and submit the manuscript for publication. The authors thank Vinod Singh for input regarding mutSigTools, Yizhou He for the RPE1 cell line, and Dr Bing Xia for MDA-MB-231 cell line and also assistance with the DNA fiber assay.

REFERENCES

1. Husmann,D. and Gozani,O. (2019) Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.*, **26**, 880.
2. Kandoth,C., McLellan,M.D., Vandin,F., Ye,K., Niu,B., Lu,C., Xie,M., Zhang,Q., McMichael,J.F., Wyczalkowski,M.A. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, **502**, 333–339.
3. Lawrence,M.S., Stojanov,P., Mermel,C.H., Robinson,J.T., Garraway,L.A., Golub,T.R., Meyerson,M., Gabriel,S.B., Lander,E.S. and Getz,G. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, **505**, 495–501.
4. Tubbs,A. and Nussenzweig,A. (2017) Endogenous DNA damage as a source of genomic instability in cancer. *Cell*, **168**, 644–656.
5. Ellis,M.J., Ding,L., Shen,D., Luo,J., Suman,V.J., Wallis,J.W., Van Tine,B.A., Hoog,J., Goiffon,R.J., Goldstein,T.C. *et al.* (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, **486**, 353–360.
6. Gala,K., Li,Q., Sinha,A., Razavi,P., Dorso,M., Sanchez-Vega,F., Chung,Y.R., Hendrickson,R., Hsieh,J.J., Berger,M. *et al.* (2018) KMT2C mediates the estrogen dependence of breast cancer through regulation of ERα enhancer function. *Oncogene*, **37**, 4692–4710.
7. Kaikkonen,M.U., Spann,N.J., Heinz,S., Romanoski,C.E., Allison,K.A., Stender,J.D., Chun,H.B., Tough,D.F., Prinjha,R.K., Benner,C. *et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell*, **51**, 310–325.
8. Lavery,W.J., Barski,A., Wiley,S., Schorry,E.K. and Lindsley,A.W. (2020) KMT2C/D COMPASS complex-associated diseases [K CD COM-ADs]: an emerging class of congenital regulopathies. *Clin. Epigenetics*, **12**, 10.
9. Rampias,T., Karagiannis,D., Avgeris,M., Polyzos,A., Kokkalis,A., Kanaki,Z., Kousidou,E., Tzetis,M., Kanavakis,E., Stravodimos,K. *et al.* (2019) The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. *EMBO Rep.*, **20**, e46821.
10. Chaudhuri,A.R., Callen,E., Ding,X., Gogola,E., Duarte,A.A., Lee,J.E., Wong,N., Lafarga,V., Calvo,J.A., Panzarino,N.J. *et al.* (2016) Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature*, **535**, 382–387.
11. Berti,M. and Vindigni,A. (2016) Replication stress: getting back on track. *Nat. Struct. Mol. Biol.*, **23**, 103–109.
12. Roy,S., Tomaszowski,K.H., Luzwick,J.W., Park,S., Li,J., Murphy,M. and Schlacher,K. (2018) p53 orchestrates DNA replication restart homeostasis by suppressing mutagenic RAD52 and POLθ pathways. *Elife*, **7**, e31723.
13. Salter,J.D., Bennett,R.P. and Smith,H.C. (2016) The APOBEC protein family: United by structure, divergent in function. *Trends Biochem. Sci.*, **41**, 578–594.
14. Roberts,S.A., Lawrence,M.S., Klimczak,L.J., Grimm,S.A., Fargo,D., Stojanov,P., Kiezun,A., Kryukov,G.V., Carter,S.L., Saksena,G. *et al.* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.*, **45**, 970–976.
15. Burns,M.B., Temiz,N.A. and Harris,R.S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.*, **45**, 977.
16. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Aparicio,S.A.J.R.J.R., Behjati,S., Biankin,A.V., Bignell,G.R., Bolli,N., Borg,A., Børresen-Dale,A.-L.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.

17. Alexandrov,L.B., Kim,J., Haradhvala,N.J., Huang,M.N., Tian Ng,A.W., Wu,Y., Boot,A., Covington,K.R., Gordenin,D.A., Bergstrom,E.N. *et al.* (2020) The repertoire of mutational signatures in human cancer. *Nature*, **578**, 94–101.

18. Nik-Zainal,S., Alexandrov,L.B., Wedge,D.C., Van Loo,P., Greenman,C.D., Raine,K., Jones,D., Hinton,J., Marshall,J., Stebbings,L.A. *et al.* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell*, **149**, 979–993.

19. Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Sander,C., Stuart,J.M., Chang,K., Creighton,C.J. *et al.* (2013) The cancer genome atlas pan-cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.

20. Hudson,T.J., Anderson,W., Aretz,A., Barker,A.D., Bell,C., Bernabé,R.R., Bhan,M.K., Calvo,F., Eerola,I., Gerhard,D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.

21. Cingolani,P., Platts,A., Wang,L.L., Coon,M., Nguyen,T., Wang,L., Land,S.J., Lu,X. and Ruden,D.M. (2012) 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).*, **6**, 80–92.

22. Mayakonda,A., Lin,D.C., Assenov,Y., Plass,C. and Koeffler,H.P. (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.*, **28**, 1747–1756.

23. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2014) The UCSC genome browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.

24. Nones,K., Johnson,J., Newell,F., Patch,A.M., Thorne,H., Kazakoff,S.H., De Luca,X.M., Parsons,M.T., Ferguson,K., Reid,L.E. *et al.* (2019) Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, **30**, 1071–1079.

25. Rosenthal,R., McGranahan,N., Herrero,J., Taylor,B.S. and Swanton,C. (2016) DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.*, **17**, 31.

26. Wang,S., Jia,M., He,Z. and Liu,X.S. (2018) APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene*, **37**, 3924–3936.

27. Jarvis,M.C., Ebrahimi,D., Temiz,N.A. and Harris,R.S. (2018) Mutation signatures including APOBEC in cancer cell lines. *JNCI Cancer Spectr.*, **2**, pky002.

28. Chan,K., Roberts,S.A., Klimczak,L.J., Sterling,J.F., Saini,N., Malc,E.P., Kim,J., Kwiatkowski,D.J., Fargo,D.C., Mieczkowski,P.A. *et al.* (2015) An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.*, **47**, 1067–1072.

29. Singh,V.K., Rastogi,A., Hu,X., Wang,Y. and De,S. (2020) Mutational signature SBS8 predominantly arises due to late replication errors in cancer. *Commun. Biol.*, **3**, 421.

30. Nik-Zainal,S., Davies,H., Staaf,J., Ramakrishna,M., Glodzik,D., Zou,X., Martincorena,I., Alexandrov,L.B., Martin,S., Wedge,D.C. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, **534**, 47–54.

31. Chen,X., Bahrami,A., Pappo,A., Easton,J., Dalton,J., Hedlund,E., Ellison,D., Shurtleff,S., Wu,G., Wei,L. *et al.* (2014) Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.*, **7**, 104–112.

32. Corces,M.R., Granja,J.M., Shams,S., Louie,B.H., Seoane,J.A., Zhou,W., Silva,T.C., Groeneveld,C., Wong,C.K., Cho,S.W. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science*, **362**, eaav1898.

33. Le Dily,F.L., Baù,D., Pohl,A., Vicent,G.P., Serra,F., Soronellas,D., Castellano,G., Wright,R.H.G., Ballare,C., Filion,G. *et al.* (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.*, **28**, 2151–2162.

34. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

35. Morganella,S., Alexandrov,L.B., Glodzik,D., Zou,X., Davies,H., Staaf,J., Sieuwerts,A.M., Brinkman,A.B., Martin,S., Ramakrishna,M. *et al.* (2016) The topography of mutational processes in breast cancer genomes. *Nat. Commun.*, **7**, 11383.

36. Huang,X., Wojtowicz,D. and Przytycka,T.M. (2018) Detecting presence of mutational signatures in cancer with confidence. *Bioinformatics*, **34**, 330–337.

37. Lin,L., Holmes,B., Shen,M.W., Kammeron,D., Geijsen,N., Gifford,D.K. and Sherwood,R.I. (2020) Comprehensive mapping of key regulatory networks that drive oncogene expression. *Cell Rep.*, **33**, 108426.

38. Nair,S., Sanchez-Martinez,S., Ji,X. and Rein,A. (2014) Biochemical and biological studies of mouse APOBEC3. *J. Virol.*, **88**, 3850–3860.

39. Matsuno,Y., Atsumi,Y., Shimizu,A., Katayama,K., Fujimori,H., Hyodo,M., Minakawa,Y., Nakatsu,Y., Kaneko,S., Hamamoto,R. *et al.* (2019) Replication stress triggers microsatellite destabilization and hypermutation leading to clonal expansion in vitro. *Nat. Commun.*, **10**, 3925.

40. Kanu,N., Cerone,M.A., Goh,G., Zalmas,L.P., Bartkova,J., Dietzen,M., McGranahan,N., Rogers,R., Law,E.K., Gromova,I. *et al.* (2016) DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol.*, **17**, 185.

41. Trenz,K., Smith,E., Smith,S. and Costanzo,V. (2006) ATM and ATR promote Mre11 dependent restart of collapsed replication forks and prevent accumulation of DNA breaks. *EMBO J.*, **25**, 1764–1774.

42. Meister,P., Taddei,A. and Gasser,S.M. (2006) In and out of the replication factory. *Cell*, **125**, 1233–1235.

43. Her,J. and Bunting,S.F. (2018) How cells ensure correct repair of DNA double-strand breaks. *J. Biol. Chem.*, **293**, 10502–10511.

44. Sunada,S., Nakanishi,A. and Miki,Y. (2018) Crosstalk of DNA double-strand break repair pathways in poly(ADP-ribose) polymerase inhibitor treatment of breast cancer susceptibility gene 1/2-mutated cancer. *Cancer Sci.*, **109**, 893–899.

45. Yan,Z., Xue,C., Kumar,S., Crickard,J.B., Yu,Y., Wang,W., Pham,N., Li,Y., Niu,H., Sung,P. *et al.* (2019) Rad52 restrains resection at DNA double-strand break ends in yeast. *Mol. Cell*, **76**, 699–711.

46. Lok,B.H. and Powell,S.N. (2012) Molecular pathways: Understanding the role of Rad52 in homologous recombination for therapeutic advancement. *Clin. Cancer Res.*, **18**, 6400–6406.

47. Malacaria,E., Honda,M., Franchitto,A., Spies,M. and Pichierri,P. (2020) Physiological and pathological roles of RAD52 at DNA replication forks. *Cancers (Basel.)*, **12**, 402.

48. Petljak,M., Alexandrov,L.B., Brammeld,J.S., Price,S., Wedge,D.C., Grossmann,S., Dawson,K.J., Ju,Y.S., Iorio,F., Tubio,J.M.C. *et al.* (2019) Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell, **176**, 1282–1294.

49. Jamal-Hanjani,M., Wilson,G.A., McGranahan,N., Birkbak,N.J., Watkins,T.B.K., Veeriah,S., Shafi,S., Johnson,D.H., Mitter,R., Rosenthal,R. *et al.* (2017) Tracking the evolution of non–small-cell lung cancer. *N. Engl. J. Med.*, **376**, 2109–2121.

50. Dorighi,K.M., Swigut,T., Henriques,T., Bhanu,N.V., Scruggs,B.S., Nady,N., Still,C.D., Garcia,B.A., Adelman,K. and Wysocka,J. (2017) Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell*, **66**, 568–576.

51. Wang,C., Lee,J.E., Lai,B., Macfarlan,T.S., Xu,S., Zhuang,L., Liu,C., Peng,W. and Ge,K. (2016) Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 11871–11876.

52. Shimizu,A., Fujimori,H., Minakawa,Y., Matsuno,Y., Hyodo,M., Murakami,Y. and Yoshioka,K. ichi (2018) Onset of deaminase APOBEC3B induction in response to DNA double-strand breaks. *Biochem. Biophys. Rep.*, **16**, 115–121.

53. Mladenov,E., Staudt,C., Soni,A., Murmann-Konda,T., Siemann-Loekes,M. and Iliakis,G. (2020) Strong suppression of gene conversion with increasing DNA double-strand break load delimited by 53BP1 and RAD52. *Nucleic Acids Res.*, **48**, 1905–1924.

54. Mehta,A., Beach,A. and Haber,J.E. (2017) Homology requirements and competition between gene conversion and break-induced replication during double-strand break repair. *Mol. Cell*, **65**, 515–526.

55. Jakobsdottir,G.M., Brewer,D.S., Cooper,C., Green,C. and Wedge,D.C. (2022) APOBEC3 mutational signatures are associated with extensive and diverse genomic instability across multiple tumour types. *BMC Biol.*, **20**, 117.