

APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response

Alexander P. Glaser MD, Damiano Fantini PhD, Kalen J. Rimar MD, Joshua J. Meeks MD PhD

APG, DF, KJR, JJM: Northwestern University, Department of Urology, Chicago, IL, 60607

Running title: APOBEC mutagenesis in bladder cancer

***Corresponding author:**

Joshua J. Meeks, MD PhD

303 E. Chicago Ave.

Tarry 16-703

Chicago, IL 60611

Email: joshua.meeks@northwestern.edu

Keywords (4-6):

- Urinary bladder neoplasms
- APOBEC Deaminases
- Mutagenesis
- DNA damage
- Interferon

Abbreviations and Acronyms:

TCGA – The Cancer Genome Atlas

ssDNA – single stranded DNA

APOBEC –apolipoprotein B mRNA editing catalytic polypeptide-like

GCAC – Genome Data Analysis Center

MAF – mutation annotation format

“APOBEC-high” – tumors enriched for APOBEC mutagenesis

“APOBEC-low” – tumors not enriched for APOBEC mutagenesis

Abstract:

Background:

The APOBEC family of enzymes is responsible for a mutation signature characterized by a TCW>T/G mutation. APOBEC-mediated mutagenesis is implicated in a wide variety of tumors, including bladder cancer. In this study, we explore the APOBEC mutational signature in bladder cancer and the relationship with specific mutations, molecular subtype, gene expression, and survival. We hypothesized that tumors with high levels of APOBEC-mediated mutagenesis would be enriched for mutations in DNA damage response genes and associated with higher expression of genes related to activation of the immune system.

Methods:

Gene expression (n=408) and mutational (n=395) data from the Cancer Genome Atlas (TCGA) bladder urothelial carcinoma provisional dataset was utilized for analysis. Tumors were split into “APOBEC-high” and “APOBEC-low” tumors based on APOBEC enrichment score. Analysis was performed with R.

Findings:

Patients with APOBEC-high tumors have better overall survival compared to those with APOBEC-low tumors (38.2 vs 18.5 months, p=0.005). Tumors enriched for APOBEC mutagenesis are more likely to have mutations in DNA damage response genes (*TP53*, *ATR*, *BRCA2*), and chromatin regulatory genes (*MLL*, *MLL3*), while APOBEC-low tumors are more likely to have mutations in *FGFR3* and *KRAS*. *APOBEC3A* and *APOBEC3B* expression correlates with total mutational burden, regardless of bladder tumor molecular subtype. APOBEC mutagenesis and enrichment is associated with increased expression of immune-related genes, including interferon signaling.

Interpretation:

Tumors enriched for APOBEC mutagenesis are more likely to have mutations in DNA damage response genes and chromatin regulatory genes, potentially providing more single-strand DNA substrate for *APOBEC3A* and *APOBEC3B*, leading to a hypermutational phenotype and the subsequent immune response.

Highlights:

- APOBEC enzymes, particularly *APOBEC3A* and *APOBEC3B*, are responsible for the predominant pattern of mutagenesis in bladder cancer
- Tumors enriched for APOBEC-mediated mutagenesis are more likely to have mutations in DNA damage response genes and chromatin regulatory genes, while tumors not enriched for APOBEC-mediated mutagenesis are more likely to have mutations in *KRAS* and *FGFR3*
- APOBEC enrichment is associated with upregulation of genes involved in the immune response

1. Introduction

Urothelial carcinoma has one of the highest mutation rates of any sequenced cancer to date ¹. High-throughput next generation sequencing analyses such as The Cancer Genome Atlas (TCGA) and others have identified a mutational signature characterized by a TCW>T/C mutation thought to be attributable to the apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) family of enzymes ¹⁻³. This mutational pattern is the predominant pattern in bladder cancer and is also frequently found in breast, cervical, head and neck, and lung cancers ^{1,3,4}.

The APOBEC family consists of 11 members, including *AID*, *APOBEC1*, *APOBEC2*, *APOBEC3A*, *APOBEC3B*, *APOBEC3C*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G*, *APOBEC3H*, and *APOBEC4*. These enzymes function as cytosine deaminases and are involved in C>U deamination in single-stranded DNA (ssDNA), and likely function physiologically in antiretroviral defense ⁵⁻⁸. However, in tumor cells, these enzymes are likely responsible for hypermutation at cytosine bases in exposed ssDNA, known as kataegis ⁹. The APOBEC3 family, and particularly *APOBEC3A* and *APOBEC3B* ^{5,10-13} but also *APOBEC3H* ¹⁴, are the predominant APOBEC enzymes theorized to contribute to cancer mutagenesis.

Several studies have linked *APOBEC3B* expression with mutagenesis ^{13,15}, but expression alone does not fully explain this mutational signature, and *APOBEC3A* may also play a significant role ¹⁰. In breast cancer, DNA replication stress and mutations in DNA repair genes have been linked to APOBEC-mediated mutagenesis ¹⁶, potentially due to increased availability of ssDNA substrate for enzymatic deamination ^{17,18}. However, less is known about the downstream effects of APOBEC mutagenesis in bladder cancer.

In this study, we explore the APOBEC mutational signature in bladder cancer and its relationship with specific mutations, molecular subtype, gene expression, and survival. We hypothesized that tumors

with high levels of APOBEC-mediated mutagenesis would be enriched for mutations in DNA damage response genes and express genes related to activation of the immune system at higher levels.

2. Methods:

2.1 *The Cancer Genome Atlas data*

The Cancer Genome Atlas (TCGA) bladder urothelial carcinoma data was downloaded from the Broad Institute Genome Data Analysis Center (GDAC) (<http://gdac.broadinstitute.org>)¹⁹ and from cBioPortal (<http://cbioportal.org>)²⁰. Data from GDAC was downloaded on November 8, 2016, from the analysis timestamp “analyses_2016_01_28” (doi:10.7908/C19G5M58)²¹. Downloaded data includes clinical and demographic data (age, sex, tumor stage, overall survival), mutation annotation files (MutSig 2CV v3.1; MAF file; Mutsig_maf_modified.maf.txt) and mRNA expression (Immumina HiSeq RNAseqV2). RNA-seq mRNA expression levels are presented as RNA-seq by expectation-maximization (RSEM) values²².

Clinical information was available on 412 samples, RNA-seq data was available on 408 samples, and mutation information was available on 395 samples for TCGA BLCA data version 2016_01_28. Overlap between the 412 patients with clinical information, 408 patients with RNA-seq data, and 395 patients with mutation annotation information yields 391 patients. Three outliers were removed from mutational analysis (TCGA-DK-A6AW, >150 mutations/Mb; TCGA-XF-AAN8 and TCGA-FD-A43, both with ≤ 5 total mutations).

2.2 *Mutation analysis and APOBEC enrichment*

Analysis and visualization of mutations from mutation annotation format (MAF) files was performed using R version 3.3.3, Bioconductor²³ version 3.4 (<http://www.Bioconductor.org>), and

MAFtools version 1.0.55²⁴. Mutation rates per sample were calculated using MutSig2CV v3.1 from the Broad Institute GDAC¹. APOBEC enrichment score based on the frequency of TCW>T/G mutations was calculated as previously described^{3,4,25}. Samples were classified into two groups: “APOBEC-high” based on APOBEC enrichment > 2 and Benjamini-Hochberg²⁶ false-discovery-rate corrected p-value < 0.05; and “APOBEC-low” based on an APOBEC enrichment < 2 and/or Benjamini-Hochberg false-discovery-rate corrected p-value ≥ 0.05. Survival outcomes between patients with APOBEC-high-enrichment and APOBEC-low-enrichment was performed using log-rank test and Kaplan-Meier curves (R packages survival v 2.41-2, survminer v0.3.1, ggplot2 v2.2.1).

Significantly differentially mutated genes between APOBEC-high-enrichment and APOBEC-low-enrichment groups was performed using MAFtools²⁴ as previously described²⁷ and visualized with oncoplots and forest plots.

2.3 Molecular Subtyping

Molecular subtyping of 408 bladder urothelial carcinoma samples with RNA-seq data was performed using R, Bioconductor, and multiClust version 1.4.0²⁸. Samples were classified as luminal, p53-like, basal, or claudin-low as previously described²⁹ with hierarchical clustering using Euclidean distance and Ward’s linkage method (ward.D2; hierarchical clustering analysis shown in **Supplementary Figure 1**). Differences in mutational load and expression of APOBEC3 enzymes (RSEM) between tumor subtypes was compared using ANOVA.

2.4 Differential gene expression and gene expression associated with APOBEC enrichment

Differential expression analysis between APOBEC-high and APOBEC-low tumors was performed using R, Bioconductor, limma version 3.30.13 and edgeR version 3.16.5^{30,31}. Association of gene expression with numeric APOBEC enrichment score was performed using Spearman’s rho. Functional

annotation of genes was performed with DAVID version 6.8 (<http://david.ncifcrf.gov>)^{32,33} and visualized with Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>)^{34,35}.

3. Results:

3.1. APOBEC mutagenesis in bladder cancer

In bladder cancer, as the mutation load increases, the frequency of specific nucleotide conversions changes, and C>G mutations become more common (**Figure 1A**). Many of these mutations are a specific contextual TCW>T/G mutation attributed to the APOBEC family of enzymes. Of 389 tumors in the provisional TCGA bladder urothelial carcinoma dataset, 324 are enriched for APOBEC mutagenesis (“APOBEC-high”) vs. 64 with low or no enrichment (“APOBEC-low”). APOBEC-high tumors have improved overall survival compared to APOBEC-low tumors (median overall survival 38.2 vs 18.5 months, $p=0.0050$, **Figure 1B**).

APOBEC-low tumors are more likely to be low-grade (17% vs 3%, $p<0.0001$), but there is no significant difference in age at diagnosis, tumor stage, tumor subtype, or patient smoking history category between groups (**Supplementary Table 1**). Interestingly, a higher frequency of Asian patients was noted in the APOBEC-low group (26% vs 7%), and *APOBEC3B* was expressed at a lower level in Asian patients vs. patients not of Asian ethnicity ($p<0.0001$, **Supplementary Figure 2**).

APOBEC-high tumors have a higher number of variants per sample and a higher proportion of C>T and C>G mutations (**Figure 1C-D**). Despite an association of APOBEC enrichment score with total mutations, 42% of APOBEC-high tumors have a mutational burden below the median (**Supplementary Figure 3**), and APOBEC-high and APOBEC-low tumors share many of the commonly mutated genes in bladder cancer (**Figure 1E**).

3.2. Frequency of mutations in APOBEC-high and APOBEC-low tumors

To determine what mutations were associated with APOBEC mutagenesis, we next compared significantly mutated genes between APOBEC-high and APOBEC-low tumors. After correction for multiple comparisons, APOBEC-high tumors were more likely to have mutations in *TP53*, *PIK3CA*, *ATR*, *BRCA2*, *MLL*, *MLL3*, and *ARID1A*, among others; while APOBEC-low tumors were more likely to have mutations in *KRAS* and *FGFR3* (**Figure 2A-B**; complete list of differentially mutated genes in **Supplementary Table 2**).

Functional annotation of differentially mutated genes demonstrates that APOBEC-high tumors are enriched for mutations in DNA damage repair genes (e.g. *TP53*, *ATR*, *BRCA2*, *POLQ*) and chromatin modification genes including *MLL*, *MLL3*, *ARID1A*, *NCOR1*, *BPTF*, *CHD7*, and others. A full list of gene ontology terms associated with genes mutated in APOBEC-high tumors can be found in **Supplementary Table 3**. APOBEC-low tumors were significantly more likely to harbor mutations in *KRAS* and *FGFR3*. As in the original TCGA dataset (n=131)³⁶, non-synonymous mutations in *KRAS* and *FGFR3* are mutually exclusive (**Figure 2C**).

3.3. Expression of APOBEC3A and APOBEC3B correlate with mutational burden

We next investigated the role of expression of APOBEC3 enzymes with total mutations in both the entire TCGA bladder cancer dataset and in the four molecular subtypes (luminal, p53-like, basal, and claudin-low). Expression of *APOBEC3A* and *APOBEC3B* correlate with the total mutation burden in bladder cancer (**Figure 3A-B**), while expression of *APOBEC3F* and *APOBEC3G* correlate weakly with total mutations and *APOBEC3C*, *APOBEC3D*, and *APOBEC3H* do not correlate with total mutations (**Supplementary Figure 4**).

Total mutations, APOBEC enrichment score, and the percentage of tumors classified as APOBEC-high vs APOBEC-low are not different between subtypes of bladder cancer (**Supplementary Figure 5**).

However, *APOBEC3A* is expressed at a significantly higher level in the basal subtype than in luminal, p53-like, or claudin-low subtypes (**Figure 3C**), while *APOBEC3B* is evenly expressed across subtypes (**Figure 3D**). Despite this, *APOBEC3A* expression levels correlate with total mutations in every subtype, as do *APOBEC3B* expression levels (**Supplementary Figure 6**).

3.4. Gene expression associated with APOBEC enrichment

We approached gene expression association with APOBEC enrichment by two methods. We first analyzed differentially expressed genes between APOBEC-high and APOBEC-low tumors. APOBEC-high tumors were enriched for expression of genes related to regulation of the immune response and lymphocyte-mediated immunity (**Figure 4A and Supplementary Table 4**), whereas APOBEC-low tumors with higher expression of genes related to transcription and translation (**Figure 4B and Supplementary Table 5**). Similarly, correlation between numeric APOBEC enrichment score and gene expression revealed a positive relationship between APOBEC enrichment score and gene families involved in IFN-gamma signaling, antigen presentation, and regulation of the immune response, including immune checkpoint HAVCR2 (also known as TIM-3; $r=0.229$, $p<0.0001$) (**Figure 4C and Supplementary Table 6**), while genes inversely correlated with APOBEC enrichment score were enriched in gene families related to transcription and translation (**Figure 4D and Supplementary Table 7**).

4. Discussion:

APOBEC mutagenesis is the predominant mutational pattern in bladder cancer. In this paper, we demonstrate that tumors enriched for APOBEC mutagenesis (APOBEC-high tumors) have better survival and are more likely to have mutations in DNA damage repair genes and chromatin regulation genes. Expression of *APOBEC3A* and *APOBEC3B* correlates with overall mutation load in bladder cancer, regardless of molecular subtype. In addition, APOBEC enrichment is associated with upregulation of

immune-related genes including interferon signaling. Unexpectedly, we also demonstrate that tumors not featuring the APOBEC mutational pattern (APOBEC-low tumors) are significantly more likely to harbor mutations in *FGFR3* and *KRAS*, which are mutually exclusive.

Our work is consistent with several prior studies linking APOBEC expression to mutational burden. *APOBEC3B* expression is upregulated in breast cancer and is associated with overall mutations¹⁵. Our group also demonstrated the relationship between APOBEC3B expression and mutations in bladder cancer³⁷. Overexpression of *APOBEC3B* in breast cancer cell lines results in DNA fragmentation, increased C>T mutations, delayed cell cycle arrest, and eventual cell death¹⁵. Furthermore, knockdown of *APOBEC3B* with short hairpin RNA in breast cancer cell lines decreases total number of uracil lesions, *TP53* mutations, and C>T mutations¹⁵.

APOBEC3A expression was not initially detectable in breast cancer cell lines¹⁵, and *APOBEC3B* expression correlates strongly with overall mutations in multiple malignancies^{13,15}, leading many to believe that APOBEC3B is responsible for the majority of these mutations. However, *APOBEC3A* expression is also correlated with mutational burden³, as we demonstrate again here, *APOBEC3A* is highly proficient at cytidine hypermutation and creation of DNA double-strand breaks^{38,39}, and may have a larger role in mutagenesis than previously recognized¹⁰.

Expression alone of APOBEC enzymes is insufficient to explain APOBEC-mediated mutagenesis. In this study, we demonstrate that although the basal subtype of bladder cancer has significantly higher expression of *APOBEC3A* than other subtypes, basal tumors do not have greater number of mutations or APOBEC enrichment compared to the other subtypes, and *APOBEC3A* expression correlates with total mutation burden even in subtypes with lower levels of *APOBEC3A* expression. This suggests that a baseline level of APOBEC expression is required for APOBEC mutagenesis, but above a certain threshold other factors are also influential⁴. Post-translational modification regulating APOBEC enzymatic activity

may play a role⁴⁰. Alternatively, as with any enzymatic reaction, increasing substrate availability may move equilibrium toward the accumulation of mutated DNA products.

The proposed substrate for APOBEC mutagenesis is ssDNA, a common DNA repair intermediate that may accumulate in cells with defects in DNA repair pathways⁴¹. Interestingly, APOBEC-high tumors are more likely to have mutations in genes related to DNA repair and chromatin regulation. In breast cancer, tumor cell lines with high levels of *APOBEC3B* expression are more likely to have mutations in *TP53*¹⁵. Furthermore, APOBEC-high breast cancers are more likely to have mutations in *TP53*, *NCOR1*, *MLL3* and other genes involved in DNA replication stress²⁵. In this study, we demonstrate that APOBEC-high bladder tumors are more likely to have mutations in *TP53*, *NCOR1*, *MLL3* (*KMT2C*), *MLL* (*KMT2A*), *ATR*, *BRCA2*, and other genes related to DNA repair and chromatin modification.

We also demonstrated a higher frequency of *PIK3CA* mutations in APOBEC-high bladder tumors. *PIK3CA* has been previously reported to be mutated at a high frequency in specific TCW-containing helical motifs across a number of tumor types⁴². Our analysis supports these results, with the majority of *PIK3CA* mutations in APOBEC-high tumors occurring in the helical domain at E542 and E545; no mutations in the helical domain were seen in APOBEC-low tumors (**Supplementary Figure 7**). These specific mutations in E542 and E545 have also been reported as hotspot mutations in breast cancer⁴³.

Interestingly, APOBEC-low tumors in this study were more likely to have mutations in the oncogenes *FGFR3* and *KRAS*, which are mutually exclusive. This suggests that tumors not enriched for the APOBEC mutational pattern may be driven by oncogenes which may dysregulate cellular homeostasis via mechanisms that do not result in accumulation of ssDNA intermediates used as substrate for APOBEC mutagenesis.

APOBEC enrichment was associated with overall survival and expression of immune related genes. *APOBEC3A* expression measured by Nanostring has previously been associated PD-L1 expression on

tumor-infiltrating mononuclear cells in bladder cancer⁴⁴. APOBEC mutagenesis is associated with overall mutational burden in bladder cancer, which likely reflects downstream neoantigen burden and subsequent immune response⁴⁵⁻⁴⁷.

Based on our results and the above discussion, we propose a working model of mutagenesis and the immune response in bladder cancer (**Figure 5**), in which a urothelial cell acquires one or more driver mutation(s). Accumulation of mutations in *TP53*, *ATR*, *BRCA2*, and/or other DNA damage response genes or chromatin regulation genes may result in the accumulation of ssDNA substrate for to *APOBEC3A* and *APOBEC3B*, leading to a high level of APOBEC-mediated mutagenesis and a hypermutation phenotype. This hypermutation in turn leads to a large neoantigen burden and the subsequent immune response generated from this increase in neoantigens. In addition, APOBEC enzymes, especially *APOBEC3A*, may be induced and overexpressed in response to interferon^{4,8,39,48}, potentially causing a positive feedback loop. In contrast, other tumors with mutations in the *FGFR3*/RAS pathway or other oncogenes may not expose sufficient substrate ssDNA to APOBEC enzymes to undergo significant APOBEC mutagenesis. These tumors have poor survival, despite an enrichment for *FGFR3* mutations and low-grade tumors, which were classically considered more benign phenotypes.

Several limitations of this study warrant mention. We utilized the provisional TCGA dataset, for which mutational and expression data is readily available. Results warrant replication in other datasets. The TCGA does not currently include any systemic treatment-related information. However, mutations in *ERCC2*^{49,50} and other DNA repair genes^{51,52} are associated with response to platinum-based therapy, and further investigation into the role of APOBEC mutagenesis and response to both cytotoxic chemotherapy and immunotherapy is warranted. Another limitation is the lack of a specific gene expression signature observed in APOBEC-low tumors other than transcription- and translation-related genes, potentially due to the heterogeneity of this group. In addition, gene expression correlations with

APOBEC enrichment score in APOBEC-low tumors would not be expected to generate a strong signal, as these tumors by definition have a low and heterogeneous numerical APOBEC enrichment score.

In summary, APOBEC enzymes are a major source of mutation in bladder cancer. Tumors enriched for APOBEC mutagenesis have better survival and are more likely to have mutations in DNA damage repair genes and chromatin modifying genes. The APOBEC mutagenesis signature is associated with increased expression of immune-related genes. Bladder tumors not enriched for APOBEC mutagenesis are more likely to have mutations in *KRAS* and *FGFR3*, which are mutually exclusive, and these tumors have poor overall survival. Further study of the regulation of APOBEC enzymes, mutagenesis, and response to subsequent therapy may provide further insight into the mutational landscape and potential therapeutics for bladder cancer.

5. Acknowledgements:

The results published here are in whole or part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. These results were presented in part at the American Urological Association 2017 meeting (Boston, MA; Abstract #17-3454).

6. Funding Sources

JJM is funded by a Veterans Health Administration Merit grant BX0033692-01 and the John P. Hanson Foundation for Cancer and Cellular Research at the Robert H. Lurie Comprehensive Cancer Center of Northwestern University. Funding sources had no role in writing of the manuscript or the decision to submit it for publication.

7. Conflicts of Interest

The authors declare no competing interest.

8. Author Contributions

APG and JJM conceived the project ideas. JJM was responsible for project final design, supervision, interpretation of the data. APG, DF, and KR performed data collection and analysis. APG drafted the manuscript. All authors were involved in critical revision of the article and final approval of the version to be published.

9. References:

1. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-8 (2013).
2. Roberts, S.A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell* **46**, 424-35 (2012).
3. Roberts, S.A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**, 970-6 (2013).
4. Middlebrooks, C.D. *et al.* Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat Genet* (2016).
5. Shi, K. *et al.* Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* (2016).
6. Harris, R.S., Petersen-Mahrt, S.K. & Neuberger, M.S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-53 (2002).
7. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103 (2003).
8. Bonvin, M. *et al.* Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364-74 (2006).
9. Lada, A.G. *et al.* AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biology Direct* **7**, 47 (2012).
10. Chan, K. *et al.* An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* **47**, 1067-72 (2015).
11. Henderson, S. & Fenton, T. APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol Med* **21**, 274-84 (2015).
12. Swanton, C., McGranahan, N., Starrett, G.J. & Harris, R.S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704-12 (2015).
13. Burns, M.B., Temiz, N.A. & Harris, R.S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* **45**, 977-83 (2013).
14. Starrett, G.J. *et al.* The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* **7**, 12918 (2016).
15. Burns, M.B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-70 (2013).
16. Cescon, D.W. & Haibe-Kains, B. DNA replication stress: a source of APOBEC3B expression in breast cancer. *Genome Biol* **17**, 202 (2016).
17. Haradhvala, N.J. *et al.* Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* **164**, 538-49 (2016).
18. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656 (2017).
19. Marx, V. Drilling into big cancer-genome data. *Nat Methods* **10**, 293-7 (2013).

20. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
21. Center, B.I.T.G.D.A. Analysis Overview for Bladder Urothelial Carcinoma (Primary solid tumor cohort) - 28 January 2016. in *Broad Institute of MIT and Harvard* (2016).
22. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
23. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115-21 (2015).
24. Mayakonda, A. & Koeffler, P.H. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *BioRxiv* (2016).
25. Kanu, N. *et al.* DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol* **17**, 185 (2016).
26. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289-300 (1995).
27. Madan, V. *et al.* Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. *Leukemia* **30**, 1672-81 (2016).
28. Lawlor, N., Fabbri, A., Guan, P., George, J. & Karuturi, R.K. multiClust: An R-package for Identifying Biologically Relevant Clusters in Cancer Transcriptome Profiles. *Cancer Inform* **15**, 103-14 (2016).
29. Dadhania, V. *et al.* Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. *EBioMedicine* (2016).
30. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
31. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-97 (2012).
32. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
33. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
34. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
35. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-7 (2016).
36. Network, C.G.A.R. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-22 (2014).
37. Glaser, A.P., Fantini, D., Shlatifard, A., Schaeffer, E.M. & Meeks, J.J. The evolving genomic landscape of urothelial carcinoma. *Nat Rev Urol* **14**, 215-229 (2017).
38. Landry, S., Narvaiza, I., Linfesty, D.C. & Weitzman, M.D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep* **12**, 444-50 (2011).
39. Mussil, B. *et al.* Human APOBEC3A Isoforms Translocate to the Nucleus and Induce DNA Double Strand Breaks Leading to Cell Stress and Death. *PLoS One* **2013**, 8 (2013).
40. Matsumoto, T. *et al.* PKA-Mediated Phosphorylation of APOBEC3B Suppresses Its DNA Mutagenic Potential in Myeloma Cells. *Blood* **128**, 4427-4427 (2016).
41. Zou, L. & Elledge, S.J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542-8 (2003).
42. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, T.R. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep* **7**, 1833-41 (2014).

43. Thirumal Kumar, D. & George Priya Doss, C. Role of E542 and E545 missense mutations of PIK3CA in breast cancer: a comparative computational approach. *J Biomol Struct Dyn*, 1-13 (2016).
44. Mullane, S.A. *et al.* Correlation of Apobec Mrna Expression with overall Survival and pd-l1 Expression in Urothelial Carcinoma. *Sci Rep* **6**, 27702 (2016).
45. Brown, S.D. *et al.* Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res* **24**, 743-50 (2014).
46. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **371**, 2189-99 (2014).
47. Hackl, H., Charoentong, P., Finotello, F. & Trajanoski, Z. Computational genomics tools for dissecting tumour-immune cell interactions. *Nat Rev Genet* **17**, 441-58 (2016).
48. Refsland, E.W. *et al.* Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Research* **38**, 4274-84 (2010).
49. Van Allen, E.M. *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**, 1140-53 (2014).
50. Liu, D. *et al.* Clinical Validation of Chemotherapy Response Biomarker ERCC2 in Muscle-Invasive Urothelial Bladder Carcinoma. *JAMA Oncol* **2**, 1094-6 (2016).
51. Plimack, E.R. *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* **68**, 959-67 (2015).
52. Evans, J.R. *et al.* Patient-Level DNA Damage and Repair Pathway Profiles and Prognosis After Prostatectomy for High-Risk Prostate Cancer. *JAMA Oncol* **2**, 471-80 (2016).

10. Figure Legends

Figure 1. APOBEC-mediated mutagenesis in the TCGA bladder cancer cohort (n=389). (A) Percentage of single nucleotide variations (SNVs) as a function of mutation load. Genomes were binned in groups of 20 samples according to mutation load. (B) Kaplan-Meier survival curve of APOBEC-high and APOBEC-low bladder tumors. (C) Summary of mutagenesis in APOBEC-high tumors, including number of variants per sample, variant classification, and class of SNV. Stacked barplot demonstrates percentage of SNVs in a representative sample of 64 tumors for comparison with APOBEC-low tumors. (D) Summary of mutagenesis in APOBEC-low tumors, including number of variants per sample, variant classification, class of SNV, and stacked barplot of the percentage of SNV per tumor. (E) Oncoplot of the top genes commonly mutated in bladder cancer in APOBEC-high and APOBEC-low tumors.

Figure 2. Significantly differentially mutated genes in APOBEC-high and APOBEC-low tumors. (A) Oncoplot of genes differentially mutated in APOBEC-high and APOBEC-low tumors. (B) Forestplot of differentially mutated genes in APOBEC-high and APOBEC-low tumors with log10 odds ratio and 95% confidence intervals and adjusted p-value (MutSig 2CV 3.1; MAFtools v1.0.55). (C) Oncoprint of nonsynonymous mutations in *FGFR3* and *KRAS* (n=395).

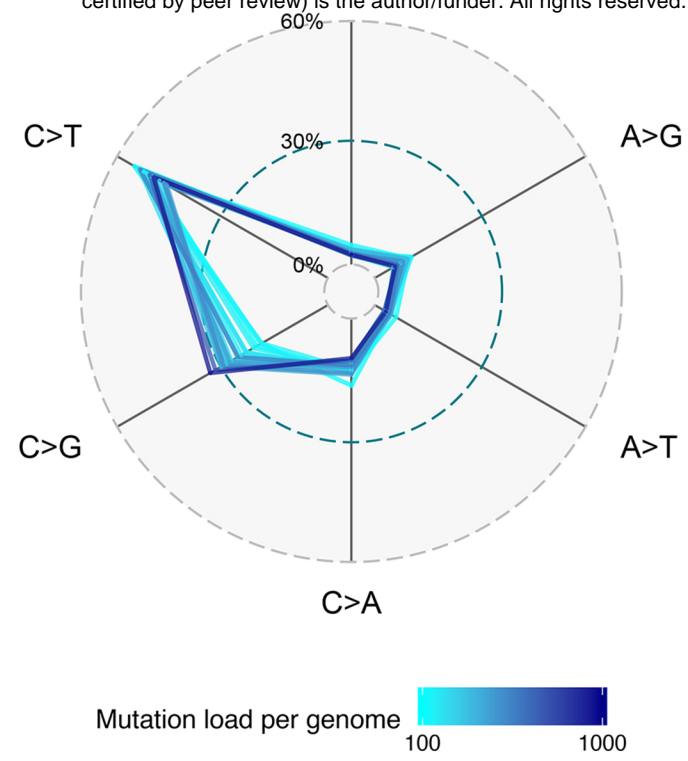
Figure 3. Correlations of *APOBEC3A* and *APOBEC3B* expression with mutational burden and molecular subtype. (A) Spearman correlation between total mutations and *APOBEC3A* expression. (B) Spearman correlation between total mutations and *APOBEC3B* expression. (C) Expression of *APOBEC3A* in the molecular subtypes of bladder cancer. (D) Expression of *APOBEC3B* in the molecular subtypes of bladder cancer.

Figure 4. Barplot of gene ontology biological processes for: (A) genes highly expressed in APOBEC-high tumors (B) genes highly expressed in APOBEC-low tumors (C) genes positively correlated with APOBEC enrichment score (D) genes negatively correlated with APOBEC enrichment score. Figure generated with Enrichr.^{34,35} Bar size based on combined score based on p-value and deviation from expected rank.

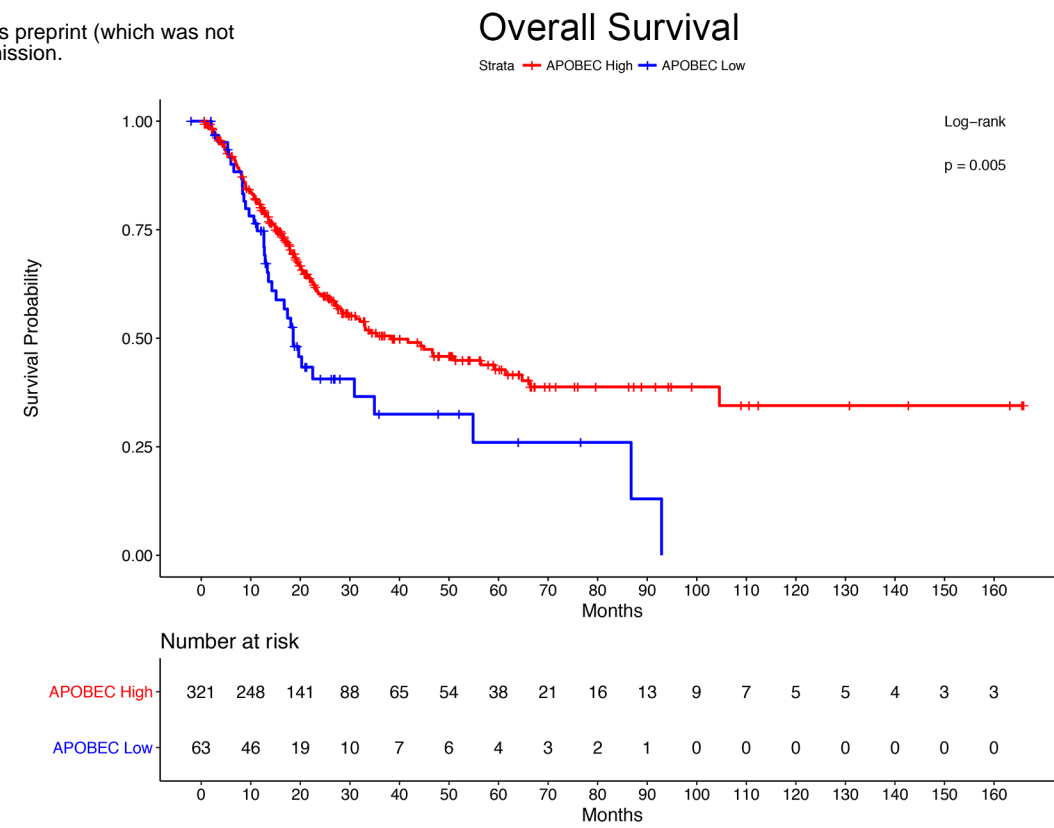
Figure 5. Working model of APOBEC-mediated mutagenesis in bladder cancer. Accumulation of mutations in *TP53*, *ATR*, *BRCA2*, and/or other DNA damage response genes or chromatin regulation genes may expose more substrate ssDNA to *APOBEC3A* and *APOBEC3B*, leading to a high level of APOBEC-mediated mutagenesis and a hypermutation phenotype, with subsequent neoantigen burden, immune response, and survival benefit. Tumors with mutations in *FGFR3* and *KRAS* may not expose enough substrate to APOBEC enzymes to promote APOBEC-mediated mutagenesis.

A

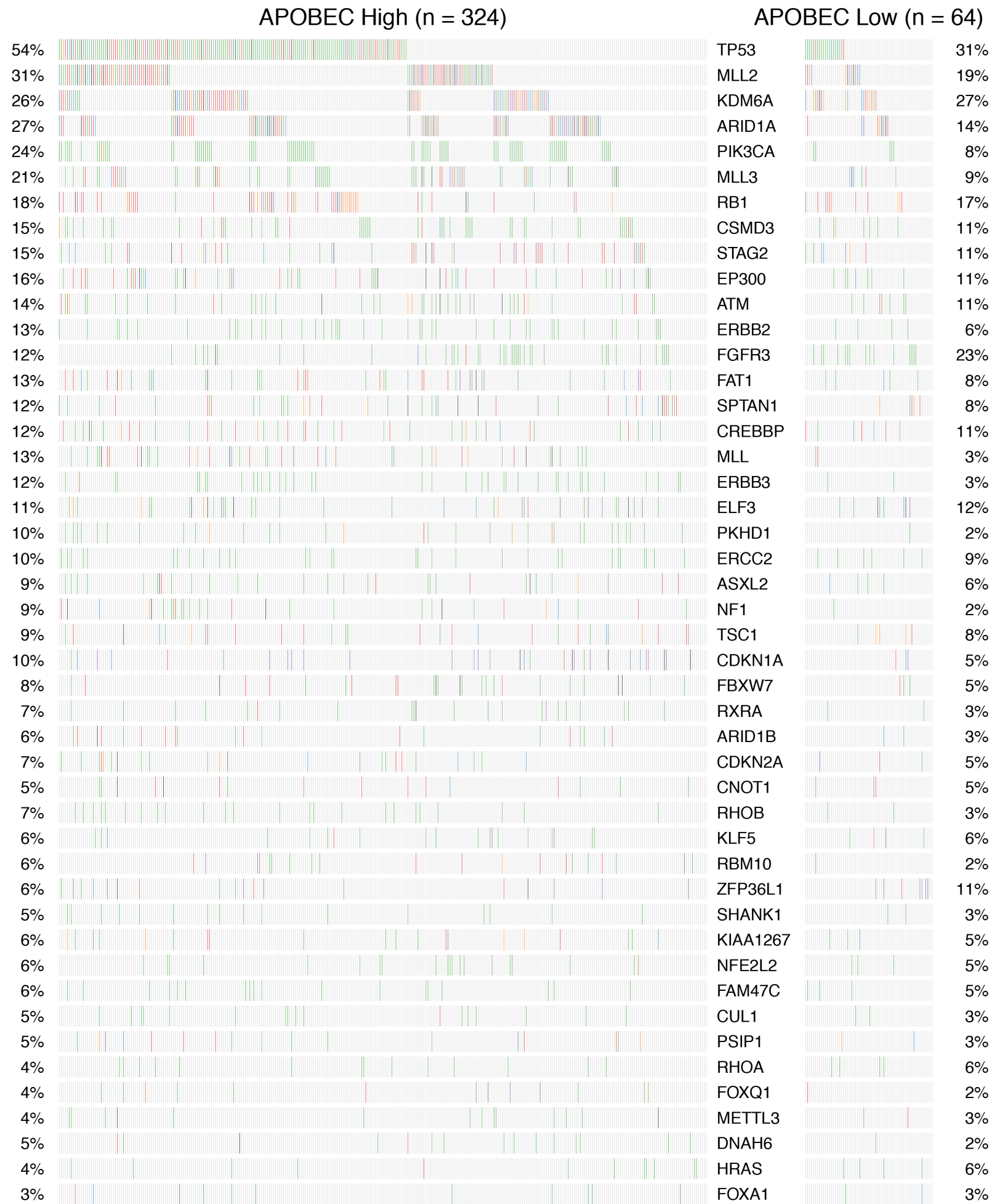
bioRxiv preprint doi: <https://doi.org/10.1101/128802>; this version posted April 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



B

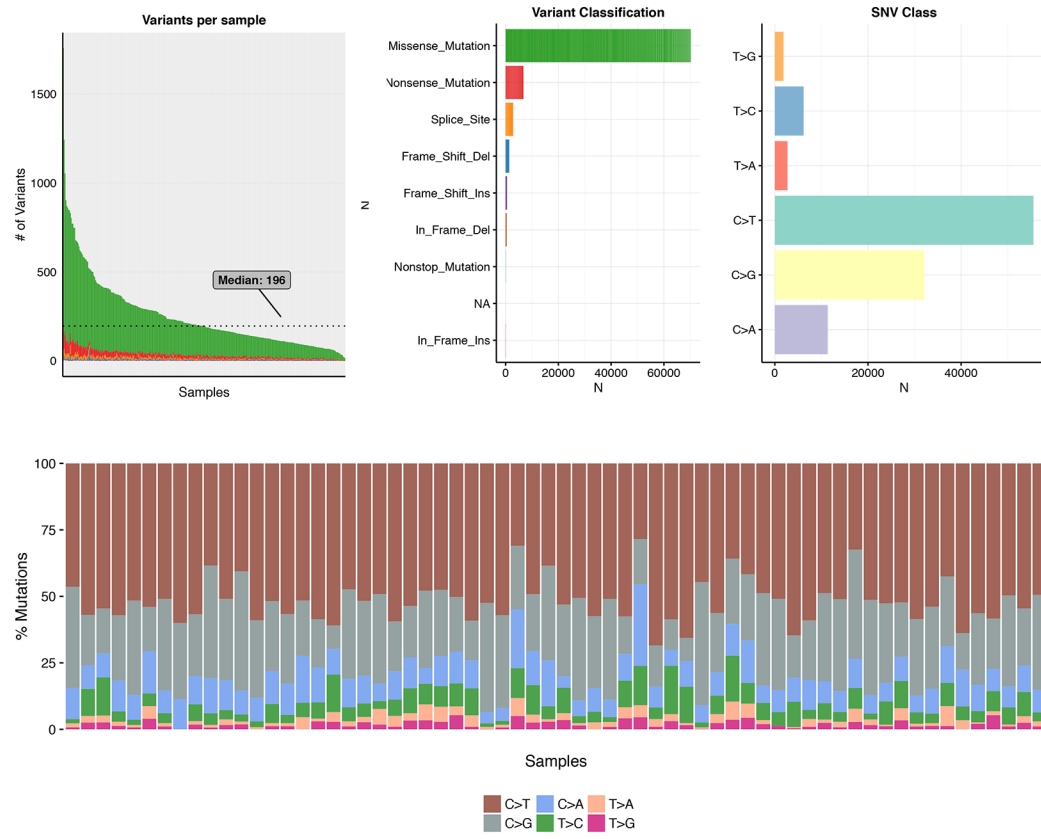


E



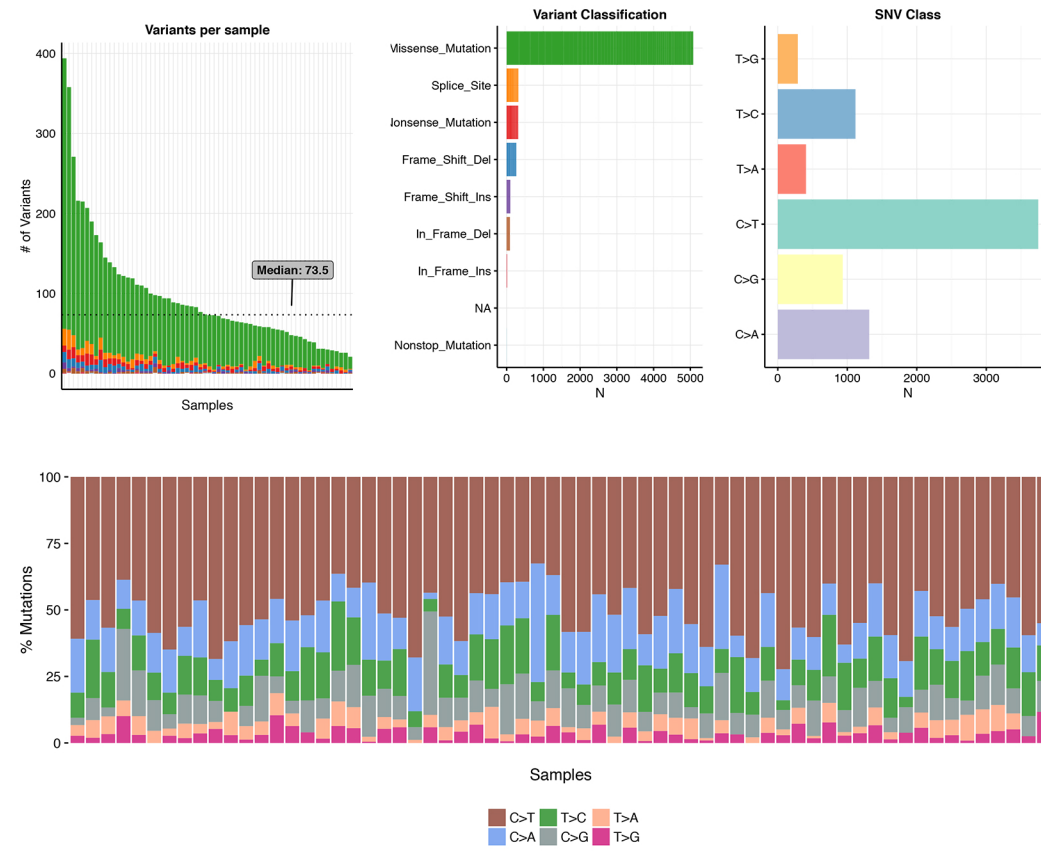
C

APOBEC High



D

APOBEC Low



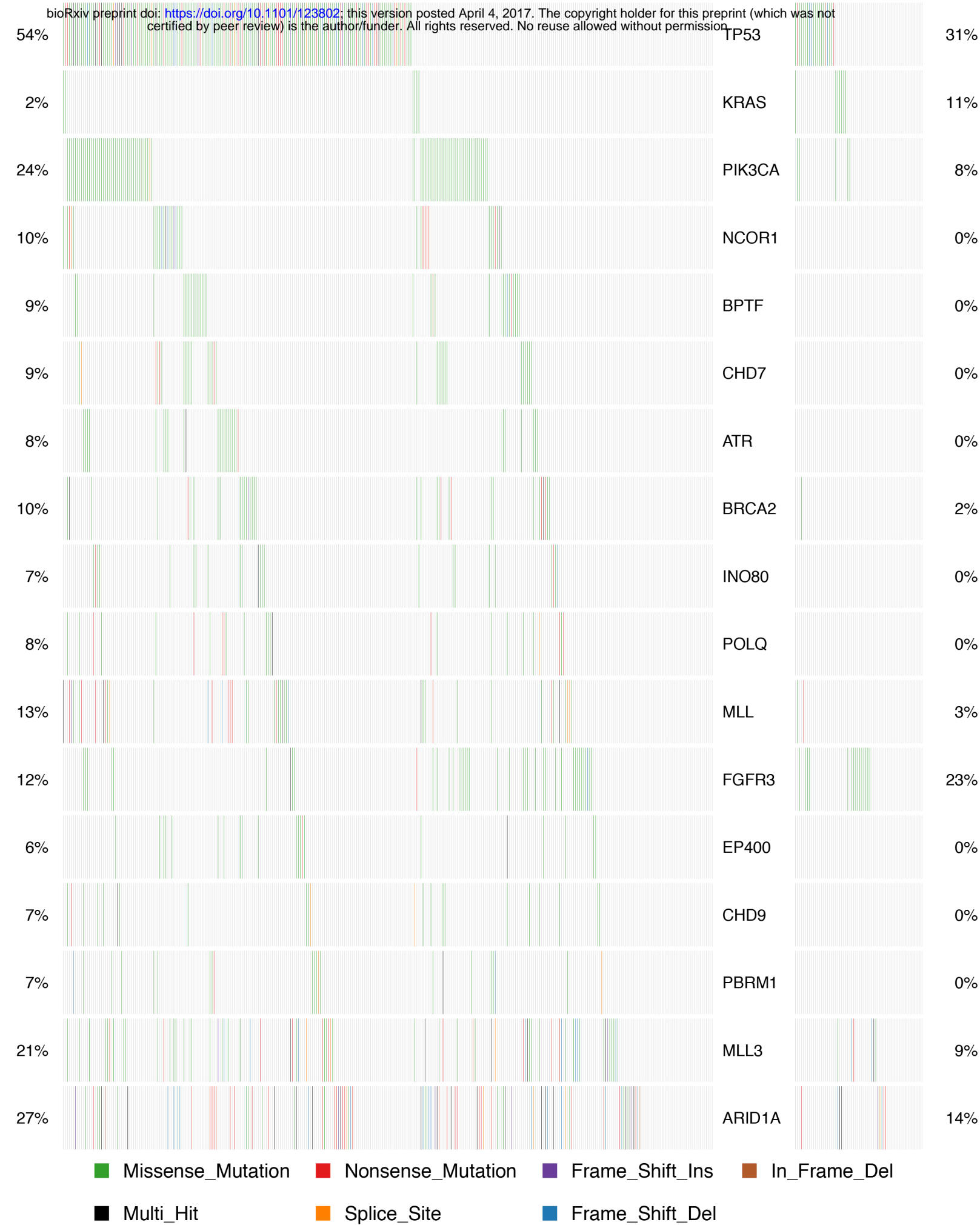
Missense_Mutation Frame_Shift_Ins Nonsense_Mutation In_Frame_Del
Frame_Shift_Del Splice_Site Multi_Hit In_Frame_Ins

A

APOBEC High (n = 324)

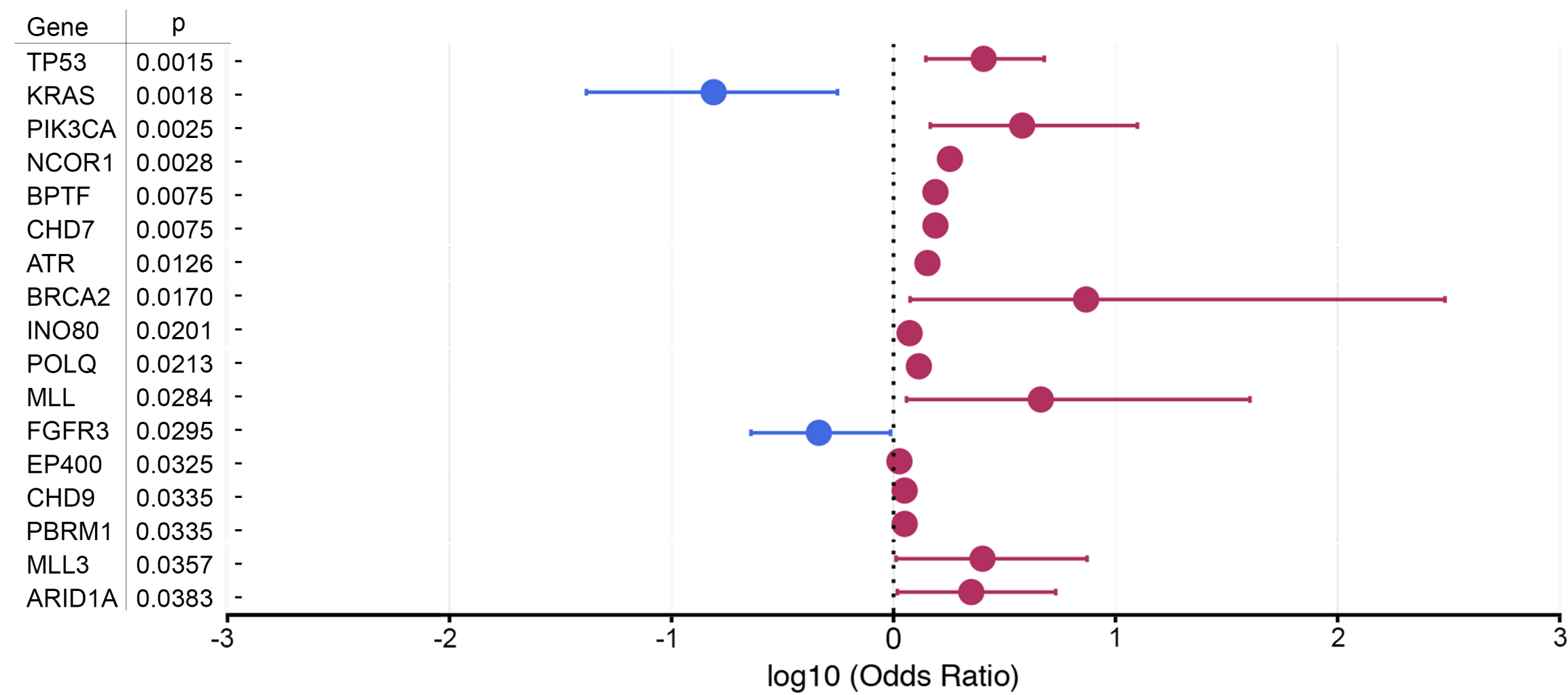
APOBEC Low (n = 64)

bioRxiv preprint doi: <https://doi.org/10.1101/123802>; this version posted April 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



B

APOBEC Low (n = 64) v/s APOBEC High (n = 324)

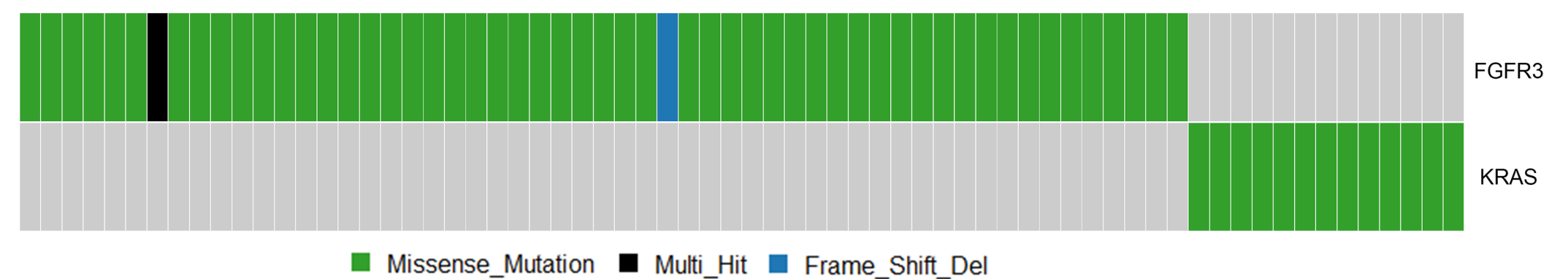


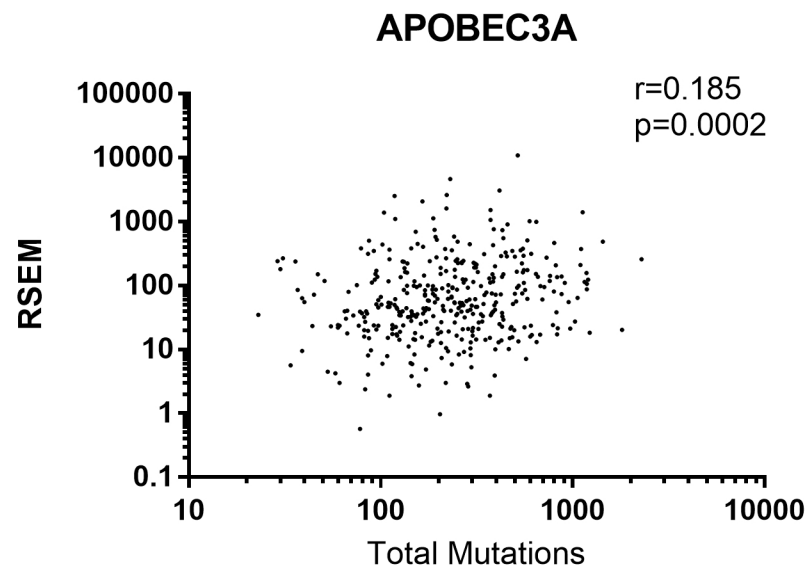
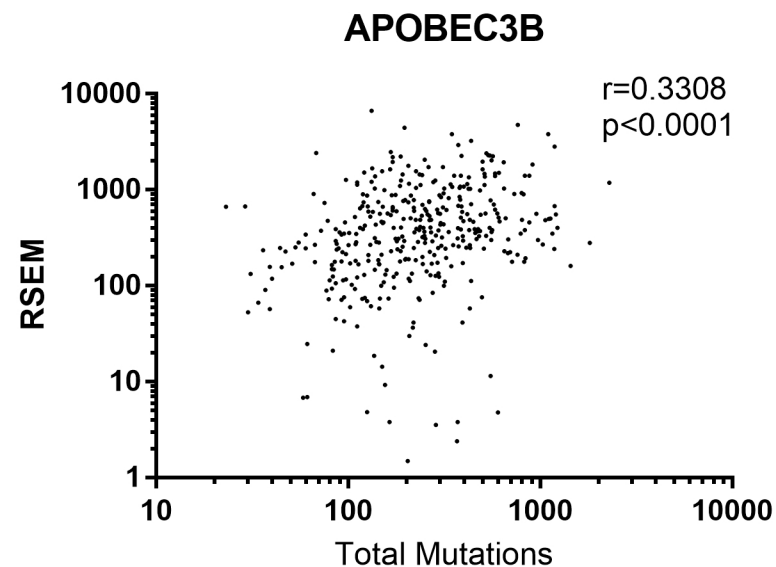
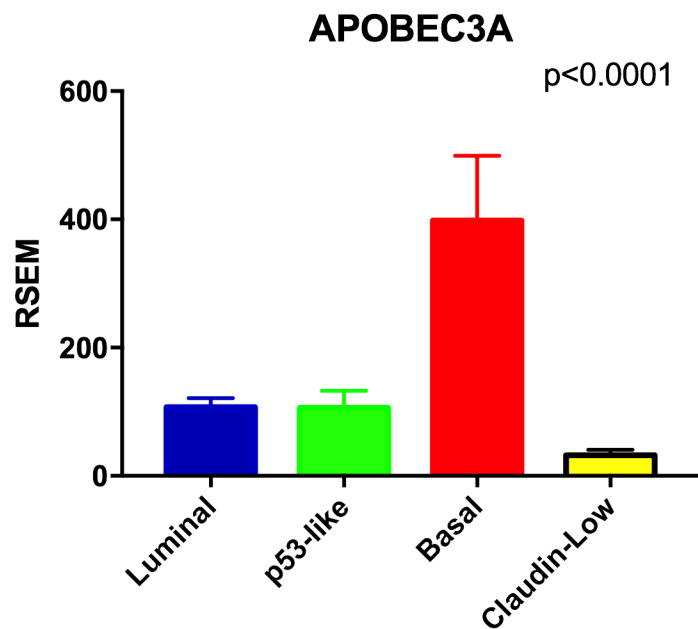
● APOBEC Low ● APOBEC High

Mutations in:
FGFR3 / RAS pathway

Mutations in:
PI(3)K/mTOR pathway
DNA damage response
Chromatin modification

C



A**B****C****D**