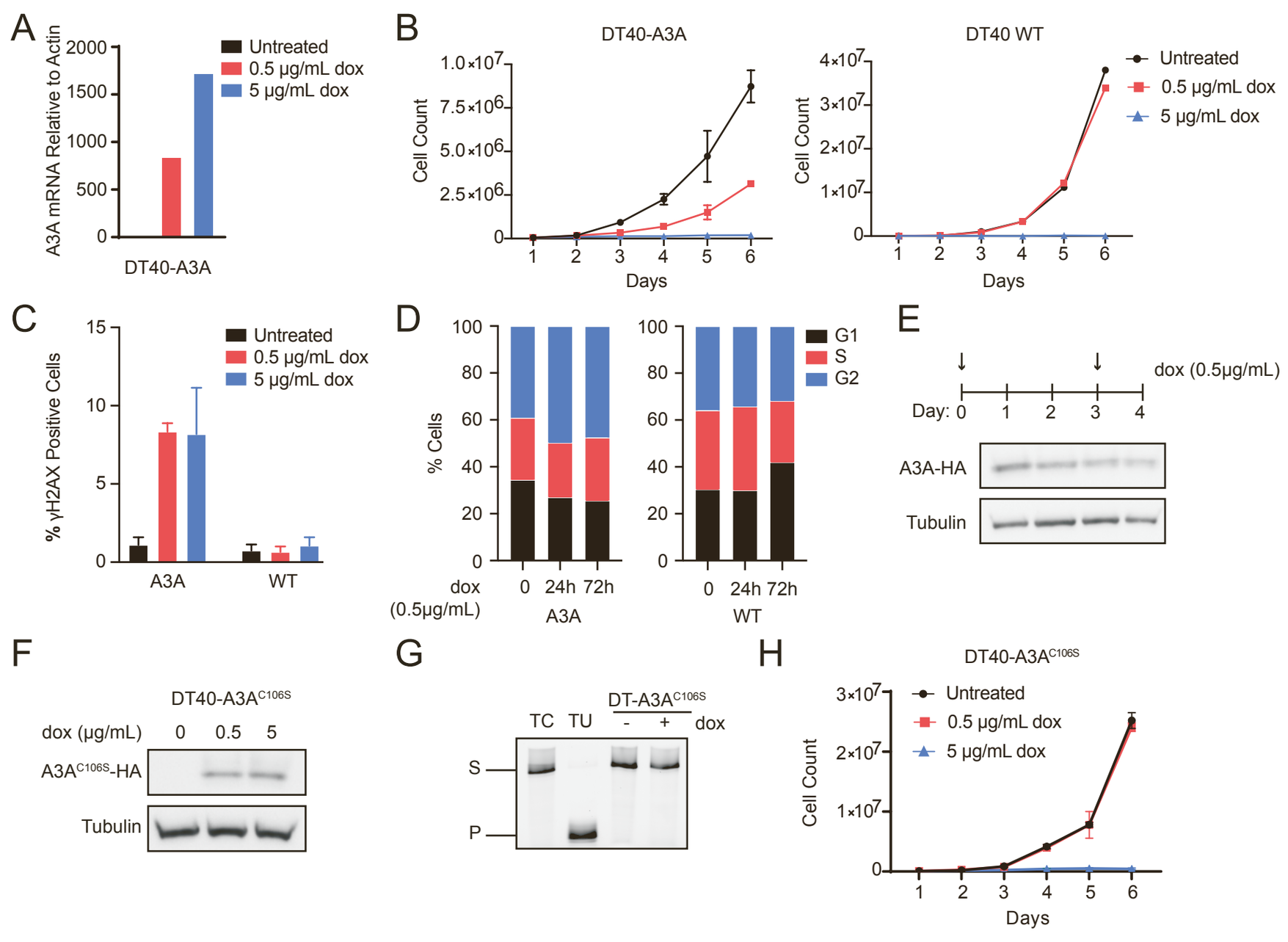
Cell Reports, Volume *38*

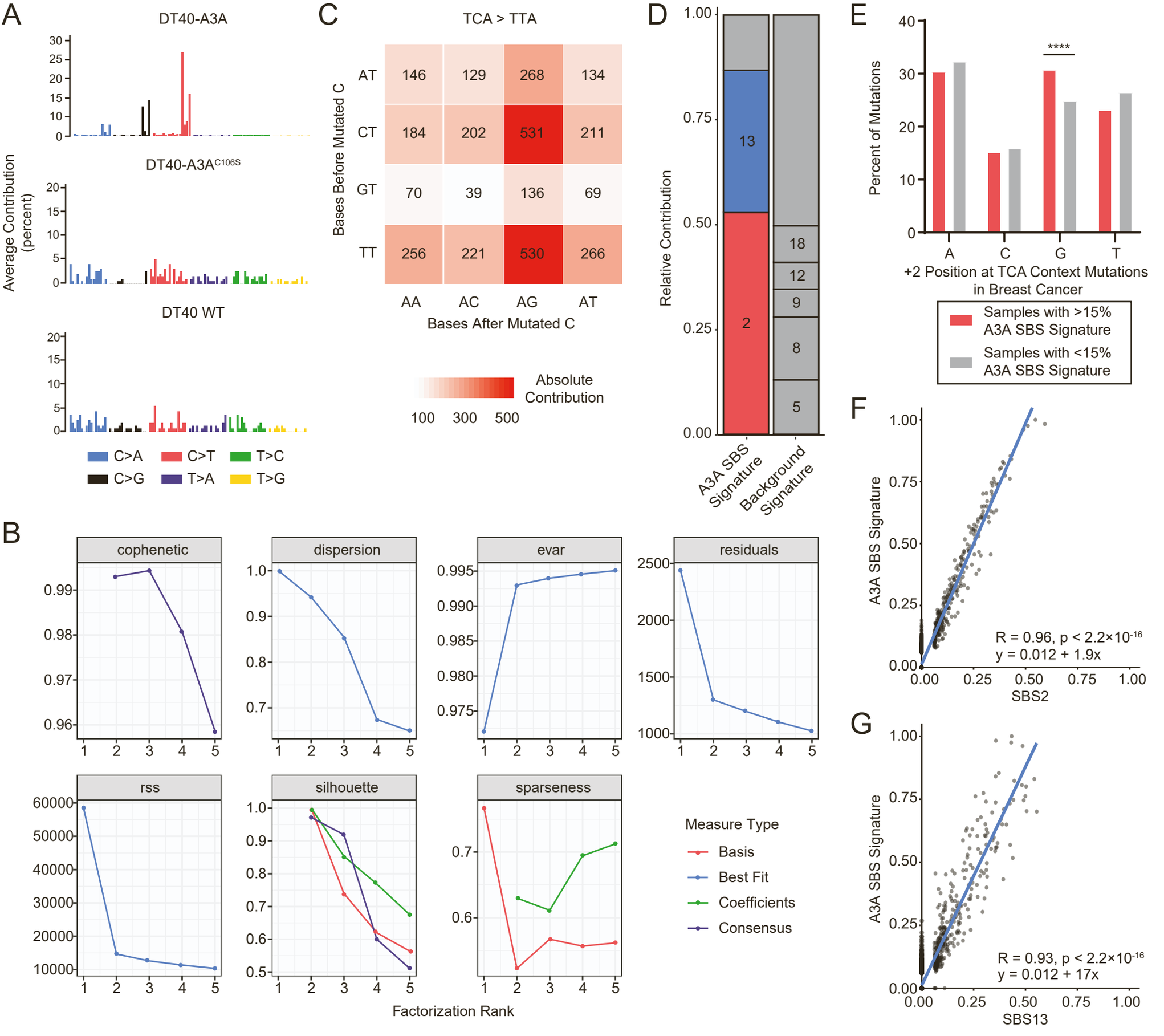
Supplemental information

Prospectively defined patterns of APOBEC3A mutagenesis are prevalent in human cancers

Rachel A. DeWeerd, Eszter Németh, Ádám Póti, Nataliya Petryk, Chun-Long Chen, Olivier Hyrien, Dávid Szüts, and Abby M. Green

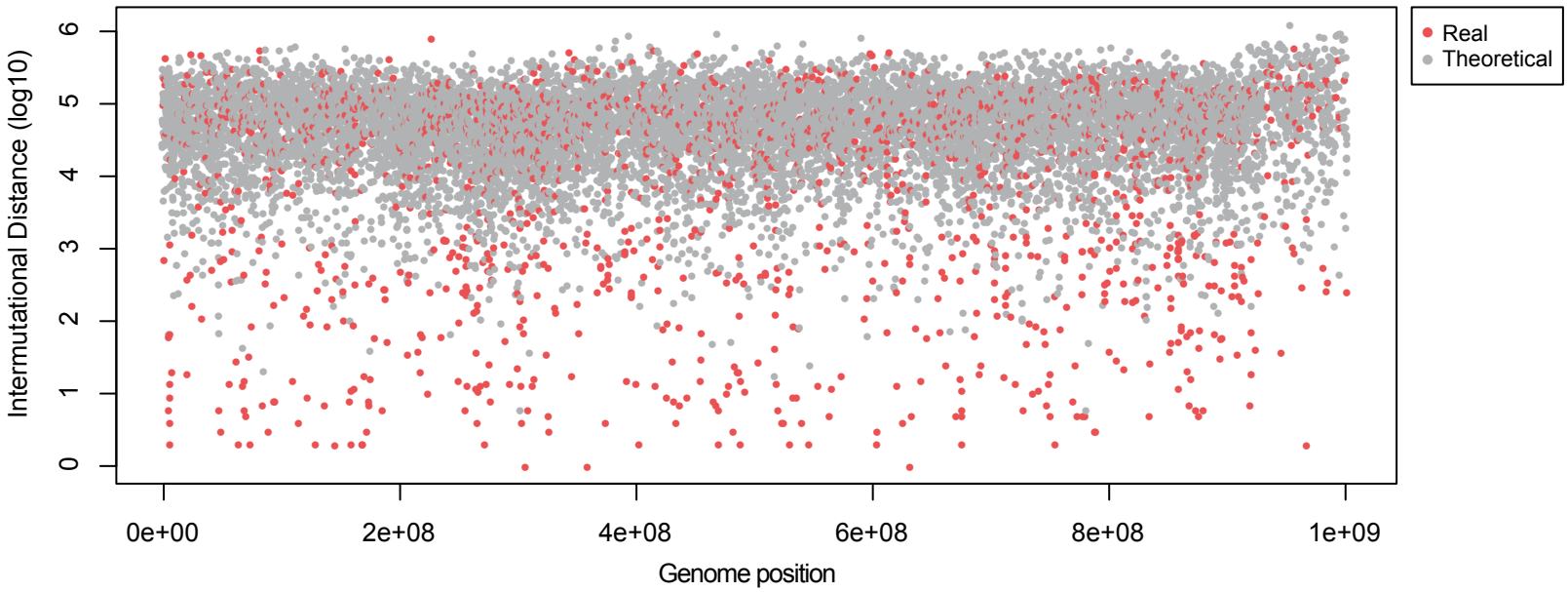

A
A3A mRNA Relative to Actin
2000
1500
1000
500
0
DT40-A3A
Untreated
0.5 μg/mL dox
5 μg/mL dox
B
DT40-A3A
Cell Count
1.0×10^7
7.5×10^6
5.0×10^6
2.5×10^6
0
1 2 3 4 5 6
Days
DT40 WT
Cell Count
4×10^7
3×10^7
2×10^7
1×10^7
0
1 2 3 4 5 6
Days
Untreated
0.5 μg/mL dox
5 μg/mL dox
C
% γH2AX Positive Cells
15
10
5
0
A3A
WT
Untreated
0.5 μg/mL dox
5 μg/mL dox
D
% Cells
100
80
60
40
20
0
dox (0.5μg/mL)
0 24h 72h
A3A
0 24h 72h
WT
G1
S
G2
E
dox (0.5μg/mL)
Day: 0 1 2 3 4
A3A-HA
Tubulin
F
DT40-A3A^C106S
dox (μg/mL) 0 0.5 5
A3A^C106S-HA
Tubulin
G
DT-A3A^C106S
TC TU - + dox
S
P
H
DT40-A3A^C106S
Cell Count
3×10^7
2×10^7
1×10^7
0
1 2 3 4 5 6
Days
Untreated
0.5 μg/mL dox
5 μg/mL dox

Supplementary Figure 1. Human APOBEC3A is active in avian DT40 cells. A) Expression of doxycycline (dox)-induced APOBEC3A transcripts was detected by quantitative real-time PCR. The average of three technical replicates is shown. **B**) APOBEC3A expression causes a proliferative defect in DT40 cells. Cells were treated with dox every 3 days and daily total cell counts were recorded. Average and SD of two independent experiments is shown. **C**) DNA damage in DT40-A3A cells. Cells were treated with dox for 3 days prior to intracellular γH2AX staining and evaluation by FACS. Average and SD of three independent experiments is shown. **D**) Cell cycle analysis in DT40-A3A cells. Cells were treated with dox prior to propidium iodide staining and FACS analysis. Proportion of cells in each cell cycle phase are displayed as an average of three independent experiments. **E**) APOBEC3A expression in DT40-A3A cells treated with dox every 3 days was evaluated by immunoblot. Image representative of two biological replicates. **F**) Dox-induced expression of the catalytically inactive $A3A^{C106S}$ mutant in DT40 cells is evaluated by immunoblot. The $A3A^{C106S}$ transgene is detected by C-terminal HA tag. Image representative of two biological replicates. **G**) Deaminase activity in DT40-$A3A^{C106S}$ cells. Experiments were performed as described in Figure 1B. **H**) Proliferation of DT40-$A3A^{C106S}$ cells. Cells were treated with indicated dox doses and proliferation was measured by daily cell counts. Average and SD of three independent experiments is shown. Related to Figure 1.

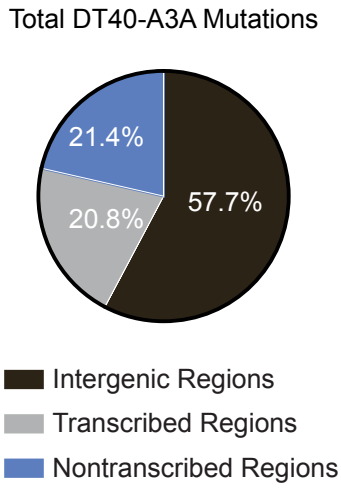

A
DT40-A3A
DT40-A3A^C106S
DT40 WT
Average Contribution (percent)
C>A
C>T
T>C
C>G
T>A
T>G
C
TCA > TTA
Bases Before Mutated C
AT
CT
GT
TT
146
129
268
134
184
202
531
211
70
39
136
69
256
221
530
266
AA
AC
AG
AT
Bases After Mutated C
Absolute Contribution
100
300
500
D
Relative Contribution
13
2
18
12
9
8
5
A3A SBS Signature
Background Signature
E
Percent of Mutations
****
A
C
G
T
+2 Position at TCA Context Mutations in Breast Cancer
Samples with >15% A3A SBS Signature
Samples with <15% A3A SBS Signature
B
cophenetic
dispersion
evar
residuals
rss
silhouette
sparseness
Factorization Rank
Measure Type
Basis
Best Fit
Coefficients
Consensus
F
A3A SBS Signature
R = 0.96, p < 2.2×10^-16
y = 0.012 + 1.9x
SBS2
G
A3A SBS Signature
R = 0.93, p < 2.2×10^-16
y = 0.012 + 17x
SBS13

Supplementary Figure 2. The genome-wide spectrum of APOBEC3A-mediated mutations. **A**) Spectra of mutations of descendent clones from DT40-A3A (top), DT40-A3A^{C106S} (middle), and DT40 wild-type treated with dox (bottom). Mutations identified by WGS are characterized by base substitution (color-coded key below graphs) and trinucleotide context (lower x-axis, in the same order as is labeled in Fig. 2A). The average contribution of each mutation among all sequenced descendent clones is displayed. **B**) Non-negative matrix factorization output for deriving signatures A3A and background, from the R package MutationalPatterns. A factorization rank of 2 was chosen based on the cophenetic correlation coefficient and the residual sum of squares (rss) values. **C**) Heatmap of the pentanucleotide context around T<u>C</u>A>TTA mutations in all DT40-A3A genomes. Numbers indicate absolute quantity of mutated cytidines within each 5 bp context. **D**) Deconstruction of SBS mutational signatures from DT40-A3A genomes into COSMIC v3 SBS signatures. The contribution of each COSMIC SBS signature to both experimentally-defined signatures from DT40-A3A genomes is shown. Signatures previously attributed to APOBEC3 activity, SBS2 and SBS13, are highlighted. Empty gray bar denotes combined remaining COSMIC SBS signatures with individual contributions below 6%. **E**) T<u>C</u>A context mutations in breast cancers were evaluated for the extended nucleotide context. The frequency of +2 bases in samples in which the A3A SBS Signature contributed to greater or less than 15% of total mutations. The two distributions were significantly different ($p<0.0001$, χ^2 test). The correlation between the A3A SBS Signature and SBS2 (**F**) or SBS13 (**G**) in all PCAWG tumors was assessed. Each point represents a single tumor sample. The slope, R and p values of linear regression are shown. Related to Figure 2.

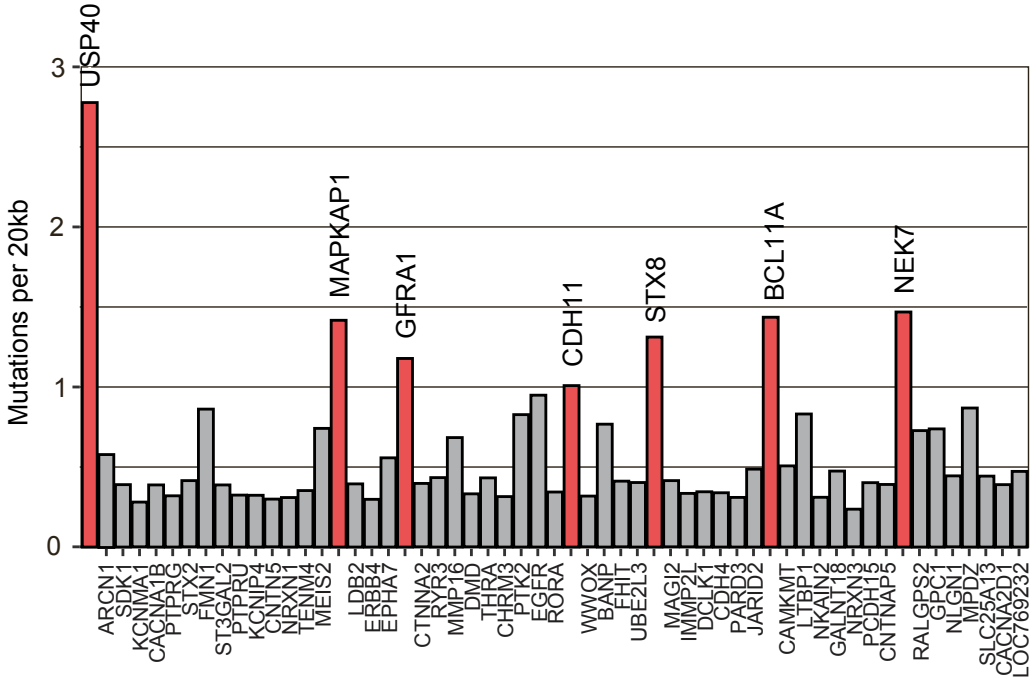
A


All DT40-A3A mutations
Intermutational Distance (log10)
0
1
2
3
4
5
6
0e+00
2e+08
4e+08
6e+08
8e+08
1e+09
Genome position
Real
Theoretical

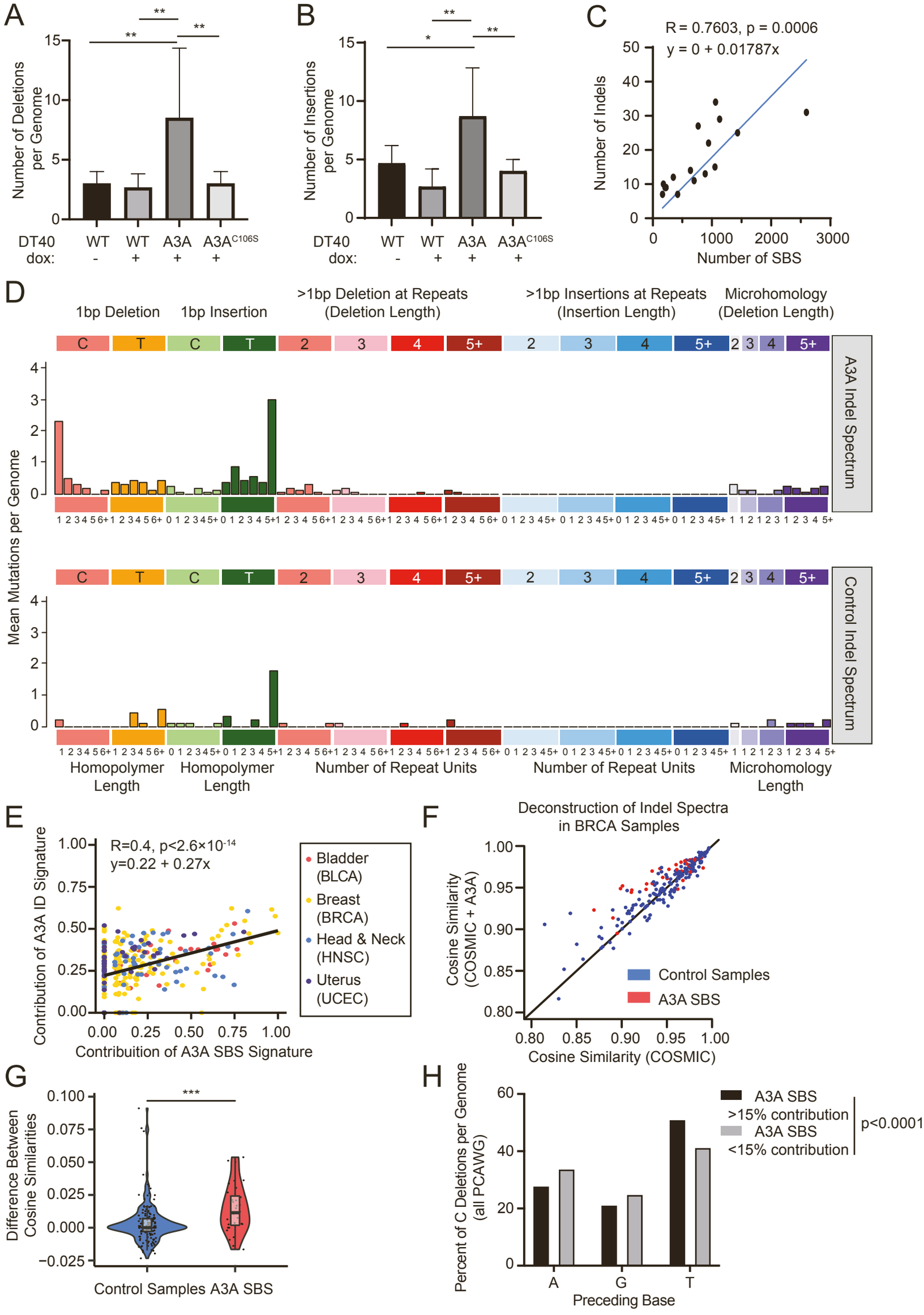
B


Total DT40-A3A Mutations
21.4%
20.8%
57.7%
Intergenic Regions
Transcribed Regions
Nontranscribed Regions

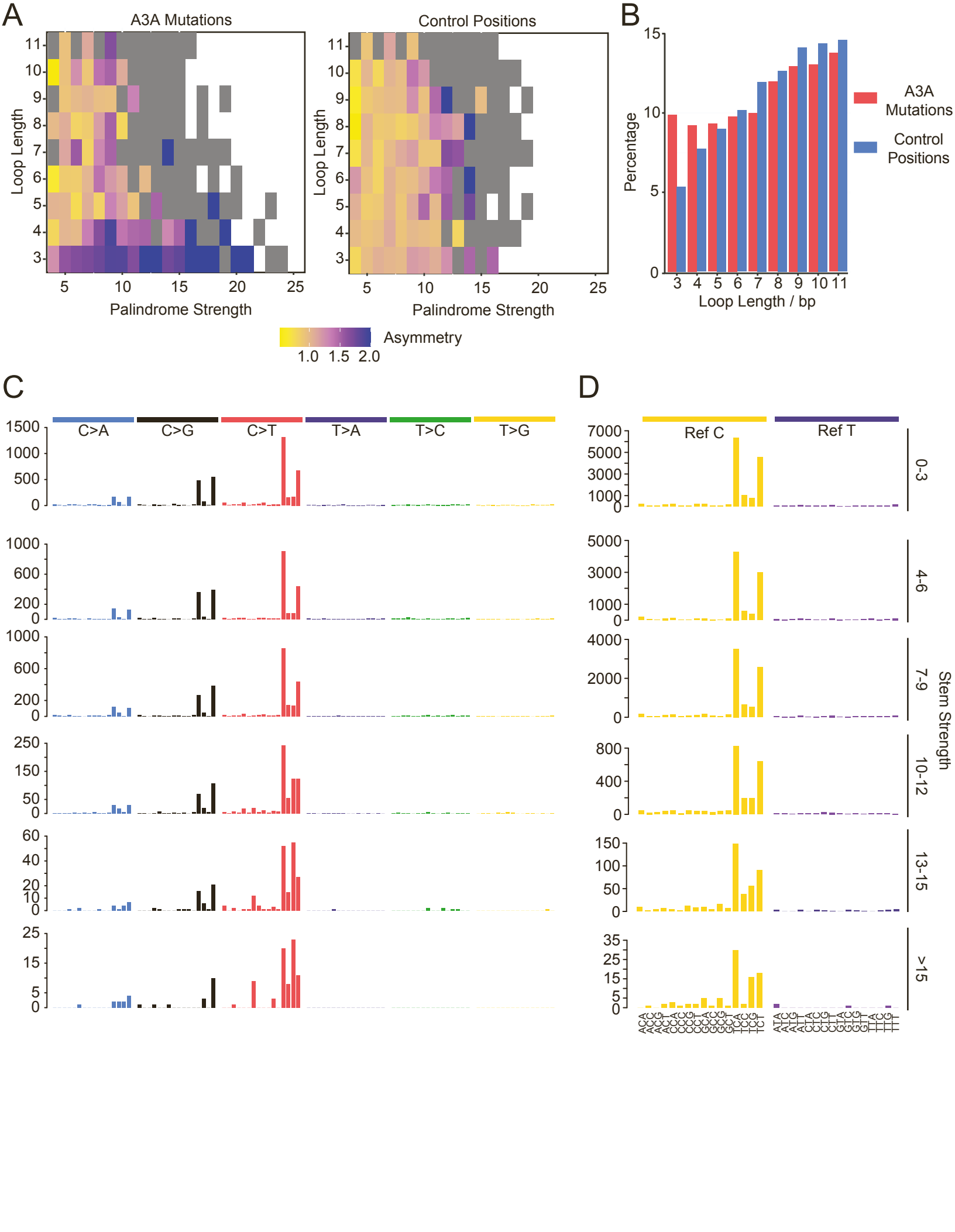
C


Mutations per 20kb
0
1
2
3
USP40
MAPKAP1
GFRA1
CDH11
STX8
BCL11A
NEK7
ARCN1
SDK1
KCNMA1
CACNA1B
PTPRG
STX2
FMN1
ST3GAL2
PTPRU
KCNIP4
CNTN5
NRXN1
TENM4
MEIS2
LDB2
ERBB4
EPHA7
CTNNA2
RYR3
MMP16
DMD
THRA
CHRM3
PTK2
EGFR
RORA
WWOX
BANP
FHIT
UBE2L3
MAGI2
IMMP2L
DCLK1
CDH4
PARD3
JARID2
CAMKMT
LTBP1
NKAIN2
GALNT18
NRXN3
PCDH15
CNTNAP5
RALGPS2
GPC1
NLGN1
MPDZ
SLC25A13
CACNA2D1
LOC769232

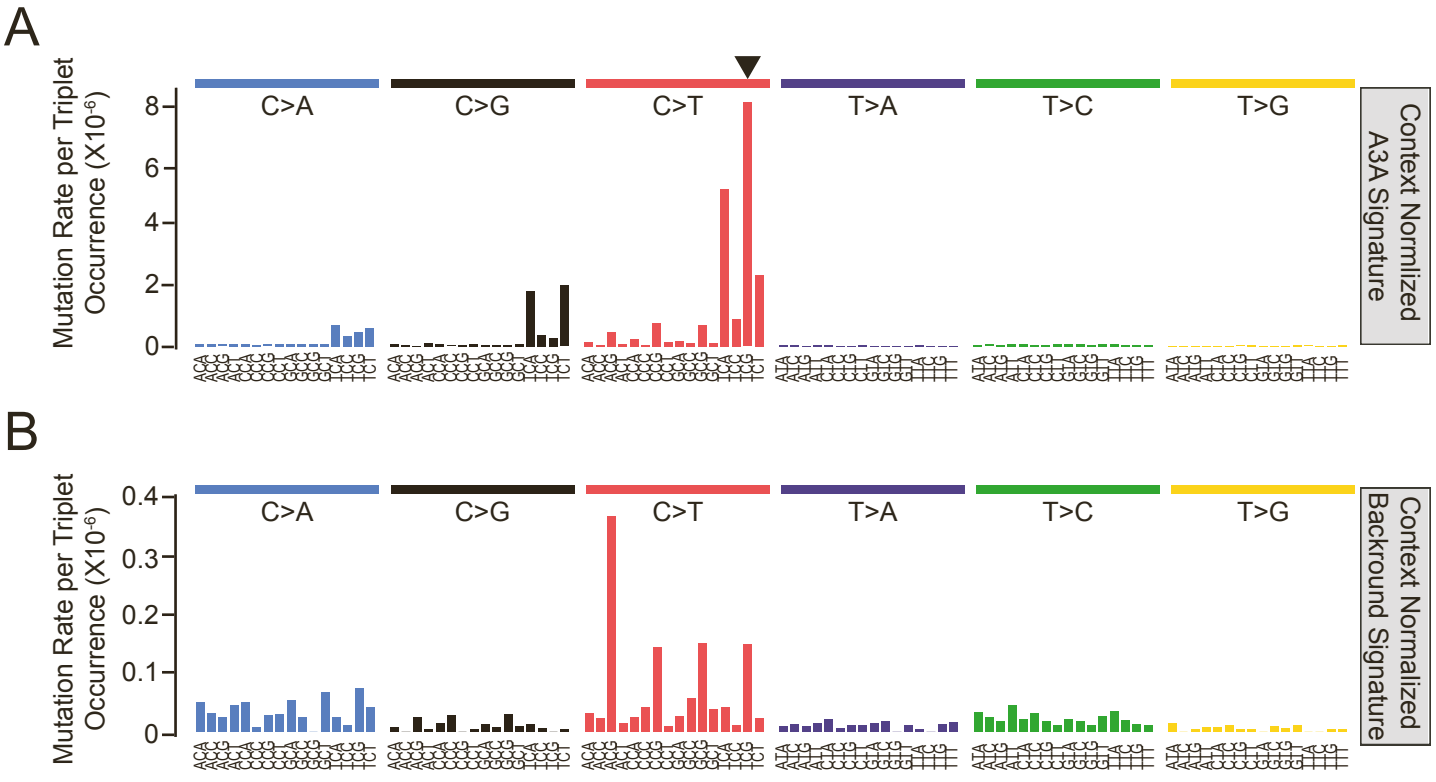
Supplementary Figure 3. Mutagenesis by APOBEC3A is stochastic. A) Rainfall plot of mutations in all DT40-A3A descendent clone genomes. Each dot represents a mutation plotted by genome position (x-axis) and distance from the previous mutation (y-axis). Mutation distances within the same genome are plotted in red (real distances), those in different genomes are plotted in grey (theoretical distances). **B**) Mutations in intergenic, transcribed, and nontranscribed regions are shown as a percentage of all mutations found in DT40-A3A descendent clones. **C**) Genes longer than 1000bp and mutated in more than 20% ($n \geq 3$) of the DT40-A3A descendent clones are shown. Genes containing more than one mutation per 20kb are highlighted in red. Related to Figure 2.


A
Number of Deletions per Genome
DT40 dox:
WT -
WT +
A3A +
A3A C106S +
B
Number of Insertions per Genome
C
R = 0.7603, p = 0.0006
y = 0 + 0.01787x
Number of Indels
Number of SBS
D
1bp Deletion
1bp Insertion
>1bp Deletion at Repeats (Deletion Length)
>1bp Insertions at Repeats (Insertion Length)
Microhomology (Deletion Length)
Mean Mutations per Genome
A3A Indel Spectrum
Control Indel Spectrum
Homopolymer Length
Homopolymer Length
Number of Repeat Units
Number of Repeat Units
Microhomology Length
E
R=0.4, p<2.6×10-14
y=0.22 + 0.27x
Contribution of A3A ID Signature
Contribuition of A3A SBS Signature
Bladder (BLCA)
Breast (BRCA)
Head & Neck (HNSC)
Uterus (UCEC)
F
Deconstruction of Indel Spectra in BRCA Samples
Cosine Similarity (COSMIC + A3A)
Cosine Similarity (COSMIC)
Control Samples
A3A SBS
G
Difference Between Cosine Similarities
Control Samples
A3A SBS
H
Percent of C Deletions per Genome (all PCAWG)
A3A SBS >15% contribution
A3A SBS <15% contribution
p<0.0001
Preceding Base

Supplementary Figure 4. APOBEC3A activity results in genomic indels. Indels were quantified for all descendent clone genomes. The average number of deletions (**A**) or insertions (**B**) per genome for each group of descendent clones is shown. Statistical analysis was performed using a two-tailed t-test, error bars indicate SD, *p<0.05, **p<0.01. (**C**) The number of indels correlated with the number of SBS in DT40-A3A genomes. Linear regression was used to determine slope, R value and significance. **D**) The full indel spectra in DT40-A3A (top) and control samples (wild-type with and without dox and DT40-A3A^{C106S}, bottom) are shown as an average number of indels per genome characterized by length and context. **E**) The experimentally-defined A3A SBS and ID signatures in human cancers. Tumor genomes from PCAWG were analyzed for the presence of the A3A SBS Signature (x-axis) and the A3A ID Signature (y-axis). The slope, R and p values of linear regression on the combined dataset are shown. **F**) Supplementation of the COSMIC indel signature reference set with the experimental A3A SBS Signature results in an improvement of signature deconstruction in samples with A3A SBS Signature mutations (red) but not in samples without A3A SBS signature mutations (blue). The graph shows the correlation of the cosine similarities of the original and reconstructed spectrum in case of the deconstruction with the reference set only (horizontal axis) or the reference set plus A3A ID Signature (vertical axis). The success of deconstruction was measured as the cosine similarity between the detected and the reconstructed mutational spectrum. **G**) The violin plot shows that cosine similarities of the deconstructed indel spectra (see Methods) versus the original indel spectra were significantly higher when the reference signature set included the A3A ID Signature (p=0.00027, median and quartiles are indicated). **H**) Single C deletions from all PCAWG genomes were evaluated for nucleotide context. The base preceding single C deletions in cancers with and without the A3A SBS Signature is depicted as an average (p<0.0001, X^2 test). Related to Figure 3.


A
A3A Mutations
Control Positions
Loop Length
Palindrome Strength
Asymmetry
1.0 1.5 2.0
B
Percentage
Loop Length / bp
A3A Mutations
Control Positions
C
C>A
C>G
C>T
T>A
T>C
T>G
D
Ref C
Ref T
Stem Strength
0-3
4-6
7-9
10-12
13-15
>15

Supplementary Figure 5. APOBEC3A mutates cytidines within DNA stem loops with strong stems and small loops. A) C/G asymmetry of mutations at the first and last loop positions, defined as the proportion of G in the first position plus the proportion of C in the last position. Because of the sequence preference of A3A, mutations in DNA stem loops show an asymmetric distribution of G and C residues when the mutated position is at the first or last position of the loop. **B**) Compared to the randomly selected control positions, the stem loops mutated by APOBEC3A are represented by a higher fraction of loops as short as 3-5 bases, and less by longer loops. This can also indicate that a fraction of the predicted large loops are not stable and APOBEC3A selects for the loops that are truly formed. **C)** The spectrum of mutations found in loops of stem loops with strong stems differs from the spectrum of APOBEC mutations that are not in a loop or are in a loop with a weaker stem. Lower x-axis is as in Figure 2A. **D**) Random genomic positions selected to match the overall A3A spectrum show the same phenomenon suggesting that the mutation preferences of A3A are the same in the loops as in other parts of the genome. The different spectrum may be caused by the GC rich composition of DNA sequences prone to stem loop formation. Related to Figure 5.


A
Mutation Rate per Triplet Occurrence (X10-6)
8
6
4
2
0
C>A
C>G
C>T
T>A
T>C
T>G
Context Normlized A3A Signature
B
Mutation Rate per Triplet Occurrence (X10-6)
0.4
0.3
0.2
0.1
0
C>A
C>G
C>T
T>A
T>C
T>G
Context Normalized Backround Signature

Supplementary Figure 6. Frequency of all cytidine mutations within a trinucleotide context. Mutations in DT40-A3A genomes (**A**) and control genomes (**B**) were normalized to the frequency of each trinucleotide occurrence within the reference genome. The normalized spectrum is shown, arrow denotes TCG context. Related to Figure 5.

		IsoMut only		Both		MuTect2 only		**Tool**
		Good	Bad	Good	Bad	Good	Bad	**Validation**
A3A-scD	DEL	1	3	6	0	5	2	
	INS	3	2	10	0	11	7	
A3A-scE	DEL	7	0	11	0	8	7	
	INS	4	0	8	0	4	9	
A3A-scJ	DEL	4	0	6	0	7	4	
	INS	5	2	5	0	3	10	
WT-scE	DEL	1	0	1	0	1	1	
	INS	0	2	1	0	3	7	
WT-scH	DEL	0	0	4	0	1	2	
	INS	0	0	1	0	8	3	
WT-scO	DEL	0	0	2	0	1	3	
	INS	2	1	1	0	1	0	
Sample	**Type**							
	sum(DEL)	13	3	30	0	23	19	
	sum(INS)	14	7	26	0	30	36	
	sum(total)	**27**	**10**	**56**	**0**	**53**	**55**	

Supplementary Table 1. Comparison of indel detection using IsoMut or Mutect2 on 6 representative samples. IsoMut column indicates same events as reported in the manuscript (Figure 3). MuTect2 column shows indel detection using 29 WT DT40 whole genome sequenced samples as a 'panel of normals', the ancestral clone of the sample as normal, and filter settings PASS, TLOD > 10, TLOD > NLOD. Validation was done by manually checking each indel in IGV. Related to Figure 3.