Identification of a tobacco-smoking induced mutation signature with biological and clinical significance in bladder cancer

Xiang-Yu Meng
Wuhan University Zhongnan Hospital
Shuo Li
Wuhan University Zhongnan Hospital
Ming-Jun Shi
Capital Medical University Affiliated Beijing Friendship Hospital
Wei-Huang Liu
Wuhan University
Hao Chen
Wuhan University Zhongnan Hospital
Jian Song
Capital Medical University Affiliated Beijing Friendship Hospital
Francois Radvanyi
Institut Curie
Isabelle Bernard-Pierrot
Institut Curie
Fubing Wang (wangfubing@znhospital.cn)
Wuhan University Zhongnan Hospital https://orcid.org/0000-0003-4171-1389
Xing-Huan Wang
Wuhan University Zhongnan Hospital

Research Article

Keywords: mutation signature, tobacco-smoking, bladder cancer, APOBEC, immune suppression, AHR, CYP1A

Posted Date: February 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1324265/v1

Abstract

Background: Tobacco-smoking is a known risk factor for bladder cancer (BCa). However, a typical tobacco-smoking induced mutation signature, COSMIC SBS4, established in lung and head-and-neck cancer, has not yet been virtually found in BCa.

Methods: De novo mutation signature extraction was performed. Association of SBS4-like signature with tobacco-smoking and clinicopathologic and molecular features were analyzed with statistical and bioinformatics analysis. The role of AHR/CYP1A in BaP-induced carcinogenic effects was functionally investigated in RT4 and T24 cell lines.

Results: We identified a mutation signature resembling the COSMIC SBS4 in a cohort of whole exome sequenced BCa tumors, with moderate contribution to the overall mutation load. Its relationship with tobacco-smoking was demonstrated given its transcriptional strand bias and association with tobacco-smoking history and other proven tobacco-induced mutation signatures, DBS2 and ID3. Load of this signature showed anti-correlation with APOBEC mutagenesis, enrichment in luminal unstable and depletion in stroma-rich molecular subtypes, and association with patients' prognosis, tumor grade, tumor proliferation, immune infiltration, inflammation, IFN-response, and urothelial versus squamous differentiation. We computationally identified and functionally validated AHR/CYP1A axis as a pivot of tobacco-induced carcinogenesis in BCa. We proposed a hypothetic mechanistic model of tobacco toxicants-related BCa development and evolution in summary of the findings.

Conclusions: A tobacco-smoking induced mutation signature with important biological and clinical significance is present in BCa. As a direct measurement of chronic tobacco genotoxic effects, it can be used for further investigations on BCa etiology and clinical management. AHR/CYP1A axis is essential in tobacco-induced carcinogenic process of BCa, which can be a potential target in BCa prevention and treatment.

Background

Bladder cancer (BCa) is one of the most frequent malignancy of the human urinary tract [1, 2]. Association between tobacco-smoking and BCa risk has long been investigated and has been epidemiologically established as analyzed in large-scale meta-analysis of observational studies, and is independent of gender, racial, and directness of exposure [3–6]. A recent Mendelian randomization study has further suggested a causal relationship between tobacco-smoking and bladder cancer, reinforcing the link [7]. Importantly, tobacco-smoking was associated with poor short-term and long-term BCa patient outcome, with undetermined mechanism [8]. DNA damages caused by tobacco-smoking related carcinogens may leave characteristic somatic alteration patterns, the so-called mutation signature, like many other mutagenesis mechanisms [9, 10]. Recent efforts of large-scale whole genome and whole exome tumor somatic mutation signature characterization along with validation in experimental models have identified signatures attributable to DNA misreplication caused by with tobacco carcinogens, ie. the

COSMIC SBS4, DBS2 and ID3. The COSMIC SBS4 has been established in tumors of tobacco-smoking related cancer types such as lung cancer and head-and-neck cancer, as well as benzo[a]pyrene (BaP, a major and potent carcinogen of cigarette-smoke) treated experiment models including normal urothelium cultured *in vitro*, as direct evidence of active exposure of tobacco carcinogens [11–14]. However, the presence of this mutation signature and association with tobacco-smoking has paradoxically not been shown in BCa tumors. Instead, a COSMIC SBS5-like signature has been repeatedly found in BCa, and reported as associated with tobacco-smoking behavior and loss-of-function mutation of *ERCC2*, an important component of nucleotide excision repair (NER) machinery which is responsible for eliminating covalent DNA adducts caused by metabolically activated BaP products [15–17]. Nevertheless, given the wide presence and clock-like nature of COSMIC SBS5 signature in pan-cancer, its association with tobacco carcinogenesis in BCa seems indirect and less straightforward [17].

We aimed here to investigate whether the cigarette-smoking induced COSMIC SBS4 signature is truly absent in BCa tumors or it could have been due to limitations of mutation signature extraction and interpretation in previous studies [18, 19]. We performed *de novo* mutation signature extraction using TCGA BLCA somatic variants of whole exome sequencing (WES). We found a mutation signature highly similar to the COSMIC SBS4. Its enrichment in current and late reformed smokers compared to life-long non-smoker and early reformed smokers and its link with DBS2 and ID3 strongly supported its association with tobacco-smoking. Further exploration revealed its correlation with patients' clinicopathologic factors and prognosis, as well as transcriptomic features and molecular classifications. Importantly, we further identified AHR as a likely pivot and demonstrated a potential role of AHR-mediated xenobiotic metabolism in tobacco-induced bladder cancer carcinogenesis. Our findings allow to propose a general model of tobacco toxicants-induced bladder cancer development and evolution, and highlight the significance of smoking-quitting and modulation of AHR/CYP1A xenobiotic metabolism axis in bladder cancer prevention and management.

Methods

Data collection

The TCGA BLCA WES somatic substitution variant calls (n = 410; #mutations = 131,660), WES somatic copy-number variant (CNV) calls (GISTIC categorized, n = 408), RNA-seq transcriptomes (n = 410, RSEM normalized counts) and patient clinical data were downloaded from cBioportal database (TCGA Cell 2017 data package) [20]. COSMIC mutation signatures (version 3.1) in numeric format was downloaded from COSMIC Mutation Signature database [21]. Raw RNA-seq counts of TCGA BLCA tumors and genome-wide CpG methylomes of TCGA BLCA tumors measured with Illumina 450K array were downloaded from UCSC Xena data portal [22]. APOBEC mutagenesis (COSMIC SBS2 and SBS13) fraction scores for TCGA BLCA tumors (n = 410) was downloaded and extracted from supplements of Shi et al [23]. Single-cell RNA-seq data of a BaSq MIBC tumor was downloaded from the GEO database (accession number, GSE145137) [24]. Markers for human skin basal / suprabasal keratinocytes and urothelial cells were extracted from The Single Cell Type Atlas of the The Human Protein Atlas (THPA) database [25, 26].

De novo mutation signature extraction

Non-negative matrix factorization (NMF) based *de novo* mutation signature extraction was applied to the WES single base substitutions (SBS) calls of the TCGA BLCA tumors [16]. The optimal number of signatures was estimated automatically using the cophenetic correlation coefficients and residual sum of squares (RSS) as described [27]. The *de novo* mutation signatures were compared against known COSMIC signatures by calculating cosine similarity as described [27]. The contribution of the signatures to each sample was obtained through NMF-based deconvolution. The above analysis was performed using the Palimpsest Bioconductor package v2.0 with default parameters [28].

Mutation origin analysis

The probability that a given mutation signature gave rise to a given mutation event in a given tumor was calculated as described in Letouzé et al. using the Palimpsest Bioconductor package v2.0 with default parameters [27].

Transcriptional strand bias analysis

The distribution of mutated G against C bases on the untranscribed strands of genes was analyzed to test if more mutated G than C on the untranscribed strands suggesting damage to guanine and activity of transcription-coupled nucleotide excision repair, particularly for C > A and T > A transversions [10, 29].

Transcriptome-based molecular subtyping

The consensus clustering classifiers by Kamoun et al. was applied to the log2 transformed RSEM transcriptome of TCGA BLCA tumors to classify the tumors to a total of 6 molecular subtypes (Luminal Papillary, LumP; Luminal Unstable, LumU; Luminal Non-specified, LumNS; Basal/Squamous, Ba/Sq; Stroma-rich; and Neuroendocrine-like, NE-like) [30]. Samples that were unable to be confidently assigned to any molecular subtype were discarded.

Transcriptome differential analysis

Raw RNA-seq counts of TCGA BLCA tumors were downloaded from UCSC Xena data portal and compared against the SBS4-like mutation burden group dichotomized at median (10 SBS4-like mutations). The DESeq2 R package was used for the transcriptome-wide differential analysis. Hypergeometric test based enrichment analysis, as well as gene set enrichment analysis (GSEA) based on the pre-ranking by log2 fold-change (log2FC), were performed using the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt, http://www.webgestalt.org/) [31].

CpG methylome differential analysis

Genome-wide CpG methylomes of TCGA BLCA tumors measured with Illumina 450K array were downloaded from UCSC data portal and compared against the SBS4-like mutation burden group dichotomized at median (10 SBS4-like mutations). The limma R package was used for the methylome-wide differential analysis, where the beta-values were first logit transformed [32].

Transcriptome-based microenvironment infiltration inference

Constitution of the tumor stroma, including infiltrated immune and mesenchymal lineages, was computationally estimated using Microenvironment Cell Populations-counter (MCP-counter), based on the tumor transcriptomes, with default parameters [33]. Inferred were the relative abundance of T cells, CD8 T cells, NK cells, cytotoxic lymphocytes, B lineage, monocytic lineage, myeloid dendritic cells, neutrophils, endothelial cells, and fibroblasts.

Single cell RNA-seq analysis

Log2 transcripts-per-million normalized gene expression of single cells from a BaSq subtype MIBC tumor was downloaded from GSE146137 dataset [34]. Genes detected in less than 3 cells were discarded. Quality control excluded cells with less than 200 genes. Data were scaled and the top 2000 variable genes were used as features for subsequent principal component analysis based linear dimension reduction. The first 9 principal components were used for a graph-based clustering approach plus Louvain modularity optimization, and uniform manifold approximation and projection (UMAP) embedding for visualization (Supplementary Figure S1). Differentially expressed genes of each cluster, ie. cluster markers were identified and used for annotation of the clusters. The highly expressed genes of the differentiated luminal tumor cells against the basal tumor cells were used as a signature of urothelial differentiation in tumor (Supplementary Table S1). All these analyses were performed using the Seurat v4 package with default parameters unless otherwise specified [35].

Cell lines and culturing

RT4 and T24 bladder transitional cell carcinoma cell lines were purchased from Procell Life Science & Technology Co., Ltd. (Wuhan, China). The cell lines were authenticated by the supplier using short tandem repeat (STR) profiling analysis. RT4 and T24 cells were cultured in McCoy’s 5A medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin per standard culture conditions suggested by the supplier. The cells were free of mycoplasma contamination.

Cell viability assay

RT4 and T24 cells were seeded in triplicate in 96-well plates and left to adhere overnight. Before treated 2.5uM or 5uM BaP for 24h and 48h, the cells were pretreated with 0.1% DMSO, 1.25uM, 2.5uM, or 5uM CH-223191, a potent selective AHR small molecule inhibitor. Control cells were treated only with DMSO. Cell viability was assessed with the Cell Counting Kit-8 (CCK-8) assay per manufacturer’s instructions. The CCK-8 kit, BaP, and CH-223191 were purchased from MedChemExpress (Princeton, NJ, USA). The assays were done in three replicates for each treatment.

Ethoxyresorufin-O-deethylase assay

Ethoxyresorufin-O-deethylase (EROD) assay was performed to measure the induction of xenobiotic-metabolizing enzyme cytochrome P-450 (CYP) 1A1 / 1A2 activity by AHR in RT4 and T24 cells pre-treated

with 0.1%DMSO or 5uM CH-223191 for 6h and then treated with 5uM BaP for 24h. The assays were done using human EROD enzyme linked immunosorbent assay (ELISA) kit (GMS18017.1; Genemed Biotechnologies, Inc.; USA). The assays were done in three replicates for each treatment.

Benzo(a)pyrene- trans-7,8-dihydrodiol-9,10-epoxide assay

The benzo(a)pyrene- trans-7,8-dihydrodiol-9,10-epoxide (BPDE), a DNA-binding genotoxic metabolite of BaP by *CYP1A1* / *CYP1A2*, was measured in RT4 and T24 cells pre-treated with 0.1%DMSO or 5uM CH-223191 for 6h and then treated with 5uM BaP for 24h. The assays were done using human BPDE ELISA kit (NBP2-70012; Novus Biologicals, LLC; USA), per manufacturer's instructions. The assays were done in three replicates for each treatment.

DNA damage assays

Assays for two tobacco-smoke associated DNA damage markers, the 8-hydroxydeoxyguanosine (8-OHdG) and phosphorylation of the Ser-139 residue of the histone variant H2AX (γH2AX) were performed to detect the genotoxic effect in RT4 and T24 cells pre-treated with 0.1%DMSO or 5uM CH-223191 for 6h and then treated with 5uM BaP for 24h. The assays were done using 8-OHdG DNA damage quantification direct kit (P-6003-48; Epigentek Group Inc; USA) and HT Gamma H2AX ELISA kit (4418-096-K; Trevigen; USA), respectively, per manufacturer's instructions. The assays were done in three replicates for each treatment.

RNA preparation and reverse transcription-polymerase chain reaction assays

RNA isolation and real-time fluorescence quantitative reverse transcription polymerase chain reaction assays (RT-PCR) was done with conventional protocols. In brief, collected cell line cells were lysed with 1ml Trizol LS (Invitrogen, USA) before purification with RNAeasy mini kits (Qiagen, Germany). RNA was subjected to cDNA synthesis and analysis using sequence specific primers, as shown in Supplementary Table S2.

Bulk RNA-sequencing

Total RNA extracted from RT4 cells was subjected to quality control for purity, integrity, contamination, and concentration. Total RNA free of degradation, DNA and protein contamination was processed for library preparation, which was then quantified with Qubit 2.0, assessed with Agilent 2000 for insert size, and then quantified with qPCR. The quality controlled library was then sequenced using 150bp pair end reads on Illumina NovaSeq platform. The fastq files were trimmed for adapter sequences and quality, and then mapped to the human reference genome hg19 using STAR package v2.5.4b [36]. Raw read counts were then calculated using the subread featureCounts v1.6.0 package [37].

Statistical analysis

Unless otherwise specified, continuous variables were described as median with inter-quartile range (IQR) and compared using Wilcoxon rank sum test, and categorical variable as count with proportion and

compared using Fisher's exact test. Survival analysis were performed with Kaplan-Meir curves, log-rank test and Cox proportional hazard modeling. GraphPad Prism 8 was used for visualization of data associated with functional assays. All statistical analyses were conducted using R version 3.5.3. Two-sided $P \leq 0.05$ was considered statistically significant.

Results

SBS4-like signature identified in BCa tumors

The optimal number of *de novo* SBS signatures was determined as k = 7 (Figure S2, upper). In the mutation signatures extracted, we found one with high similarity to the COSMIC SBS4 signature (SBS de novo 2, cosine similarity = 0.86). Mutation signatures equivalent to or resembling COSMIC SBS1 (SBS de novo 7, cosine similarity = 0.97), COSMIC SBS2 (SBS de novo 6, cosine similarity = 0.99), COSMIC SBS13 (SBS de novo 1, cosine similarity = 0.89), COSMIC SBS5 (SBS de novo 3, cosine similarity = 0.83), and COSMIC SBS10b (SBS de novo 5, cosine similarity = 0.79, showing contamination of COSMIC SBS10a) were extracted as well. One signature, SBS de novo 4, with minor contribution to the overall mutation load, showed moderate similarity (cosine similarity = 0.66) with COSMIC SBS3 related with homologous recombination deficiency (HRD). *De novo* signatures with ~0.9 or higher cosine similarity to established COSMIC SBS signatures were thereafter considered as the corresponding COSMIC signature, those with ~0.8 cosine similarity was named as the corresponding COSMIC signature followed by a '-like' suffix, except the SBS de novo 4 signature which was re-named as "Moderate similarity to SBS3, HRD' (Figure 1A). Namely, SBS de novo 1 was re-named as SBS13, SBS de novo 2 as SBS4-like, SBS de novo 3 as SBS5-like, SBS de novo 4 as SBS3 moderate similarity, SBS de novo 5 as SBS10b-like, SBS de novo 6 as SBS2, and SBS de novo 7 as SBS1. The heat map showing systematic cosine similarity analysis of the mutation signatures were shown in Figure S2 lower panel.

APOBEC-associated mutagenesis contributed most to the overall SBS mutation load (SBS13, #mutations = 55,828, 43.3%; and SBS2, #mutations = 31,244, 24.2%) same as previously reported [38], followed by SBS5-like (#mutations = 11,788, 9.1%), SBS4-like (#mutations = 8,436, 6.6%), and SBS1 (#mutations = 7,682, 6.0%). The SBS10b-like signature associated with certain *POLE* exonuclease domain mutations were basically limited to one hypermutator sample with *POLE* P286R mutation (TCGA-DK-A6AW-01; Figure S3). The distribution of sample-wise mutation signature fraction scores was shown in Figure 1B.

SBS4-like signature associated with smoking history and other smoking somatic signatures

TCGA BLCA tumors without documented tobacco-smoking history (n = 13) or current reformed smoker with unspecified duration (n = 10) were excluded from this analysis. The remaining TCGA BLCA patients were classified to four tobacco-smoking history categories, namely lifelong non-smoker (n = 111), current smoker (n = 89), early reformed smoker (current reformed smoker for >15 years, n = 114), and late reformed smoker (current reformed smoker for ≤ 15 years, n = 73). Applying a mutation etiology

estimation algorithm by Letouzé and colleagues, SBS4-like mutations were further refined here to those with SBS4-like signature as the most probable origin, totaling 5,057 mutations. We found a significantly higher proportion of tumors with detected SBS4-like mutations as well as higher number of SBS4-like mutations in current or late reformed smokers than early reformed or lifelong non-smokers (Fisher's exact test, $P = 0.017$; Wilcoxon test, $P = 0.008$; Figure 2A and 2B). All other signatures expect SBS5-like showed null association with tobacco-smoking behavior, consistent with previous reports (Figure S4) [15, 17].

We further demonstrated the correlation of SBS4-like signature with other known tobacco-smoking induced mutation signatures, ie ID3 and DBS2, as tumors with detected DBS2 (characterized by CC > AA mutations) and ID3 (characterized by 1bp deletion of C) mutations showed significantly higher number of SBS4-like mutations (Wilcoxon test, $P = 6.9 \times 10^{-9}$ and 0.038, respectively; Figure 2C and 2D).

SBS4-like mutations showed transcriptional strand bias

We found transcriptional strand bias featured by more mutated G than C bases on the untranscribed strands of genes for C > A and T > A mutations (binomial test, $P = 1.2 \times 10^{-25}$ and P = 0.005, respectively), a characteristic coherent with what has been observed in the COSMIC SBS4 mutations (Figure 2E).

SBS4-like signature anti-correlated with APOBEC mutagenesis

We looked at the Spearman's correlation coefficients among APOBEC mutagenesis fraction scores calculated in our previous work [23] and that of the SBS4-like signature which was calculated as the sample-wise proportion of NMF-deconvolution derived SBS4-like signature mutation counts against overall mutation load. The APOBEC fraction scores were as expected highly correlated with each other, and both showed significant anti-correlation with the SBS4-like signature fraction (SBS2 vs SBS4-like, Spearman's rho = -0.58, $P = 2.9 \times 10^{-37}$; SBS13 vs SBS4-like, rho = -0.50, $P = 4.8 \times 10^{-37}$; APOBEC vs SBS4-like, Spearman's rho rho = -0.60, $P = 1.9 \times 10^{-41}$; Figure 3A), suggesting potential competing relationship between these two types of mutagenesis. Coherently, the SBS4-like high tumors (dichotomized at median of SBS4-like mutation count, 10) showed significantly lower APOBEC mutation fraction and load both, compared with SBS-like low tumors (Wilcoxon test, $P = 4.2 \times 10^{-13}$ and 4.8×10^{-6}, respectively; Figure 3B and 3C).

SBS4-like signature associated with BCa molecular subtype

We further asked whether there was a link between SBS4-like mutagenesis and TCGA BLCA molecular subtypes. The MIBC consensus classes by Kamoun and colleagues was considered [30]. We found that the LumU tumors showed a strong enrichment of SBS4-like mutations (Figure 4A; Kruskal-Wallis test, $P = 5.6 \times 10^{-5}$; Wilcoxon test, $P = 3.1 \times 10^{-6}$). With categorical analysis, the SBS4-like high tumors showed significantly higher proportion of LumU subtype tumors and lower proportion of stroma-rich and LumP subtype tumors, compared to SBS4-like low tumors (Figure 4B; Chi-squared test, $P = 3.7 \times 10^{-4}$). This

enrichment was independent of overall tumor mutation load exclusively for the SBS4-like mutations, thus ruling out the possibility that the observation was merely a side-effect of higher global genomic instability in LumU tumors (Figure 4C).

SBS4-like mutation load correlated with tumor grade and prognosis

We then asked whether SBS4-like mutation load was correlated with patient's clinicopathologic factors and prognosis. The SBS4-like high tumors showed significantly higher proportion of high-grade tumors, and the two groups had similar tumor stage, and patient's age and gender profiles (Figure 4D and Figure S5). The SBS4-like high group tumors had significantly worse disease-free survival (DFS) and overall survival (OS) compared with SBS4-like low group tumors (Figure 4E; median DFS, 24.8 months in SBS4-like high and 54.8 months in SBS4-like low tumors, log-rank test, $P = 0.02$; median OS, 24.3 months in SBS4-like high and 46.8 months SBS4-like low tumors, log-rank test, $P = 0.03$). SBS4-like mutation load showed independent prognostic significance in a multivariate Cox model with age, gender, tumor stage, tumor grade, and consensus classification subtypes as adjustment covariates (SBS4-like high vs low, hazard ratio = 1.42, 95% confidence interval = [1.04, 1.94]; $P = 0.03$).

SBS4-like mutation load associated with enhanced cell cycle, suppressed IFN response and immune infiltration, and squamous differentiation

We performed transcriptome differential analysis comparing SBS4-like high and low tumors. The genes were pre-ranked in decreasing order by log2FC for GSEA, considering the Broad Institute Hall Mark 50 gene sets. Among the gene sets with statistically significant enrichment (false discovery rate (FDR) < 0.05), the cell cycle related ontologies (eg., cell cycle progression: E2F targets, normalized enrichment score (NES) = 2.29, FDR = 0; cell cycle progression: G2/M checkpoint, NES = 1.90, FDR = 6.2×10^{-4}) were top enriched for genes up-regulated in SBS4-like high tumors, and the immune/inflammation/interferon related ontologies (eg., interferon gamma response, NES = -2.47, FDR = 0; allograft rejection, NES = -2.47, FDR = 0; inflammation, NES = -2.12, FDR = 0) for those down-regulated (Figure 5A). We then looked at the difference in tumor microenvironment infiltration between SBS4-like high and low tumors, using transcriptome-derived cell population abundance estimated with the MCP-counter. Similar with findings of gene expression differential analysis, we found significantly lower abundance of T cells, myeloid dendritic cells, neutrophils, and endothelial cells in SBS4-like high tumors than SBS4-like low tumors, as well as a trend for CD8 T cells, cytotoxic lymphocytes, NK cells, B lineage cells and monocytic lineage cells (Figure 5B).

We also noted among the top genes up-regulated in SBS4-like high tumors an enrichment of skin / epidermis development / keratinization related genes, including squamous cytokeratin (*KRT1*, *KRT6B*, etc) and keratinization/cornification related proteins (squamous Kallikrein-related peptidase, small proline rich

protein family, late cornified envelope family, desmoglein family, etc; Figure 5C, Figure S6, and Supplementary Table S4). Given expression of these genes was basically found in Ba/Sq subtype tumors but no difference in the proportion of Ba/Sq tumors was observed between the SBS4-like high and low groups, we further performed transcriptomic differential analysis within the Ba/Sq tumors. Consistent with what was observed in global analysis, the Gene Ontology (GO) terms enriched for the top 200 genes up-regulated in SBS4-like high Ba/Sq tumors were basically epidermis/keratinization related (eg., GO:0031424, keratinization, FDR = 0; GO:0009913, epidermal cell differentiation, FDR = 0; Figure 5D). This was further validated by pre-ranked GSEA analysis with single cell RNA-seq derived skin supra-basal/basal keratinocyte markers (supra-basal keratinocyte markers, NES = 2.22, FDR = 0; basal keratinocyte markers, NES = 1.82, FDR = 0; Figure 5E and 5F). On the other hand, the SBS4-like high Ba/Sq tumors showed loss of urothelial differentiation in GSEA analysis with both single cell RNA-seq derived urothelial cell markers and markers of luminal epithelial cells found in a Ba/Sq tumor (urothelial cell markers, NES = -1.45, FDR = 0; markers of luminal epithelial cells found in Ba/Sq tumor GSM4307111, NES = -2.12, FDR = 0; Figure 5G and 5H).

In contrast, CpG methylome wide comparison revealed only a few differentially methylated CpG sites (Limma $P_{adj} < 0.05$) between these two groups with no difference in global CpG methylation status (Figure 5I and Table S3).

SBS4-like mutations enriched in tumors with AHR amplification and high CYP1A enzyme expression

Given the key role of AHR-mediated xenobiotic metabolism in the activation of tobacco-smoking carcinogens, we investigated the association between *AHR* copy number alterations and SBS4-like signature. We here did not consider the *AHR* mutations as we demonstrated the *AHR* Q383H hotspot was induced by APOBEC thus less likely to be associated with SBS4-like given the anti-correlation between the two, and the other *AHR* mutations were basically non-recurrent of minor significance. We found that tumors with *AHR* genomic amplifications showed significantly higher SBS4-like mutation load (Wilcoxon test, $P = 8.1 \times 10^{-4}$; Chi-squared test, $P = 5.8 \times 10^{-4}$; Figure 6A left and right), which was not observed for all other SBS signatures (Figure S7). Our genome-wide analysis of somatic CNV further confirmed this as a genuine biological effect as the *AHR* gene was found within the 7p21.1 locus with top differential amplification rate between the tumors with high and low SBS4-load (proportion of 7p21.1 amplified = 11.7% and 1.96%, number of gene-level amplification event = 287 and 35, respectively in SBS4-high and low tumors; odds ratio = 6.62, Fisher's test, $P = 8.6 \times 10^{-5}$; Figure S8). The *CYP1A1* and *CYP1A2* enzymes have been established as major canonical down-stream effectors of AHR-mediated carcinogen metabolism [39, 40], which we further confirmed by transcriptomic DEGs analysis in RT4 BCa cells pre- and post- treatment of 5uM tobacco carcinogen BaP for 24 hours as *CYP1A1* and *CYP1A2* were the top 2 up-regulated canonical xenobiotic CYP enzymes induced by BaP (log2FC = 10.8 and 4.5, $P_{adj.} = 1.5 \times 10^{-30}$ and 1.9×10^{-17}, respectively; Figure S9). We then tested the association between SBS4-like

mutation load and the collective gene expression level of the two enzymes measured by gene set variation analysis (GSVA), and we found that tumors with high *CYP1A1/A2* expression showed significantly higher SBS4-like mutation load (Figure S10), further suggesting a role of AHR in tobacco-smoking induced mutagenesis in BCa.

BaP-induced cytotoxicity and DNA-damage dependent on AHR-mediated xenobiotic metabolism in BCa cells

BaP treatment showed concentration and time-dependent inhibition on cell viability of luminal RT4 cells with high *AHR* expression (Figure S11), which was antagonized by pre-treatment of 5uM AHR inhibitor CH-223191 in a concentration dependent manner (Figure 6B, left). In contrast, the non-luminal T24 cells with low *AHR* expression (Figure S11) showed neutral effect in cell viability in response to the treatments (Figure 6B, right).

In RT4 cells, pre-treatment of 5uM AHR inhibitor CH-223191 largely attenuated the effect of 5uM BaP treatment for 24h on the induction of CYP1A1/A2 enzyme activity as measured by EROD assays, as well as the production of BPDE, a genotoxic metabolite of BaP, and DNA damage biomarkers including 8OHdG and γH2AX (Figure 6C). In contrast, all these effects were absent in T24, which showed no response to BaP treatment (Figure 6C).

No effect of short-term BaP exposure on gene expression of APOBEC3A

Our previous study showed that the APOBEC-associated hotspot mutations were basically of YTCN (Y = C or T) motif and tumors presenting these hotspots had a higher *APOBEC3A* expression-level compared with other tumors, suggesting a major role of *APOBEC3A* in APOBEC-mediated mutagenesis process in bladder cancer [41]. We in this study found an anti-correlation between APOBEC and tobacco-smoking induced mutation load, thus then tested the effect of short-term BaP exposure on *APOBEC3A* expression level in RT4 and T24 cell lines. We showed *APOBEC3A* was not among the significant DEGs between RT4 cells pre- and post-treatment of 5uM BaP for 24h in the bulk RNA-seq analysis (log2FC = 0.22, P_{adj} = 0.83; Figure S9). This was further confirmed by RT-PCR analysis (Figure 6D).

Discussion

Somatic mutations leave specific patterns corresponding to various mutagenesis processes known or undetermined, as foot-print like features called somatic mutation signatures. Linking somatic mutation signatures with exogenous and endogenous mutagenic sources, using computational and experimental approaches, offers great opportunities to trace the origin of somatic mutations with novel insights for the understanding and management of somatic diseases, particularly cancer. BCa is among the cancer types with the highest mutation burden. We and other teams have demonstrated that the majority of SBS mutations corresponds to APOBEC mutation signatures in BCa, followed by clock-like signatures SBS5-

like and SBS1 [16, 38]. Though tobacco-smoking has been well established as a causal risk factor for BCa, the smoking induced signature SBS4 that has been found widely present in lung and head-and-neck cancer tumors and detected in an *in vitro* system of BaP-treated human urothelium, has not been identified in any of the previous mutation signature works involving human BCa tumors, while a correlation between the SBS5-like signature and tobacco-smoking behavior has been suggested, which is highly likely confounded indirect association [17]. However, from a computational perspective, considering the 'flatness' of SBS5 in terms of the 96-category distribution (base substitution plus 5' and 3' bases) and its high contribution to mutations in BCa tumors, it is likely the SBS5-like signature extracted in previous studies were contaminated of mutations of other signatures including SBS4, making the absence of SBS4 a likely false negative which is a recognized issue in mutation signature analysis as discussed previously [18, 19]. Using up-to-date deconvolution strategy, we successfully *de novo* extracted a mutation signature with high cosine similarity to the COSMIC SBS4 signature, which we nominated as SBS4-like in BCa. We believe this SBS4-like mutation signature is indeed part of the finger-print of tobacco-smoking induced genomic instability, given its transcription-strand bias and significant correlation with tobacco-smoking history as well as the established tobacco-smoking induced non-SBS mutation signatures DBS2 and ID3. This identification may put a comma on the elusiveness regarding the presence of tobacco-smoking induced mutation signature in BCa and once again warns that caution is needed in the investigation and interpretation of mutation signature analysis results. The identification of this mutation signature shows the feasibility in directly measuring the active effect of tobacco-carcinogenesis in BCa, with biomarker potential superior to tobacco-smoking behavior which is inevitably at risk of confounding. It can also be used as a quantitative trait for the identification of genetic susceptibility to BCa specifically associated with tobacco-smoking. The presence of SBS signatures equivalent to or resembling COSMIC SBS4 still remains unclear in multiple cancer types with epidemiologically suggested association with tobacco-smoking, which can be either a genuine absence due to tissue-specificity in terms of metabolic reactions to carcinogen (eg. inducibility / capacity of elimination of the terminal carcinogen), DNA damage repair machinery and dominance/take-over by other competing mutagenic forces, or simply a failure in discovery due to limited power (eg. limited sample size and small sample-wise mutation load) or inappropriate computational implementation, which should be carefully explored in future studies before any conclusion can be made.

The SBS4-like signature contributed to 6.6% of all exonic SBS mutations in the cohort, with a median contribution of ~10% in individual BCa tumors. The mutation load in number of mutations per megabase was 0.33 [0, 0.91], which was ~30 times weaker compared with that in lung and head-and-neck cancer with relatively comparable global somatic mutability [42]. Besides, it showed minimal correlation with tumor mutation burden (Spearman's rho = 0.03), unlike the significant correlation found in lung and head-and-neck cancer reported previously [43]. The mechanism for these dramatic differences remains to be explored. The difference in abundance of tobacco carcinogens and tissue-specificity in metabolic transformation, elimination and DNA damage repair, and the presence of other competing mutagenic forces such as APOBEC could be likely explanations.

Like APOBEC-mutation signatures whose association with clinical and molecular tumor characteristics has been suggested [44–47], we also demonstrated SBS4-like signature of important clinical and biological significance in BCa. It showed independent prognostic significance while associated with tumor grade, molecular subtype, cell cycle and tumor stroma characteristics. The tumor cells of high SBS4-like mutations, namely those with stronger genomic instability caused by tobacco carcinogens, are of poorer differentiation and higher proliferation rate, both indicating higher aggressiveness. We observed a loss of urothelial differentiation and an increase in squamous phenotype in BaSq subtype tumors with high SBS4-like burden. The former could be a result of the stronger sensitivity to tobacco-carcinogen induced cell death in differentiated cells with CYP1A-inducibility, as what we observed in RT4 cells. The squamous histology has also been reported in BCa related with other chemical carcinogens, and shown as associated with poor prognosis in BCa patients [16, 30, 48–50]. On the other hand, the tumor stroma compartment of high SBS4-like tumors presents a non-specific pan-immune suppression phenotype, suggesting restricted potential of anti-tumor immunity. This could be a result of the non-selective cytotoxic effects induced by the tobacco carcinogens to these cell populations as well as decreased immune cell migration, as previously described [51, 52]. This is associated with weaker IFN response that has been suggested as likely source driving APOBEC mutagenesis [53], which may partly explain the reduced APOBEC mutation intensity in SBS4-like high tumors. Short-term BaP exposure is of neutral effect to APOBEC3A expression, and the full picture of the interactions between tobacco-smoking and APOBEC mutagenesis remains to be explored.

We computationally identified AHR as a likely key factor in tobacco-smoking induced mutagenesis in BCa. We tested AHR's role in the BaP metabolic reaction process in two *in vitro* cell line systems RT4 and T24. As described previously [54], these two cell lines may represent two distinct urothelial cell subtypes in terms of CYP1A inducibility, where both of them showed almost no baseline gene expression of CYP1A enzymes and comparable baseline EROD activity. We proved AHR as a pivot in the BaP-triggered induction of CYP1A enzymes and subsequent metabolic transformation, terminal toxicant generation and DNA damage, as pre-BaP-exposure AHR inhibition by small molecular antagonist significantly attenuated the effect cascade in BaP responsive RT4 cells. Though at a much lower level, the T24 also express *AHR* of the same isoform constitution as that in RT4 (CCLE RNA-seq expected transcript counts, data not shown), the presence of *AHR* mRNA seems not a sufficient condition for CYP1A inducibility. We believe at least part of the luminal differentiated urothelial cells are CYP1A activity-inducible regardless of pre-challenge CYP1A expression levels, as suggested previously the differentiation related *POR* expression may play a role [55]. However, we found no correlation between *POR* gene expression and SBS4-like mutation load in the investigated cohort, thus the exact multi-level regulatory mechanism behind this selectiveness in CYP1A inducibility remains to be clarified, with epigenetic states of the regulatory elements and AHR protein complex status among possible explanations. Further studies may leverage single-cell techniques on the systematic identification of cell populations responsible for response to xenobiotic carcinogens like BaP, as these cells could be the local *prime evil* in the chemical bladder carcinogenesis including tobacco-smoking and could be the cellular target for personalized prevention.

Conclusions

In summary of the findings, we propose a hypothetic mechanistic model regarding tobacco-smoking toxicants' role in BCa, as illustrated in Figure 7. The differentiated luminal cells with inducibility potential of AHR-mediated xenobiotic metabolism represents the biochemical reactors which transform the primary procarcinogens such as BaP to genotoxic terminal carcinogens such as BPDE, and they themselves at the same time suffer from DNA damage and stress, which predispose malignant transformation. Chronic in-situ exposure to this damage in luminal cells result in accumulation of somatic alterations and increasing genomic instability, namely toward a LumU phenotype, and will likely in long-term confer a fast proliferating phenotype and loss of urothelial differentiation. The above represents the first-hand direct scenario in the luminal path. On the other hand, the produced genotoxic metabolites trigger cytotoxic cell death particularly in the upper luminal layers which make the terminal carcinogens released to the microenvironment and deprives supra-basal protection from basal epithelial cells that further undergoes a squamous transformation under exposure to the terminal carcinogens. This microenvironment is also toxic to immune cells and inhibits infiltration and inflammatory effect, suppressing APOBEC mutagenesis that is basically APOBEC3A directed and likely IFN signaling promoted in BCa. These represent the second-hand indirect impact.

The present study has certain limitations. First, NMIBC tumors are left unstudied. Second, the mutation signature analysis was based on WES data. Third, the data used are of insufficient power to support a hotspot analysis. The NMIBC tumors were generally of mutation burden much lower than MIBC tumors, thus much larger sample size is required to reach a reasonable power for the extraction of the non-predominant SBS4-like signature in NMIBC tumors. Inclusion of WGS tumor samples will largely increase the detection power regarding recurrent / hotspot mutations or regions and be helpful for the identification of complex variant signatures and elucidation of the DNA-level mutagenesis mechanism.

In summary, we identified in bladder cancer a tobacco-smoking induced SBS mutation signature of moderate contribution to mutation burden and significant association with patient prognosis, tumor grade, molecular subtype, proliferation, squamous and urothelial differentiation, immune-infiltration, IFN-signaling, and APOBEC mutagenesis. We highlighted the AHR/CYP1A axis as a potential pivot in tobacco-smoking induced carcinogenesis in BCa. This mutation signature as direct measurement of chronic tobacco genotoxic effects, can be used for further investigations regarding BCa etiology and clinical management.

Abbreviations

AHR: Aryl Hydrocarbon Receptor;

APOBEC: Apolipoprotein B mRNA editing catalytic polypeptide-like;

BaP: Benzo(a)pyrene;

Ba/Sq: Basal/Squamous;

BCa: Bladder Cancer;

BPDE: Benzo(a)pyrene- trans-7,8-dihydrodiol-9,10-epoxide;

CCK-8: Cell Counting Kit-8;

CCLE: Cancer Cell Line Encyclopedia;

CNV: Copy Number Variant;

COSMIC: Catalogue of Somatic Mutations in Cancer;

CYP: Cytochrome P450;

DBS: Doublet Base Substitutions;

DEG: Differentially Expressed Gene;

DFS: Disease Free Survival;

DMSO: Dimethyl sulphoxide;

DNA: Deoxyribonucleic acid;

ELISA: Enzyme-linked immunosorbent assay;

EROD: Ethoxyresorufin-O-deethylase;

FBS: Fetal Bovine Serum;

FC: Fold Change;

FDR: False Discovery Rate;

GEO: Gene Expression Omnibus;

GISTIC: Genomic Identification of Significant Targets in Cancer;

GSEA: Gene Set Enrichment Analysis;

GSVA: Gene Set Variation Analysis;

HRD: Homologous Recombination Deficiency;

ID: Insertions and Deletions;

IFN: Interferon;

IQR: Inter-quartile range;

LumNS: Luminal Non-specified;

LumP: Luminal Papillary;

LumU: Luminal Unstable;

MCP-counter: Microenviroment Cell Populations-counter

MIBC: Muscle-Invasive Bladder Cancer;

NES: Normalized Enrichment Score;

NE-like: Neuroendocrine-like;

NMF: Non-negative Matrix Factorization;

NMIBC: Non-Muscle-Invasive Bladder Cancer;

OS: Overall Survival;

RNA: Ribonucleic acid;

RT-PCR: Reverse Transcription-Polymerase Chain Reaction;

RSS: Residual Sum of Squares;

SBS: Single Base Substitution;

STR: Short Tandem Repeat;

TCGA: The Cancer Genome Atlas;

THPA: The Human Protein Atlas;

UMAP: Uniform Manifold Approximation and Projection;

WES: Whole Exome Sequencing;

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Availability of data and materials

Data collected from publically available sources were described in the Data Collection part of the Materials and Methods section. The bulk RNA-seq raw count data for RT4 cells with/without BaP treatment were deposited at figshare data repository. Web links to the data sources involved are provided below:

1. cBioPortal, https://www.cbioportal.org/

2. COSMIC Mutation Signature, https://cancer.sanger.ac.uk/cosmic/signatures/index.tt

3. UCSC Xena, http://xena.ucsc.edu/

4. The Human Protein Atlas, https://www.proteinatlas.org/

5. GSE145137, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145137

6. APOBEC fraction score, https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00781-y

7. RNA-seq of RT4 cells, https://doi.org/10.6084/m9.figshare.16674439.v1

Other supporting data can be available from the corresponding author upon reasonable request.

Competing Interest

The authors declare no conflict of interest.

Authors' Contributions

All authors contributed to the design of the study. I.B.P., F.R., F.B.W, and X.H.W co-supervised the study. X.Y.M, S.L., and M.J.S. contributed equally to this work, they performed data collection, data analysis, experiments, and data interpretation. All authors contributed in the preparation of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Improvement Project for Theranostic ability on Difficulty Miscellaneous disease (Tumor) and the research fund from medical Sci-Tech innovation platform of Zhongnan Hospital (No. ZLYNXM202008 and PTXM2021001), and the SEnACLE project (2019-1-TABAC-02-1). This work was also funded by Medical Top-talented youth development project of Hubei Province and the Health Commission of Hubei Province scientific research project (WJ2021M172). MJS was supported by

National Natural Science Foundation of China (82002672). XYM was supported by a fellowship from ITMO Cancer AVIESAN within the framework of Cancer Plan.

Acknowledgment

We thank The Cancer Genome Atlas (TCGA), the Human Protein Atlas (HPA) and the Catalogue Of Somatic Mutations In Cancer (COSMIC) initiatives for providing accessibility to various types of data involved in the study.

References

1. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder Cancer: A Review JAMA. 2020;324(19):1980–91.
2. Richters A, Aben KKH, Kiemeney L. The global burden of urinary bladder cancer: an update. World J Urol. 2020;38(8):1895–904.
3. van Osch FH, Jochems SH, van Schooten FJ, Bryan RT, Zeegers MP. Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies. Int J Epidemiol. 2016;45(3):857–70.
4. Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. JAMA. 2011;306(7):737–45.
5. Masaoka H, Matsuo K, Ito H, Wakai K, Nagata C, Nakayama T, Sadakane A, Tanaka K, Tamakoshi A, Sugawara Y, et al. Cigarette smoking and bladder cancer risk: an evaluation based on a systematic review of epidemiologic evidence in the Japanese population. Jpn J Clin Oncol. 2016;46(3):273–83.
6. Yan H, Ying Y, Xie H, Li J, Wang X, He L, Jin K, Tang J, Xu X, Zheng X. Secondhand smoking increases bladder cancer risk in nonsmoking population: a meta-analysis. Cancer Manag Res. 2018;10:3781–91.
7. Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaelsson K, Burgess S. Smoking, alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and international genetic consortia participants. PLoS Med. 2020;17(7):e1003178.
8. Mori K, Mostafaei H, Abufaraj M, Yang L, Egawa S, Shariat SF. Smoking and bladder cancer: review of the recent literature. Curr Opin Urol. 2020;30(5):720–5.
9. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415–21.
10. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94–101.
11. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet. 2016;48(6):607–16.

12. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576–82.
13. Nik-Zainal S, Kucab JE, Morganella S, Glodzik D, Alexandrov LB, Arlt VM, Weninger A, Hollstein M, Stratton MR, Phillips DH. The genome as a record of environmental exposure. Mutagenesis. 2015;30(6):763–70.
14. Baker SC, Mason AS, Southgate J. Procarcinogen Activation and Mutational Signatures Model the Initiation of Carcinogenesis in Human Urothelial Tissues In Vitro. Eur Urol. 2020;78(2):143–7.
15. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea A, Getz G. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat Genet. 2016;48(6):600–6.
16. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;171(3):540–56 e525.
17. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354(6312):618–22.
18. Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L, Royo R, Ziccheddu B, Puente XS, Avet-Loiseau H, et al. A practical guide for mutational signature analysis in hematological malignancies. Nat Commun. 2019;10(1):2969.
19. Koh G, Zou X, Nik-Zainal S. Mutational signatures: experimental design and analytical framework. Genome Biol. 2020;21(1):37.
20. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.
21. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941–7.
22. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–8.
23. Shi MJ, Meng XY, Fontugne J, Chen CL, Radvanyi F, Bernard-Pierrot I. Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. Genome Med. 2020;12(1):85.
24. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
25. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al: Proteomics. Tissue-based map of the human proteome. Science 2015,

347(6220):1260419.
26. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, et al: A pathology atlas of the human cancer transcriptome. Science 2017, 357(6352).
27. Letouze E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, Bayard Q, Bacq D, Meyer V, Semhoun J, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8(1):1315.
28. Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, Zucman-Rossi J, Letouze E. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. Bioinformatics. 2018;34(19):3380–1.
29. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. Cell. 2016;164(3):538–49.
30. Kamoun A, de Reynies A, Allory Y, Sjodahl G, Robertson AG, Seiler R, Hoadley KA, Groeneveld CS, Al-Ahmadie H, Choi W, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020;77(4):420–33.
31. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199–205.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
33. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautes-Fridman C, Fridman WH, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17(1):218.
34. Lee HW, Chung W, Lee HO, Jeong DE, Jo A, Lim JE, Hong JH, Nam DH, Jeong BC, Park SH, et al. Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. Genome Med. 2020;12(1):47.
35. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–87 e3529.
36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
37. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
38. Shi MJ, Meng XY, Lamy P, Banday AR, Yang J, Moreno-Vega A, Chen CL, Dyrskjot L, Bernard-Pierrot I, Prokunina-Olsson L, et al. APOBEC-mediated Mutagenesis as a Likely Cause of FGFR3 S249C Mutation Over-representation in Bladder Cancer. Eur Urol. 2019;76(1):9–13.
39. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem. 2004;279(23):23847–50.

40. Vogel CFA, Van Winkle LS, Esser C, Haarmann-Stemmann T. The aryl hydrocarbon receptor as a target of environmental stressors - Implications for pollution mediated stress and inflammatory responses. Redox Biol. 2020;34:101530.
41. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47(9):1067–72.
42. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214–8.
43. Desrichard A, Kuo F, Chowell D, Lee KW, Riaz N, Wong RJ, Chan TA, Morris LGT. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. J Natl Cancer Inst. 2018;110(12):1386–92.
44. Wang S, Jia M, He Z, Liu XS. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. Oncogene. 2018;37(29):3924–36.
45. Faden DL, Ding F, Lin Y, Zhai S, Kuo F, Chan TA, Morris LG, Ferris RL. APOBEC mutagenesis is tightly linked to the immune landscape and immunotherapy biomarkers in head and neck squamous cell carcinoma. Oral Oncol. 2019;96:140–7.
46. Messerschmidt C, Obermayer B, Klinghammer K, Ochsenreither S, Treue D, Stenzinger A, Glimm H, Frohling S, Kindler T, Brandts CH, et al. Distinct immune evasion in APOBEC-enriched, HPV-negative HNSCC. Int J Cancer. 2020;147(8):2293–302.
47. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, Miller SD, Meeks JJ. APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. Oncotarget. 2018;9(4):4537–48.
48. Herman CJ, Vegt PD, Debruyne FM, Vooijs GP, Ramaekers FC. Squamous and transitional elements in rat bladder carcinomas induced by N-butyl-N-4-hydroxybutyl-nitrosamine (BBN). A study of cytokeratin expression. Am J Pathol. 1985;120(3):419–26.
49. Cao L, Zhou XD, Sens MA, Garrett SH, Zheng Y, Dunlevy JR, Sens DA, Somji S. Keratin 6 expression correlates to areas of squamous differentiation in multiple independent isolates of As(+3)-induced bladder cancer. J Appl Toxicol. 2010;30(5):416–30.
50. Dotson A, May A, Davaro F, Raza SJ, Siddiqui S, Hamilton Z. Squamous cell carcinoma of the bladder: poor response to neoadjuvant chemotherapy. Int J Clin Oncol. 2019;24(6):706–11.
51. Martin EM, Clapp PW, Rebuli ME, Pawlak EA, Glista-Baker E, Benowitz NL, Fry RC, Jaspers I. E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke. Am J Physiol Lung Cell Mol Physiol. 2016;311(1):L135–44.
52. de la Iglesia JV, Slebos RJC, Martin-Gomez L, Wang X, Teer JK, Tan AC, Gerke TA, Aden-Buie G, van Veen T, Masannat J, et al. Effects of Tobacco Smoking on the Tumor Immune Microenvironment in Head and Neck Squamous Cell Carcinoma. Clin Cancer Res. 2020;26(6):1474–85.

53. Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, Figueroa JD, Zhu B, Koutros S, Kubo M, et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat Genet. 2016;48(11):1330–8.
54. Plottner S, Selinski S, Bolt HM, Degen GH, Hengstler JG, Roos PH, Follmann W. Distinct subtypes of urinary bladder epithelial cells with inducible and non-inducible cytochrome P450 1A1. Arch Toxicol. 2009;83(2):131–8.
55. Baker SC, Arlt VM, Indra R, Joel M, Stiborova M, Eardley I, Ahmad N, Otto W, Burger M, Rubenwolf P, et al. Differentiation-associated urothelial cytochrome P450 oxidoreductase predicates the xenobiotic-metabolizing activity of "luminal" muscle-invasive bladder cancers. Mol Carcinog. 2018;57(5):606–18.

Figures

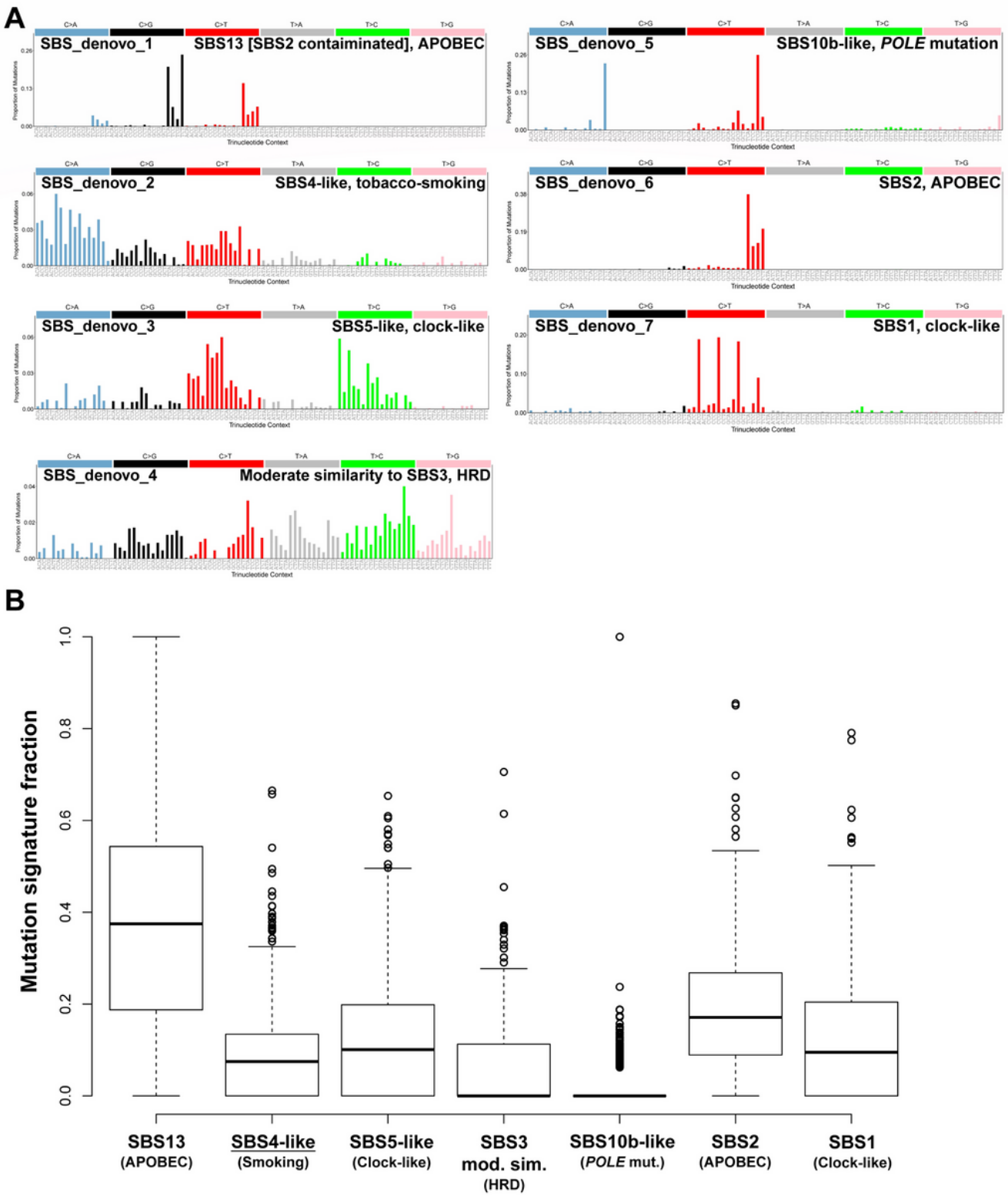


Figure 1

(A) Mutation signature *de novo* extraction identified 7 SBS mutation signatures in the TCGA BLCA cohort whole exome sequenced tumors (n = 410), including one with high cosine similarity with COSMIC SBS4, named SBS4-like. **(B)** Distribution of fraction of the 7 SBS signatures in the tumors. SBS4-like signature showed moderate contribution with ~10% as the median fraction.

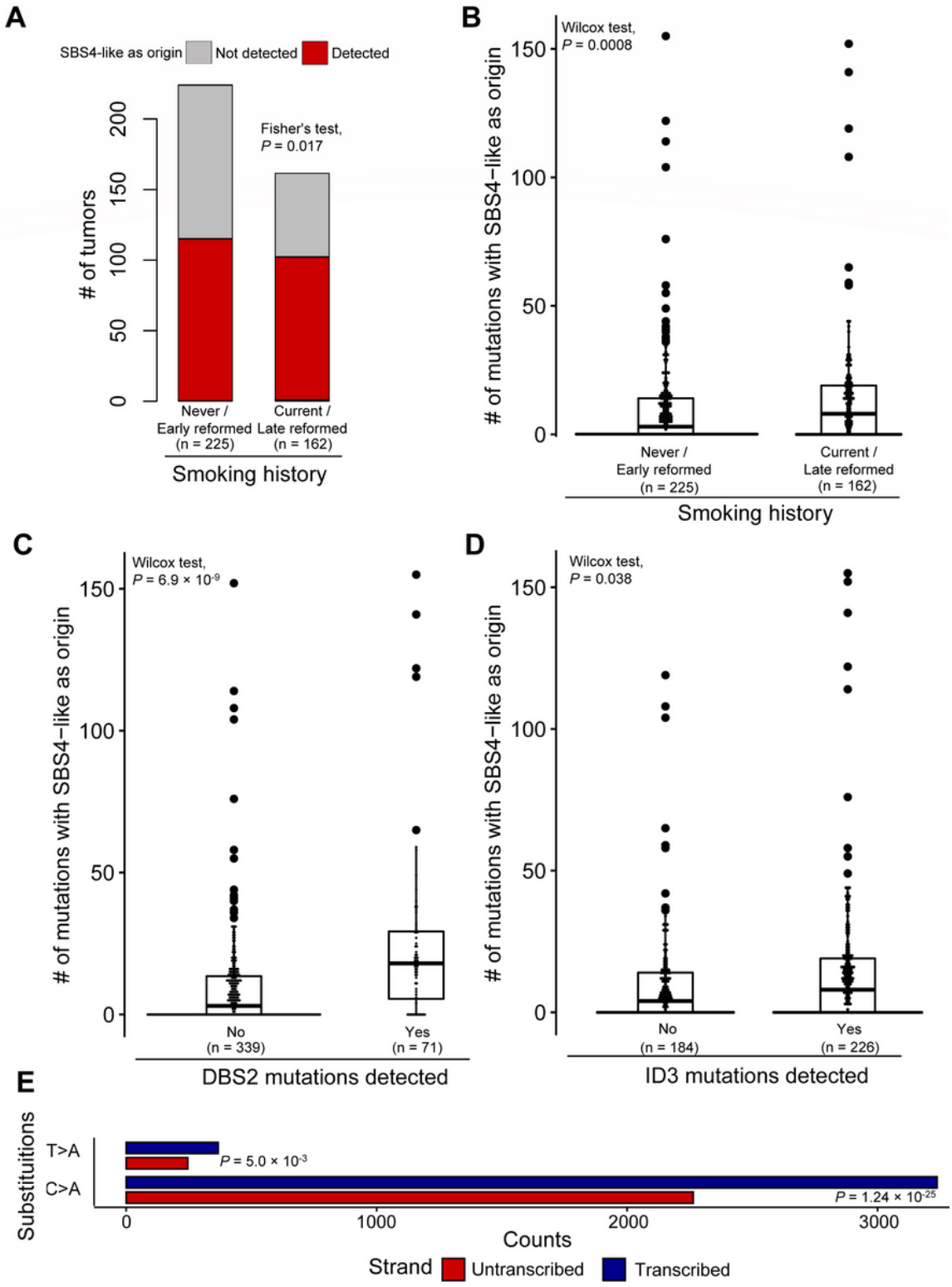


Figure 2

(A) Higher proportion of patients with SBS4-like mutation detected in current / late reformed smokers compared with never / early reformed smokers. **(B)** Higher SBS4-like mutation burden in current / late reformed smokers than never / early reformed smokers. **(C)** Higher SBS4-like mutation burden in tumors

with DBS2 mutation detected. **(D)** Higher SBS4-like mutation burden in tumors with ID3 mutation detected. **(E)** Transcription strand bias of SBS4-like mutations in BCa.

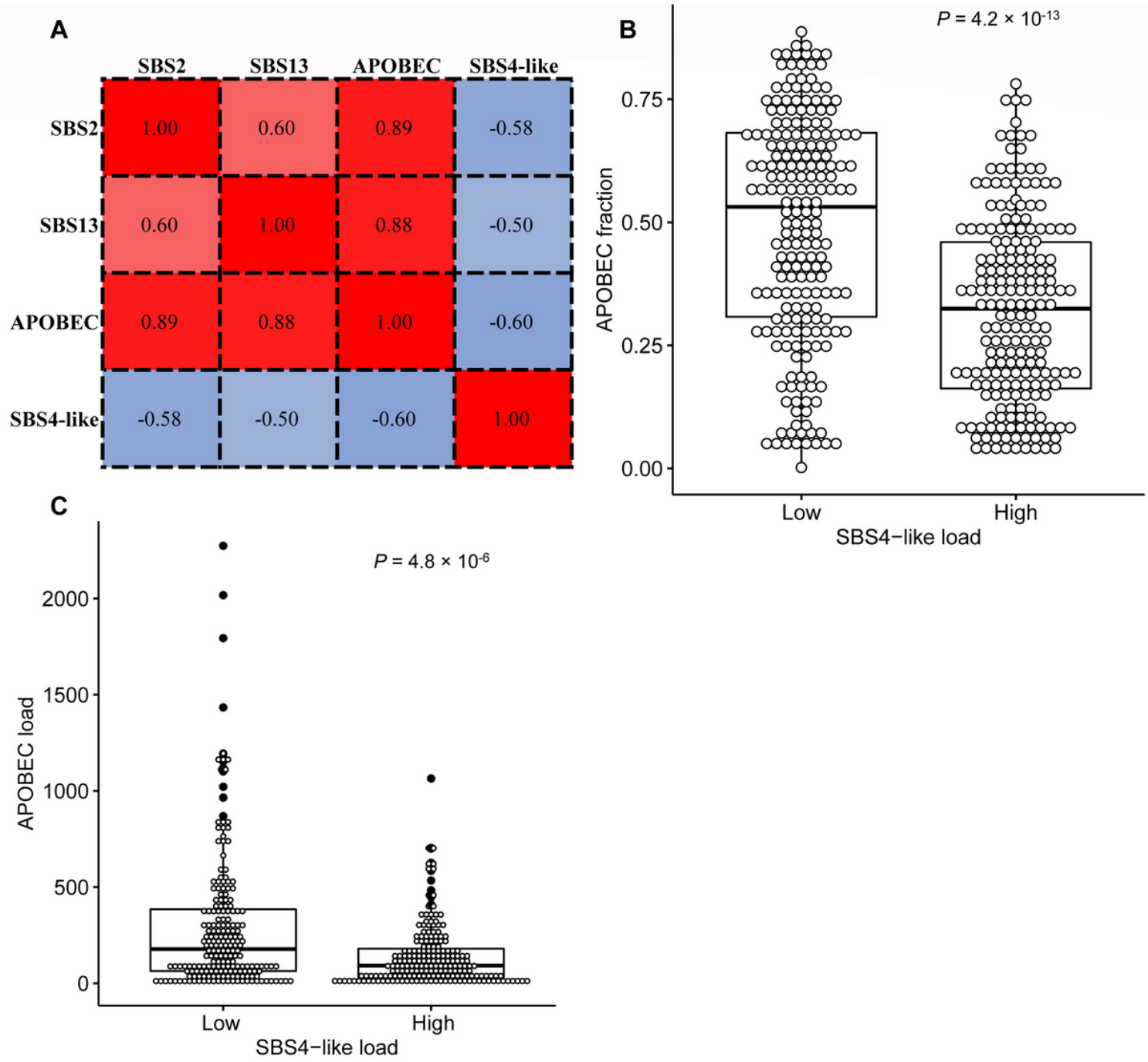
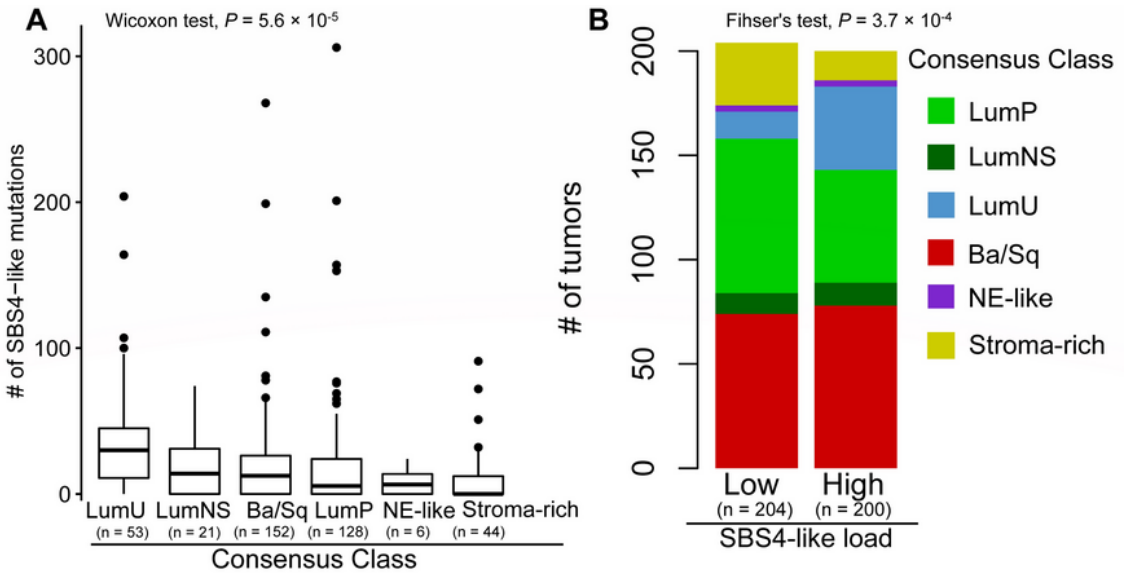


Figure 3

(A) Correlation matrix showing Spearman's correlation coefficient among sample fractions of APOBEC mutation signatures and smoking-related SBS4-like signature. **(B)** BCa tumors with high SBS4-like mutation load showed significantly lower fraction of APOBEC mutations, compared with tumors with low SBS-like mutation load. **(C)** BCa tumors with high SBS4-like mutation load showed significantly lower

APOBEC mutation load, compared with tumors with low SBS-like mutation load, suggesting anti-correlation between APOBEC and smoking related mutagenesis in BCa.



C

Signatures	Count median		Median ratio	95% CI
	LumU	Others		
SBS13 (APOBEC)	120	66	1.818	[1.075, 2.952]
SBS4-like (Smoking)	30	7	4.286	[2.154, 17.50]
SBS5-like (Clock-like)	21	18	1.167	[0.857, 1.944]
SBS3 mod.sim. (HRD)	0	0	NA	NA
SBS10b-like (*POLE* mutation)	0	0	NA	NA
SBS2 (APOBEC)	59	33	1.788	[1.135, 2.615]
SBS1 (Clock-like)	25	18	1.389	[0.000, 1.929]
Control background (All_Mutation)	321	210	1.529	[1.110, 1.984]

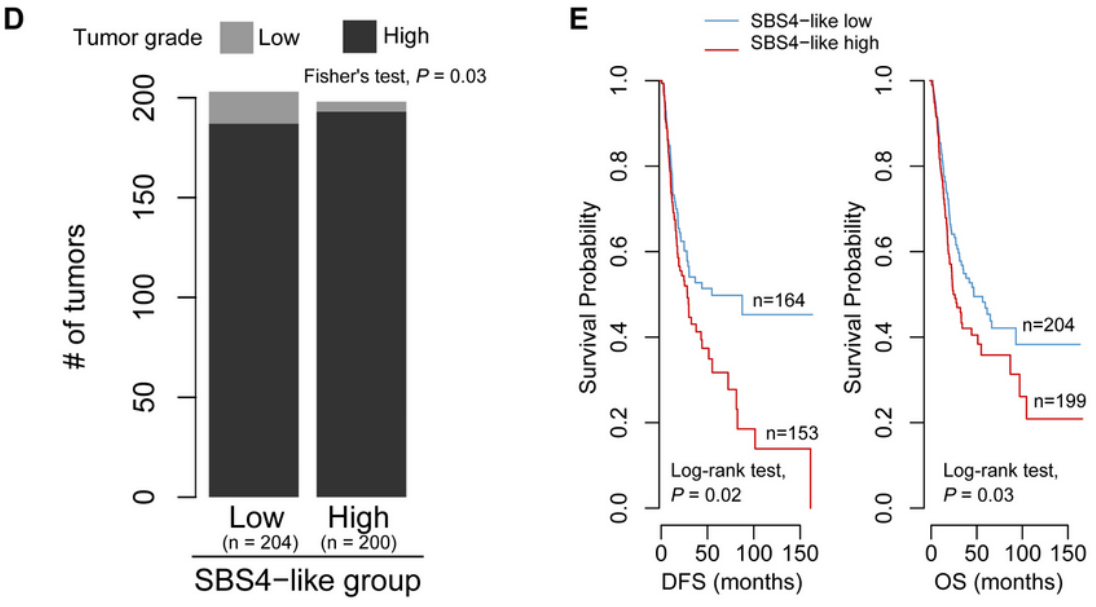


Figure 4

(A) Variable SBS4-like mutation load in tumors of different consensus molecular subtypes, with LumU the highest and Stroma-rich the lowest. **(B)** Proportion of Stroma-rich subtype in SBS4-like high tumors were half the value of that in SBS-like low tumors. Though comparable proportion of luminal and basal subtype tumors between the SBS4-like high and low tumors, significantly higher proportion of LumU tumors were found in SBS4-like high tumors. **(C)** Enrichment of SBS4-like mutations in LumU tumors was independent of the globally higher genomic instability of LumU tumors, as indicated by the non-overlap of 95% confidence interval (95% CI) of the ratio of the median count **(D)** High SBS4-like mutation burden associated with high tumor histologic grade. **(E)** Kaplan-Meier curves showing high SBS4-like mutation burden associated with reduced DFS and OS.

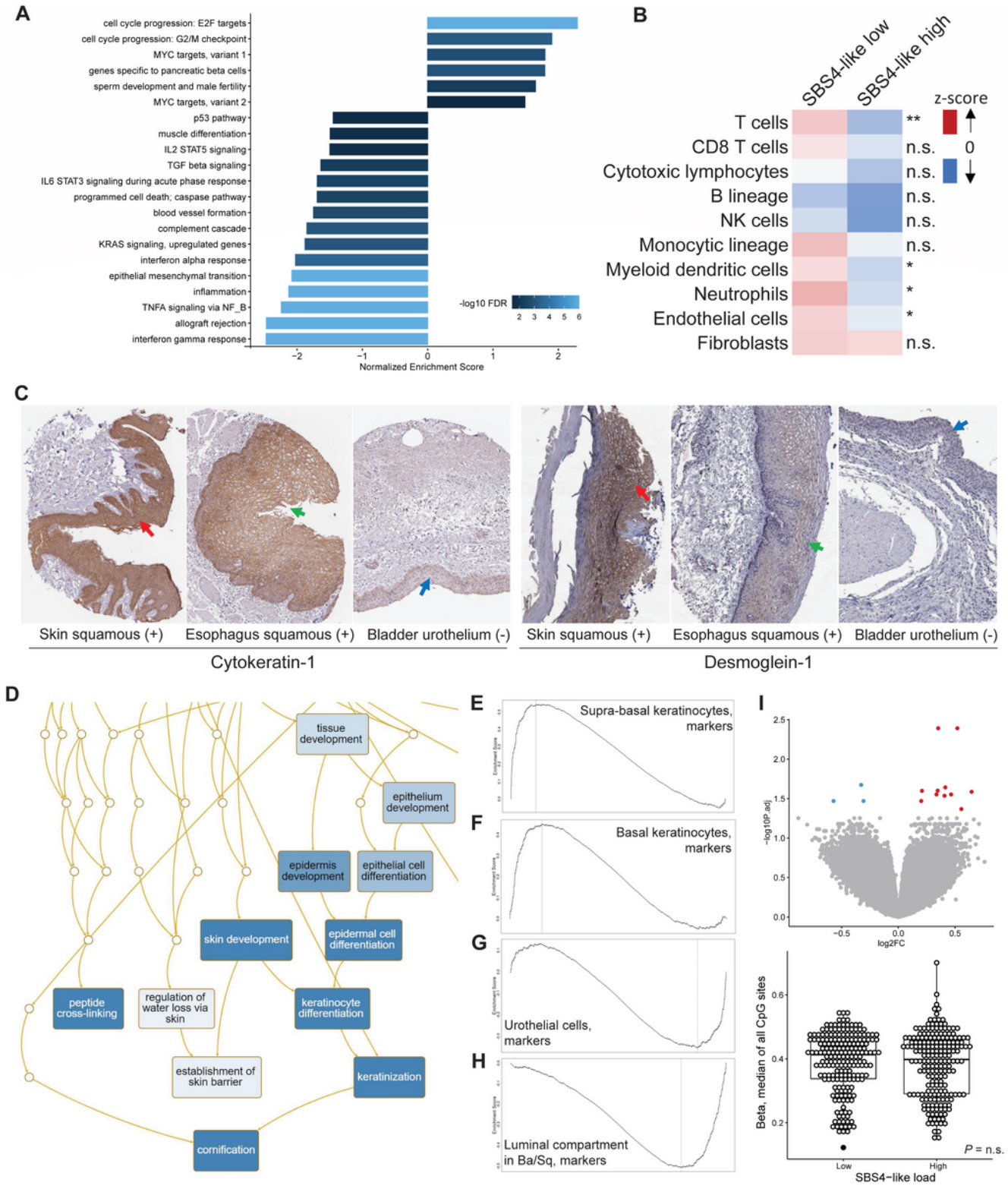


Figure 5

(A) Cancer hall mark gene-set enrichment analysis of differential gene expression between SBS4-like high and low tumors revealed up-regulation of cell cycle genes and down-regulation of interferon response and inflammation related genes. **(B)** SBS4-like high tumors present a suppressed immune microenvironment as inferred with MCP-counter. (C) Genes specifically expressed in squamous epithelium (eg. skin epidermis and esophagous squamous epithelium, taking *KRT1*/cytokeratin-1 and DSG1/Desmoglein-1 as

examples) but negative in bladder urothelium were among the genes top up-regulated in SBS4-like high tumors (see Supplementary Table S5). **(D)** Skin epidermis / keratinization related gene ontologies were enriched in the genes top up-regulated in SBS4-like high compared with SBS4-like low BaSq tumors. **(E-H)** Gene-set enrichment analysis with single-cell analysis derived cell population signatures revealed loss of urothelial differentiation and increased squamous phenotype in SBS4-like high compared with SBS4-like low BaSq tumors. **(I)** Minimal difference in CpG methylation between SBS4-like high and low tumors.

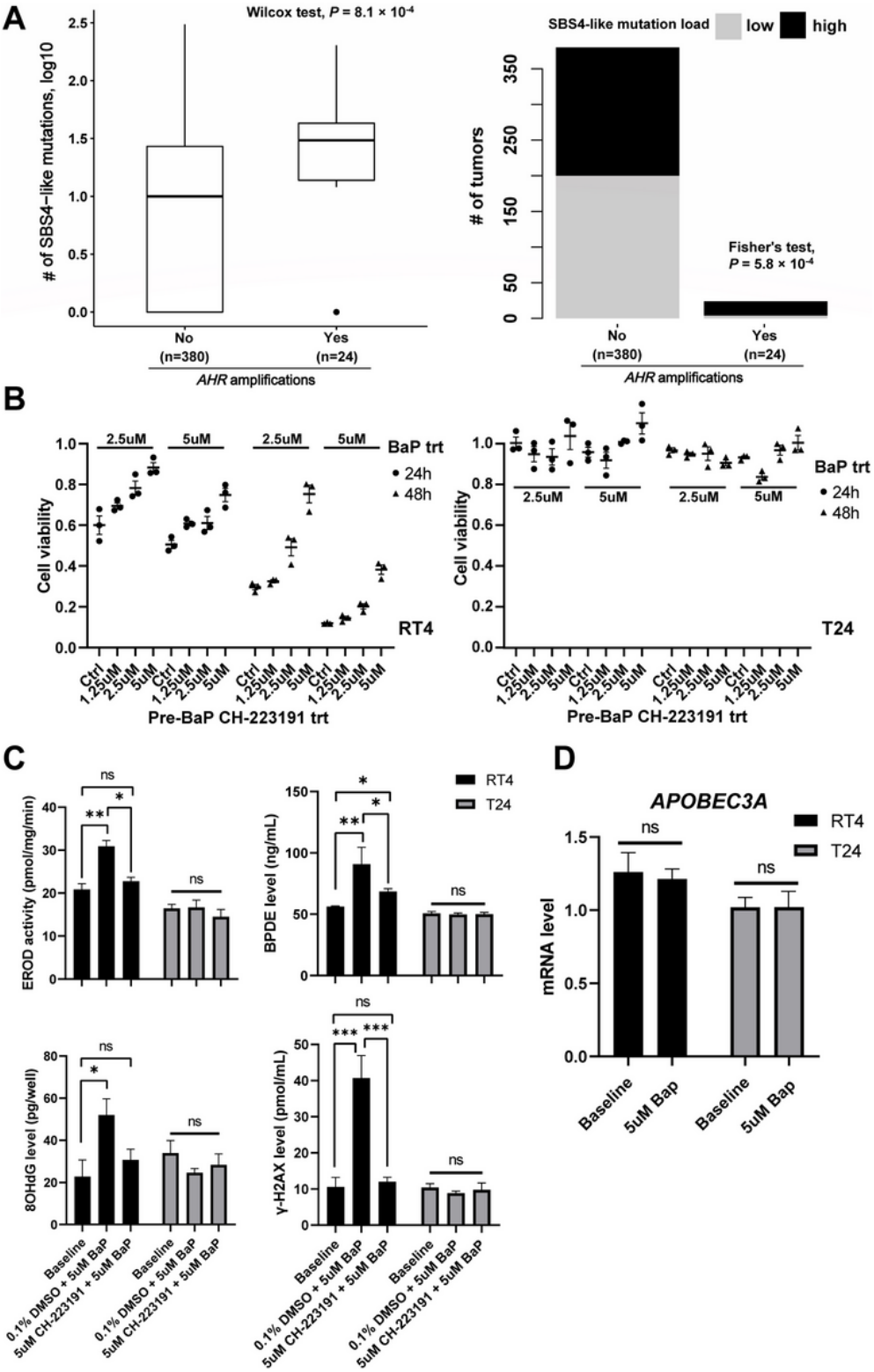


Figure 6

(A) *AHR* amplified tumors showed significantly higher SBS4-like mutation burden (left) and strong enrichment of SBS4-like high tumors (right). **(B)** Time- and concentration-dependent cytotoxic effect of BaP attenuated by AHR inhibition in a dose-dependent manner, in BaP-responsive RT4 cells (left) but not BaP non-responsive T24 cells (right). **(C)** AHR inhibition suppressed the BaP-triggered induction of CYP1A

activity, production of terminal carcinogen BPDE, and DNA damage in RT4 cells. No effect in T24 cells. **(D)** Short term BaP exposure showed neutral effect on *APOBEC3A* gene expression in RT4 cells.

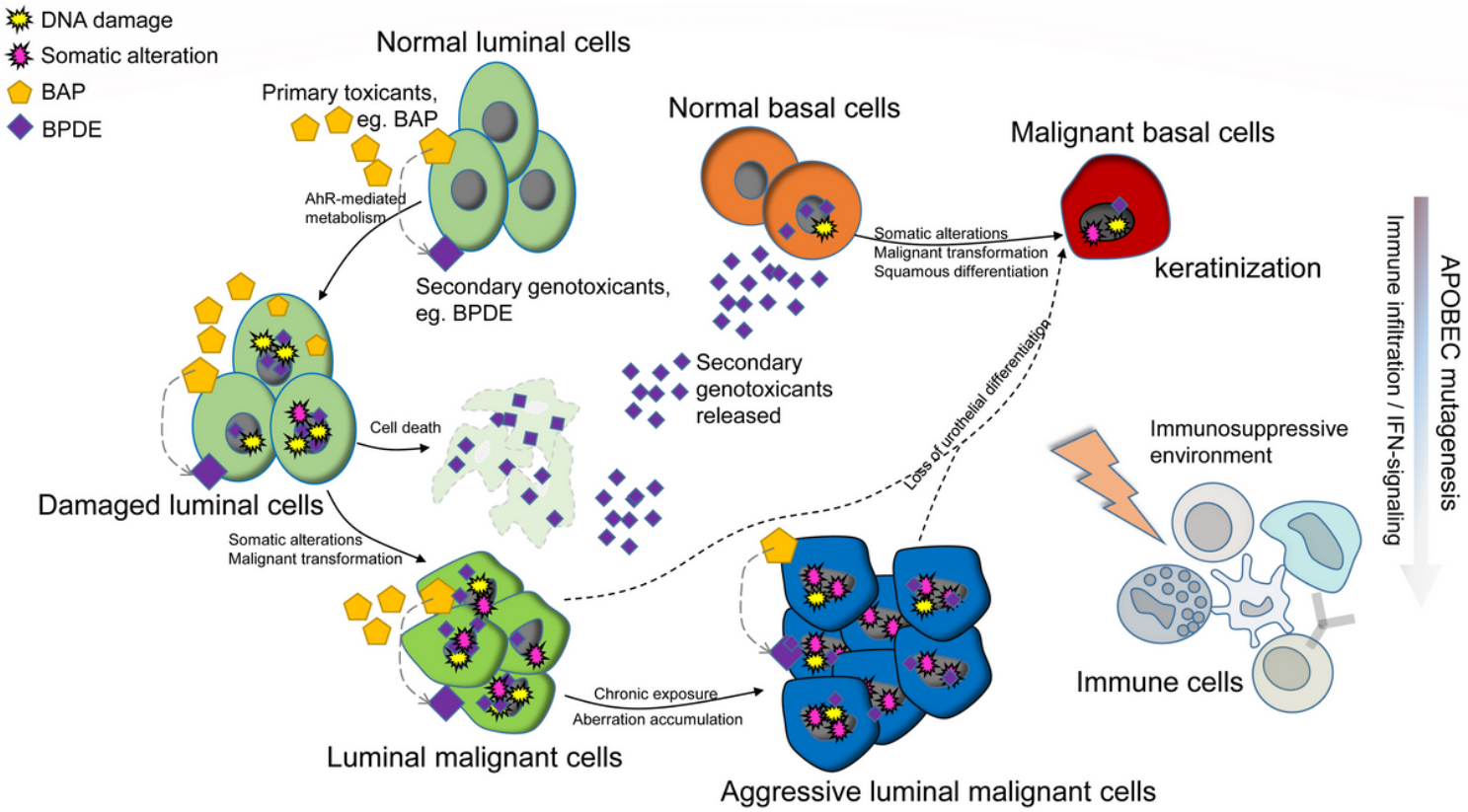


Figure 7

A hypothetic mechanistic model illustrating roles and interactions of tobacco-smoking carcinogens in BCa.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pdf
- SupplementaryTables.pdf