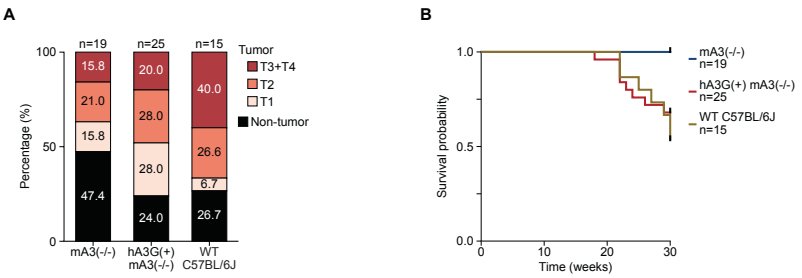
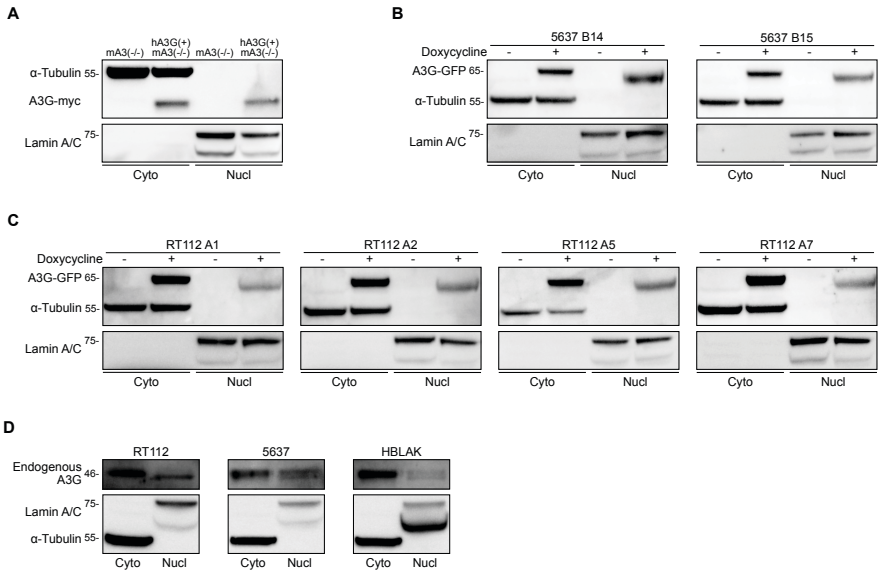
Supplementary Figure S1



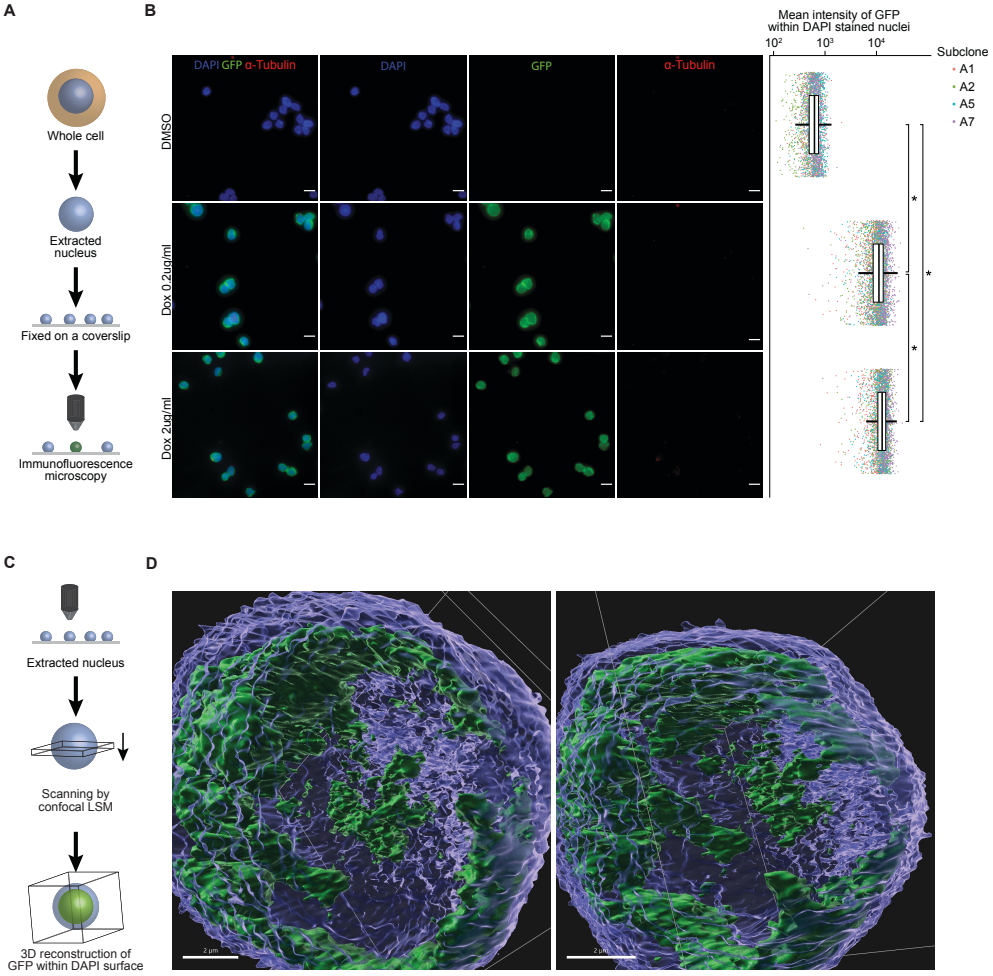
Supplementary Figure S1 | APOBEC3G contributes to carcinogenesis in a murine bladder cancer model. A) Composite bar chart representing the percentage of tumor stages in the *hA3G*(+) *mA3*(-/-), the *mA3*(-/-), and the C57BL/6J mice. The non-tumor category includes benign tissue, hyperplasia, and dysplasia. B) Survival curves for the *hA3G*(+) *mA3*(-/-), the *mA3*(-/-), and the C57BL/6J mice. The *hA3G*(+) *mA3*(-/-) mice had a similar survival curve to the C57BL/6J mice. Log-rank test. All dead mice had pathologically confirmed bladder cancer. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*. WT: wild-type.

Supplementary Figure S2



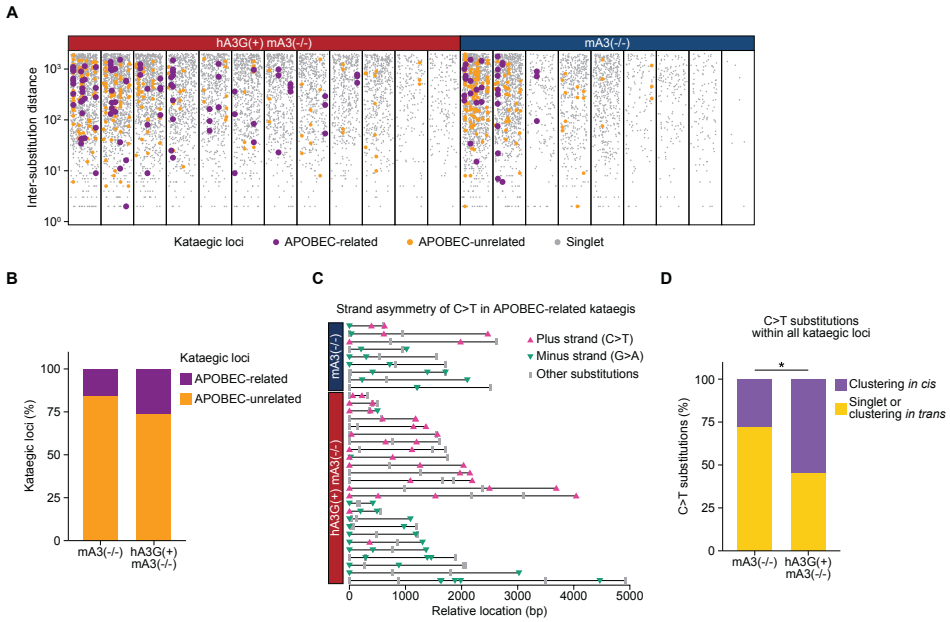
Supplementary Figure S2| APOBEC3G is present in the nuclear compartment. A) SDS-PAGE western blot for the murine tumor organoids derived from a *hA3G*(+) *mA3*(-/-) tumor and a *mA3*(-/-) tumor. B) SDS-PAGE western blot for single-cell clones from the human urothelial bladder tumor cell line 5637 with doxycycline-inducible GFP-tagged APOBEC3G. C) SDS-PAGE western blot for single-cell clones from the human urothelial bladder cancer cell line RT112 with doxycycline-inducible GFP-tagged APOBEC3G. D) SDS-PAGE western blot for two wild-type human bladder cancer cell lines (RT112 and 5637) and one immortalized human bladder epithelial cell (HBLAK). A3G-myc: APOBEC3G tagged with myc tag. A3G-GFP: APOBEC3G tagged with GFP. A3G: APOBEC3G. Cyto: Cytoplasm. Nucl: Nucleus. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3.*

Supplementary Figure S3



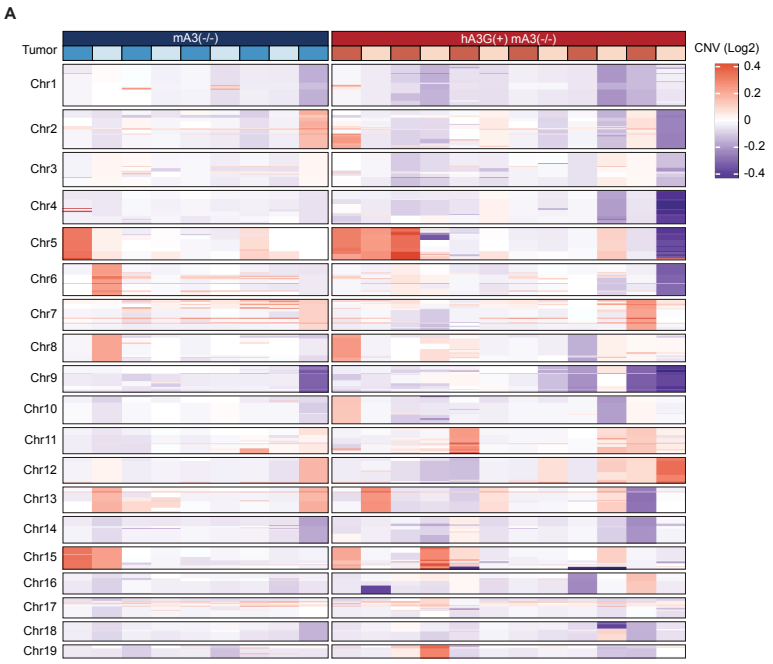
Supplementary Figure S3| APOBEC3G is capable of entering the nuclear compartment. A) Experimental schema for immunofluorescence imaging of extracted nuclei. B) Representative images of extracted nuclei from human urothelial bladder cancer cell line RT112 with integrated doxycycline-inducible GFP-tagged APOBEC3G vector. The extracted nuclei were stained with DAPI, and α-Tubulin was used as a cytoplasmic marker control. The quantitative result of the GFP signal within the nucleus is presented in a box plot as the median with IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. The dot indicates an individual nucleus. Multiple comparisons were performed. *: P<0.05. C) Experimental schema of confocal laser scanning microscopy and 3D reconstruction. D) A representative image from the 3D reconstruction of an extracted nucleus. The violet surface indicated the nuclear surface reconstructed from the DAPI staining. The green signal indicated the GFP signal within the reconstructed DAPI surface. IQR: interquartile range.

Supplementary Figure S4



Supplementary Figure S4 | APOBEC3G induces kataegic events *in vivo* identified by SeqKat. A) Rainfall plot of kataegic loci in the *hA3G(+) mA3(-/-)* and the *mA3(-/-)* tumors. The kataegic loci were identified by SeqKat. Vertical lines represent individual tumors. Grey dots represent singlet substitutions. Orange dots indicate the substitutions within the APOBEC-unrelated kataegic loci. Purple dots indicate substitutions within the APOBEC-related kataegic loci with a significant number of C>T or G>A in kataegic loci calculated by the binormal test (**Methods**). B) Bar chart representing the number of kataegic loci indicating that APOBEC-related kataegic loci were enriched in the *hA3G*(+) *mA3*(-/-) tumors. 25 out of 95 in the *hA3G*(+) *mA3*(-/-) tumors and 9 out of 56 in the *mA3*(-/-) tumors. Fisher's exact test. *: $P<0.05$. C) Strand asymmetry of C>T substitutions in APOBEC-related kataegis. Each line represents an APOBEC-related kataegic locus in different genotypes. The length of the line indicates the relative distance between substitutions. The triangles with different directions and colors indicated the strandedness of C>T substitutions. D) Bar chart representing the number of C>T substitutions occurring *in cis* and that occurring *in trans* and singlet within all kataegic loci, indicating the C>T substitutions occurring *in cis* were enriched in the *hA3G*(+) *mA3*(-/-) tumors. 60 out of 109 in the *hA3G*(+) *mA3*(-/-) tumors and 19 out of 65 in the *mA3*(-/-) tumors. Fisher's exact test. *: $P<0.05$. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3.*

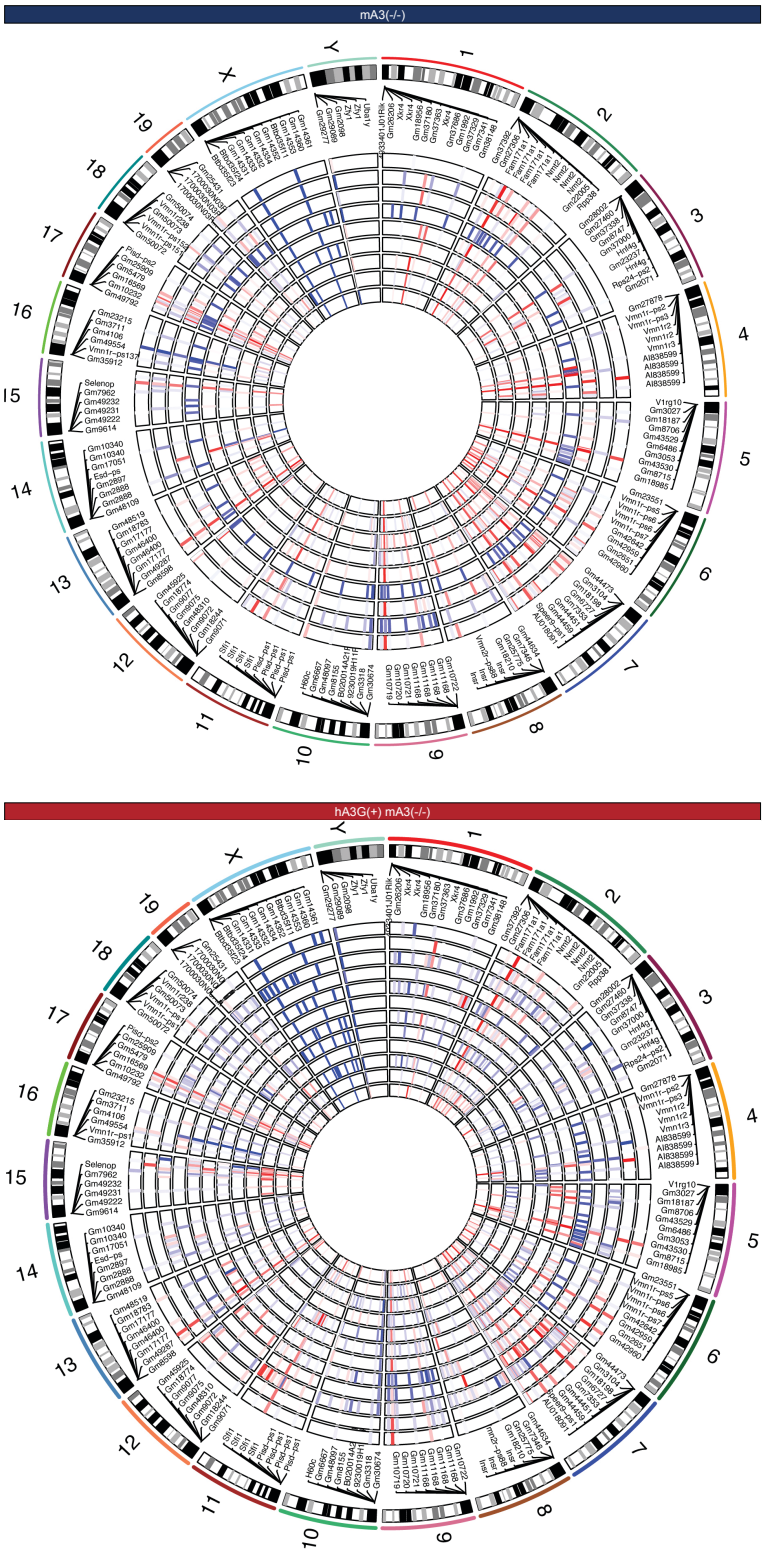
Supplementary Figure S5



Supplementary Figure S5 | Heatmap of copy number variants in bladder tumors from the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) mice. Each tumor is indicated by one vertical lane. The heat scale indicates the corrected copy number values for each segment. Chr: Chromosome. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*. CNV: copy number variant.

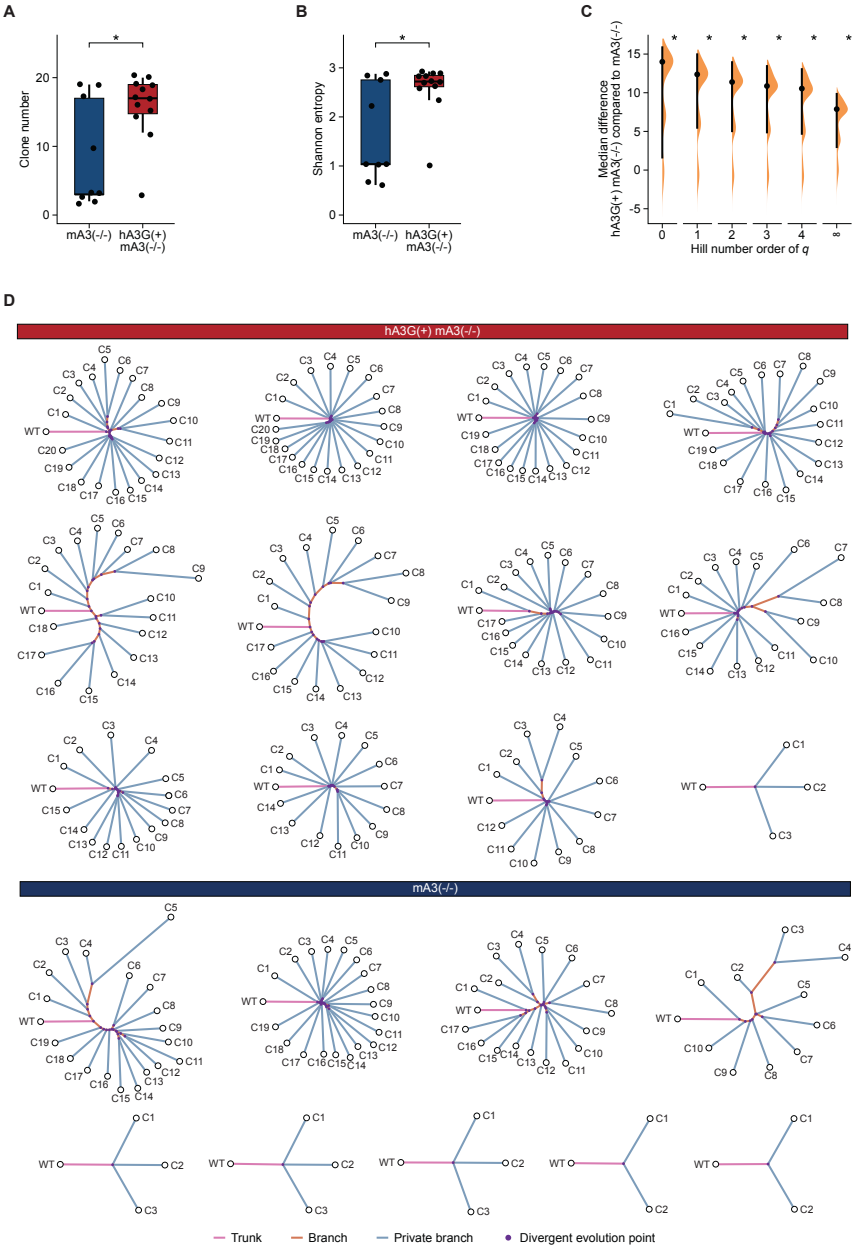
Supplementary Figure S6

A



Supplementary Figure S6 | Circos plots of CNVs in bladder tumors from the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) mice. Concentric circles represent individual tumors. CNV gains are represented in red and CNV losses are represented in blue. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*. CNV: copy number variant.

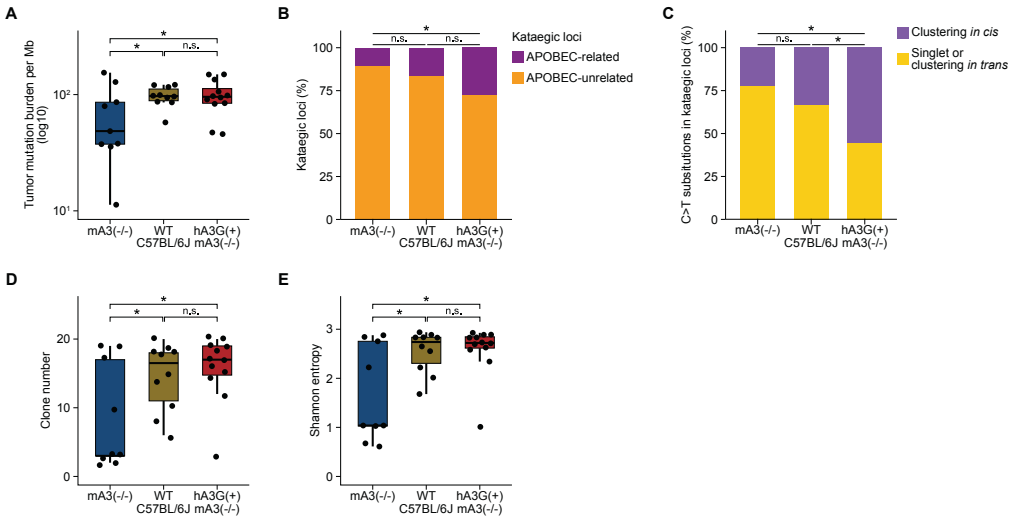
Supplementary Figure S7



Supplementary Figure S7 | APOBEC3G increases intra-tumoral heterogeneity. The figure represents analysis from the EXPANDS tool. A) The *hA3G*(+) *mA3*(-/-) tumors harbor a higher number of clones compared to the *mA3*(-/-) tumors. Mann-Whitney U test. *: *P*<0.05. Boxplots show the median and IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. Individual dots indicate individual tumors. B) The *hA3G*(+) *mA3*(-/-) tumors displayed higher Shannon entropy compared to the *mA3*(-/-) tumors. Mann-Whitney U test. *: *P*<0.05. Boxplots show the median and IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. Individual dots indicate individual tumors. C) The median differences in Hill number orders between the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) tumors are shown by the Gardner-Altman

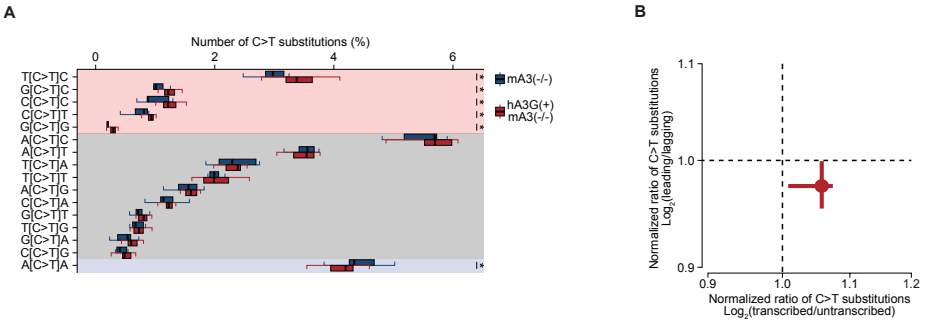
estimation plot. The mean difference is plotted as a bootstrap sampling distribution and depicted as a dot with a 90% confidence interval indicated by the ends of the vertical error bar. D) Phylogenetic trees representing the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) tumors. 'WT' represents the inferred wild-type genome. Branch length corresponds to the proportion of the number of shared variants. The length of the branches in different tumors is normalized to the same scale. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*. IQR: interquartile range.

Supplementary Figure S8



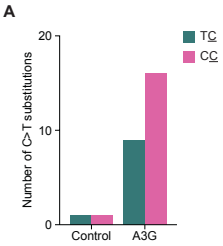
Supplementary Figure S8 | Tumor mutational burden in tumors from the *mA3* (-/-), the *hA3G* (+) *mA3* (-/-), and the C57BL/6J mice. A) Tumors from the C57BL/6J mice harbor a higher mutational burden than those from the *mA3*(-/-) mice. Mann-Whitney U test. The data is shown in the box plot as the median with IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. Each dot represents one tumor. B) Bar chart representing the number of kataegic loci indicating that there is no significant difference in APOBEC-related kataegic loci between the tumors from the C57BL/6J and the *mA3*(-/-) mice. 20 out of 73 in tumors from the *hA3G*(+) *mA3*(-/-) mice, 6 out of 37 in tumors from the C57BL/6J mice, and 4 out of 37 in tumors from the *mA3*(-/-) mice. Fisher's exact test. C) Bar chart representing the number of C>T substitutions occurring *in cis* and that occurring *in trans* or singlet within all kataegic loci, indicating there is no difference of C>T substitutions occurring *in cis* between tumors from the C57BL/6J and the *mA3*(-/-) mice. 45 out of 81 in tumors from the *hA3G*(+) *mA3*(-/-) mice, 12 out of 36 in tumors from the C57BL/6J mice, and 9 out of 40 in tumors from the *mA3*(-/-) mice. Fisher's exact test. D) Tumors from the C57BL/6J mice harbor a higher number of clones than those from the *mA3*(-/-) mice. The clone number is predicted by EXPANDS. Mann-Whitney U test. Boxplots show the median and IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. Individual dots indicate individual tumors. E) Tumors from the C57BL/6J mice displayed higher Shannon entropy than those from the *mA3*(-/-) mice. Mann-Whitney U test. Boxplots show the median and IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. Individual dots indicate individual tumors. *: P<0.05. n.s.: not significant. IQR: interquartile range. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*, WT: wild-type.

Supplementary Figure S9



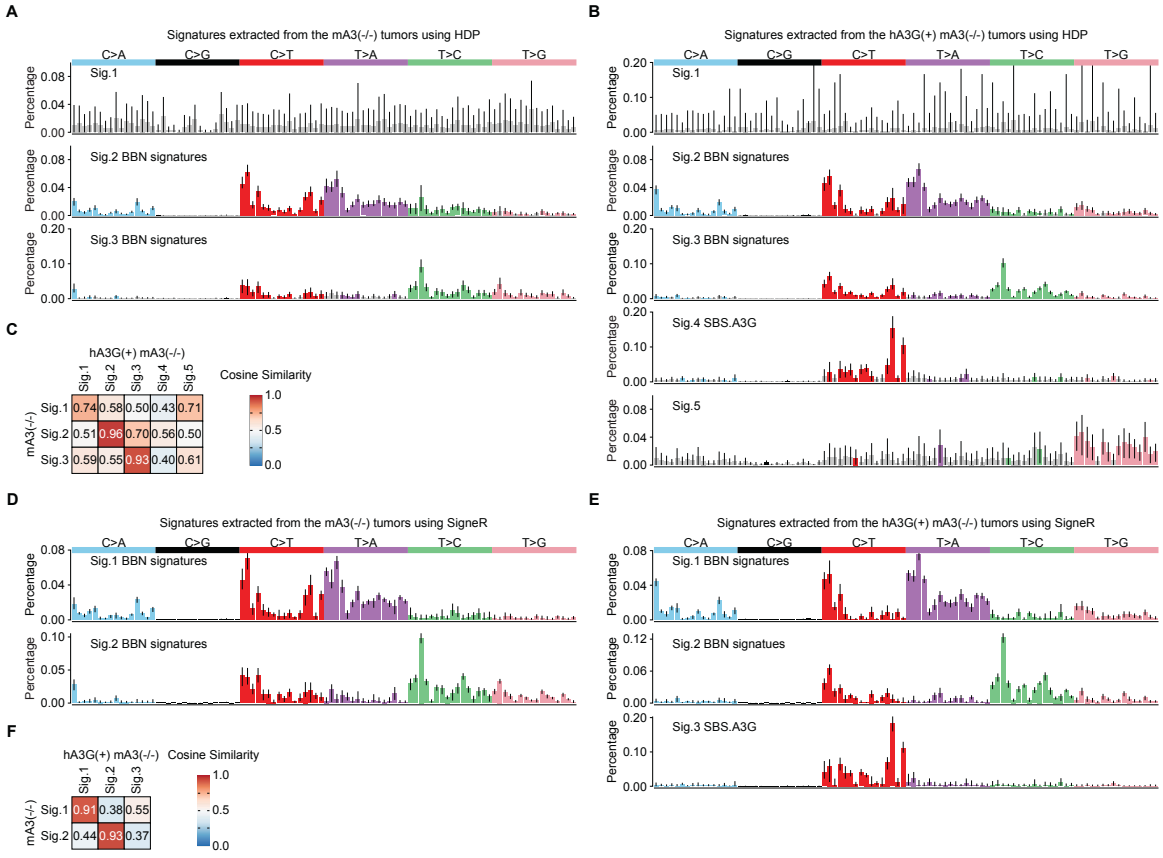
Supplementary Figure S9 | APOBEC3G generates a distinct *in vivo* mutational signature. A) The percentages of trinucleotide motifs harboring C>T substitutions of the total substitutions were compared between tumors from the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) mice. Mann-Whitney U test. *: $P<0.05$. The light red background highlights the five trinucleotide motifs that were significantly increased, and the light blue background highlights the motif that was significantly decreased in the *hA3G*(+) *mA3*(-/-) tumors compared to the *mA3*(-/-) tumors. The light grey highlights motifs with no significant differences between tumors from the two genotypes. Boxplots show the median and IQR. The lower whisker indicates Q1-1.5*IQR. The upper whisker indicates Q3+1.5*IQR. B) Strand asymmetry of APOBEC3G introduced C>T substitutions. The ratio was calculated by dividing the number of C>T substitutions in the leading or transcribed strand by the number of C>T substitutions in the lagging or untranscribed strand. Then, the ratio of the *hA3G*(+) *mA3*(-/-) tumors was normalized to the mean ratio of the *mA3*(-/-) tumors. The dot indicates the median normalized ratio of transcriptional and replicative strand asymmetry in the *hA3G*(+) *mA3*(-/-) tumors. The line represents the interquartile range.

Supplementary Figure S10



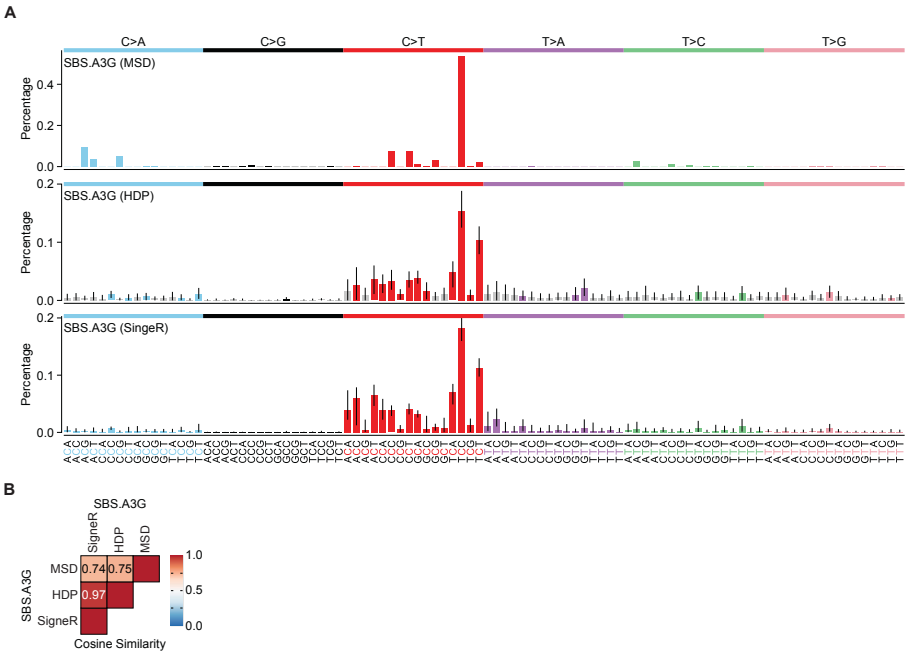
Supplementary Figure S10 | C>T substitutions in the *pol* gene of HIV. The number of C>T substitutions in the *pol* gene of HIV.

Supplementary Figure S11



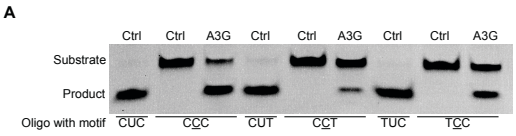
Supplementary Figure S11 | The signatures extracted from the *mA3*(-/-) and the *hA3G*(+) *mA3*(-/-) tumors using HDP and SigneR methods. The signatures were extracted using HDP method from the *mA3*(-/-) (A) and the *hA3G*(+) *mA3*(-/-) (B) tumors. C) The cosine similarity between the signatures. The signatures were extracted using SigneR method from the *mA3*(-/-) (D) and the *hA3G*(+) *mA3*(-/-) (E) tumors. F) The cosine similarity between the signatures. Sig: signature, *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*.

Supplementary Figure S12



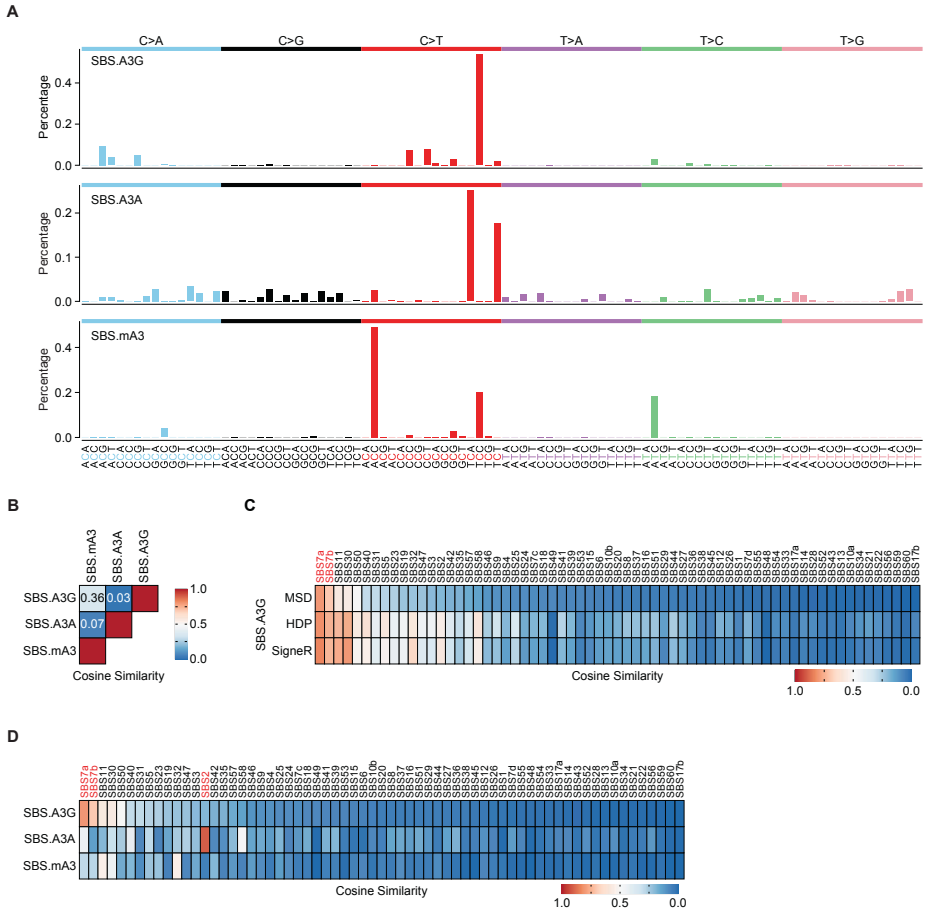
Supplementary Figure S12 | Validation of signature induced by APOBEC3G. A) The single base substitution signature induced by human APOBEC3G (SBS.A3G) was extracted using statistical frameworks based on the mutational shift distance (MSD), hierarchical Dirichlet process (HDP), and Bayesian treated non-negative matrix factorization (NMF) model (SigneR) (**Methods**). B) SBS.A3G extracted using three different tools have high cosine similarity. *hA3G*: human *APOBEC3G*, SBS: single base substitution.

Supplementary Figure S13



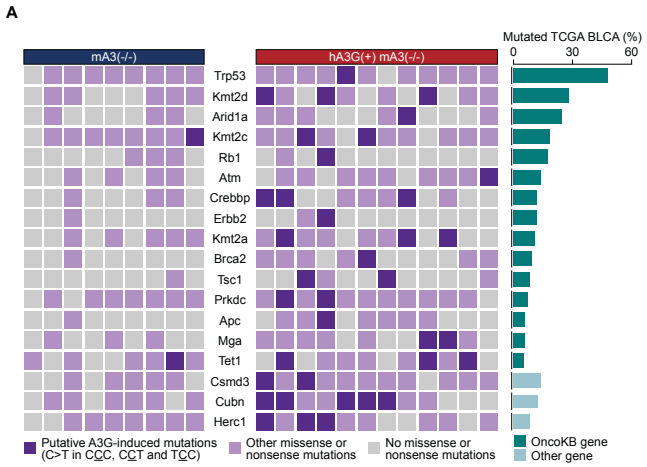
Supplementary Figure S13 | APOBEC3G deaminates cytidines in the C<u>C</u>C, C<u>C</u>T, and T<u>C</u>C motifs in the deamination assay. Ctrl: control oligonucleotide harboring U in the middle of the trinucleotide motif or target oligonucleotide without purified APOBEC3G protein. Oligo: oligonucleotide. A3G: purified APOBEC3G protein.

Supplementary Figure S14



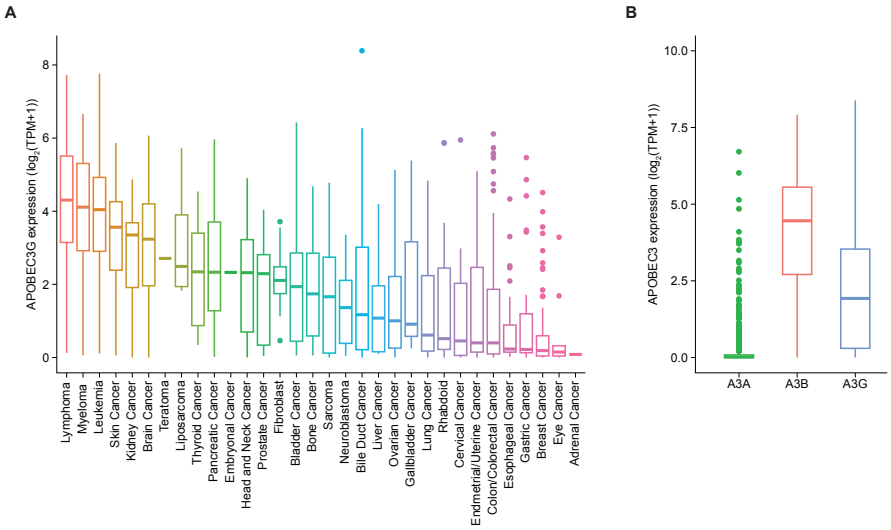
Supplementary Figure S14 | Distinct mutational signatures induced by APOBEC3A, APOBEC3G, and mouse Apobec3. A) The single base substitution signature induced by transgenic expression of human APOBEC3G (SBS.A3G), transgenic expression of human APOBEC3A (SBS.A3A), and mouse Apobec3 (SBS.mA3). B) SBS.A3G has low cosine similarity to SBS.A3A and SBS.mA3. C) Cosine similarity of the SBS.A3G extracted from experimental data using different methods with known COSMIC PCAWG single base substitution signatures. D) Cosine similarity of experimentally derived and COSMIC PCAWG single base substitution signatures. SBS.A3G has a low cosine similarity to SBS2 and SBS13. SBS.A3A had a high cosine similarity with SBS2. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*, SBS: single base substitution, COSMIC: Catalogue Of Somatic Mutations in Cancer, PCAWG: the Pan-Cancer Analysis of Whole Genomes. IQR: interquartile range.

Supplementary Figure S15



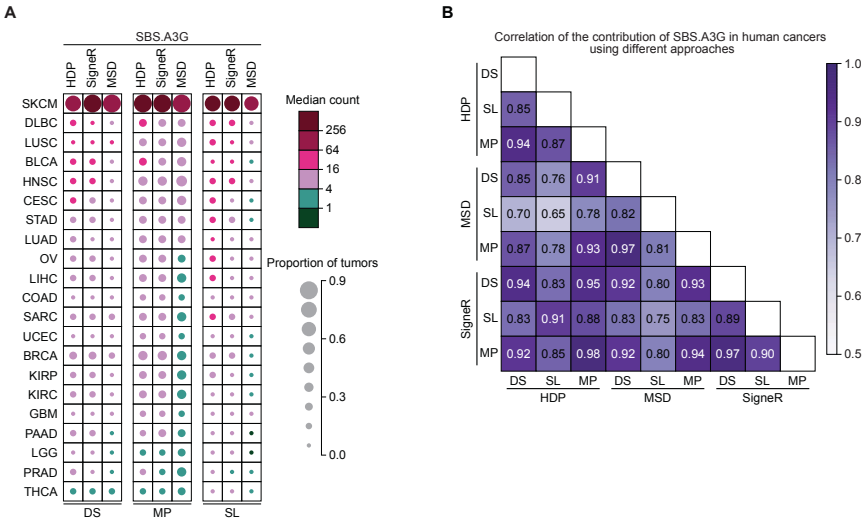
Supplementary Figure S15 | Recurrently mutated genes in bladder tumors from the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) mice and the corresponding frequency in human bladder cancer. A co-mut plot of selected mutated genes in bladder tumors from the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) mice. Dark purple squares represent putative APOBEC3G-induced missense or nonsense mutations (C>T in the CCC, CCT, and TCC motifs). Light purple squares represent other missense or nonsense mutations, which are not putative APOBEC3G-induced. Horizontal bar plots indicate the corresponding frequency in TCGA bladder cancer cohort. Dark green indicates the cancer genes, and light green indicates the other genes based on OncoKB. Due to the limited sample size, we did not undertake statistical analysis to compare the frequency of mutated genes between the *hA3G*(+) *mA3*(-/-) and the *mA3*(-/-) tumors. *hA3G*: human *APOBEC3G*, *mA3*: mouse *Apobec3*.

Supplementary Figure S16



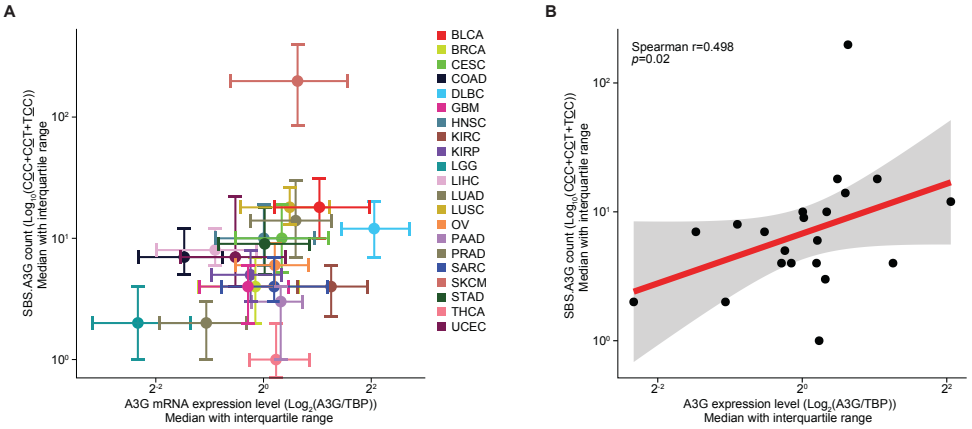
Supplementary Figure S16| ***APOBEC3A***, ***APOBEC3B***, **and** ***APOBEC3G*** **mRNA expression in CCLE cancer cell lines.** A) *APOBEC3G* mRNA expression level in different cancer types. B) *APOBEC3A*, *APOBEC3B*, and *APOBEC3G* mRNA expression in tumor cell lines. Data were obtained from CCLE. A3A: *APOBEC3A*, A3B: *APOBEC3B*, A3G: *APOBEC3G*.

Supplementary Figure S17



Supplementary Figure S17 | The SBS.A3G signature contributes to the mutational load in human cancers. A) The contribution of human APOBEC3G-induced mutagenesis to human cancers is predicted by different methods. The area of each circle represents the proportion of tumors with SBS.A3G contribution for each cancer type. The circle's color represents the median count of SBS.A3G mutations in each cancer type. B) Correlation of the mutational contribution attributed to SBS.A3G in human tumors from TCGA cohorts. The contribution in each tumor was predicted by three pipelines (DS, MP, and SL) to fit SBS.A3G extracted by three approaches (HDP, MSD, and SigneR) to each tumor with other COSMIC signatures. MSD: a statistical framework based on the mutational shift distance to extract mutational signatures. HDP: a framework to extract the signature based on the hierarchical Dirichlet process. SigneR: a framework to extract the signature based on the Bayesian treated non-negative matrix factorization (NMF) model. DS: deconstructSigs. MP: MutationalPatterns. SL: sigLASSO. Abbreviations for TCGA cancer types are available at https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations.

Supplementary Figure S18



Supplementary Figure S18 | The correlation between *APOBEC3G* mRNA and SBS.A3G mutational loads in TCGA pan-cancer cohorts. A) Correlation between *APOBEC3G* mRNA expression and SBS.A3G mutational loads in TCGA pan-cancer cohorts. Each dot represents a cancer type based on the median *APOBEC3G* mRNA expression level (normalized to *TBP*) (horizontal) and the median C>T counts in the C<u>C</u>C, C<u>C</u>T, and T<u>C</u>C motifs (vertical), with error bars representing IQR. B) A nonparametric Spearman correlation analysis was performed between *APOBEC3G* mRNA expression and SBS.A3G mutational loads in TCGA pan-cancer cohorts. Each dot represents a cancer type based on the median *APOBEC3G* mRNA expression level (normalized to *TBP*) (horizontal) and the median C>T counts in the C<u>C</u>C, C<u>C</u>T, and T<u>C</u>C motifs (vertical). A linear regression line (red line) and a 95% CI (grey) were generated. CI: confidence interval. IQR: interquartile range. Abbreviations for TCGA cancer types are available at https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations.