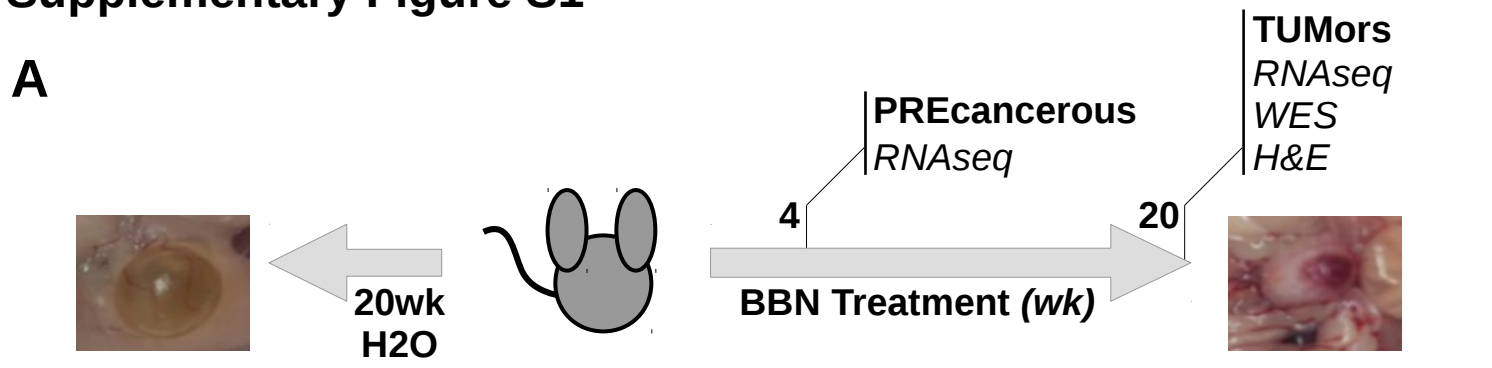
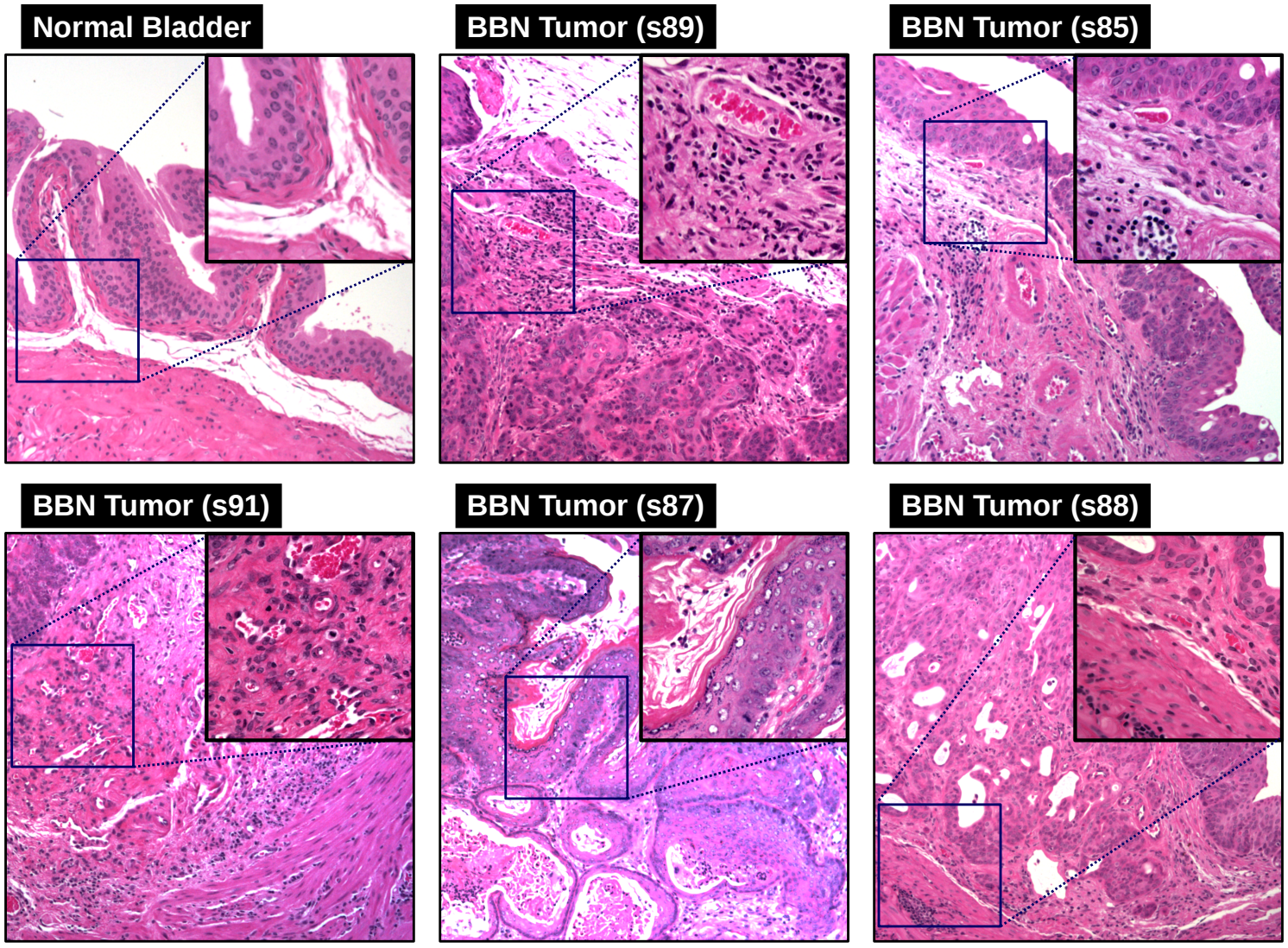
Supplementary Table ST1

Sample	Cellularity	Ploidy
s82	0.575	2.10
s83	0.580	2.10
s84	0.560	2.10
s85	0.510	2.15
s86	0.585	2.10
s87	0.495	2.10
s88	0.580	2.10
s89	0.605	2.10
s90	0.585	2.15
s91	0.570	2.10

Cellularity and Ploidy values of 10 BBN-induced mouse baldder tumors determined via the sequenza pipeline. BAM files from each tumor were analyzed against each of two controls independently, and resulting values were averaged. The following cellularity and ploidy ranges were tested: cellularity, from 0.1 to 1.0, with a step of 0.01; ploidy, from 0.4 to 3.8, with a step of 0.1.

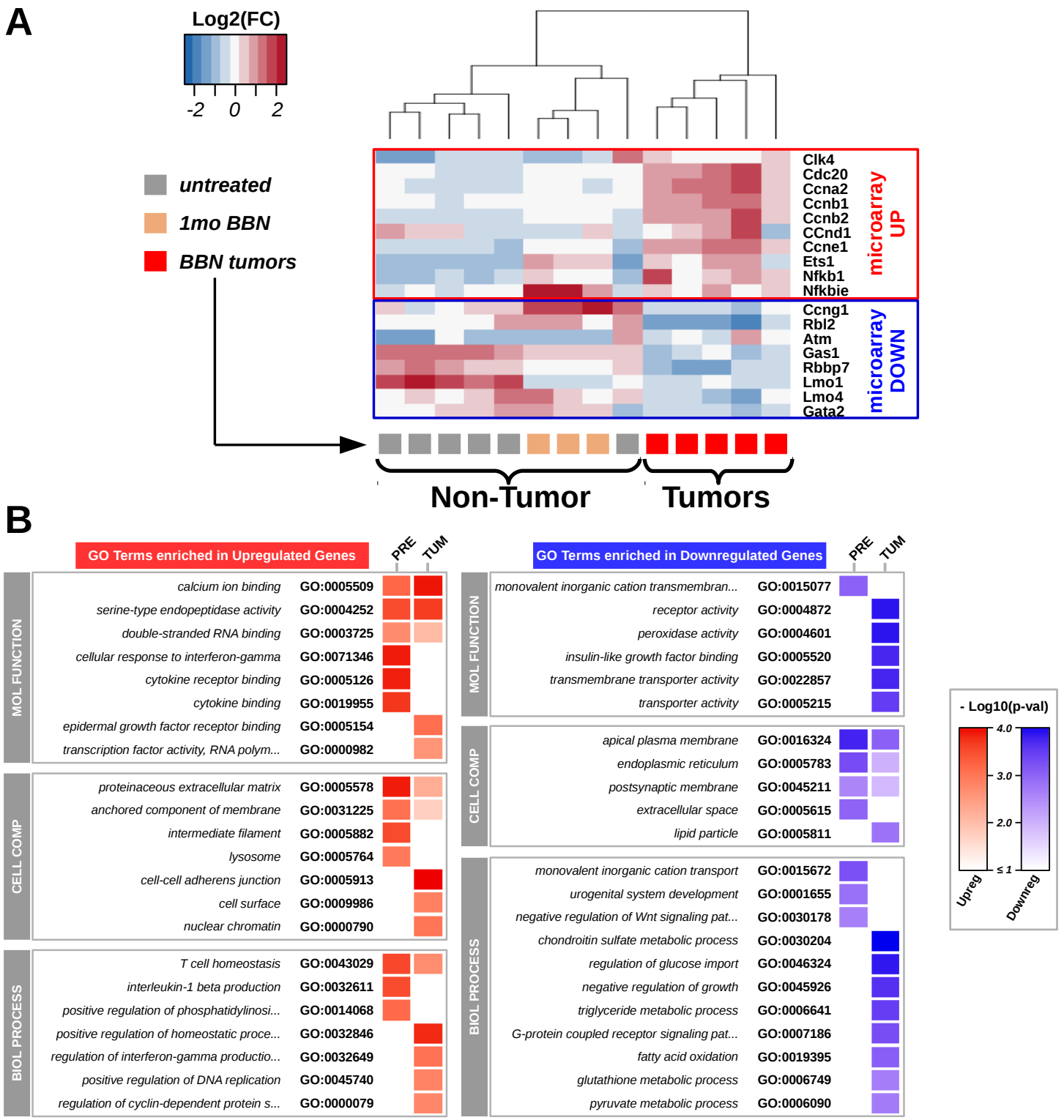
Supplementary Figure S1





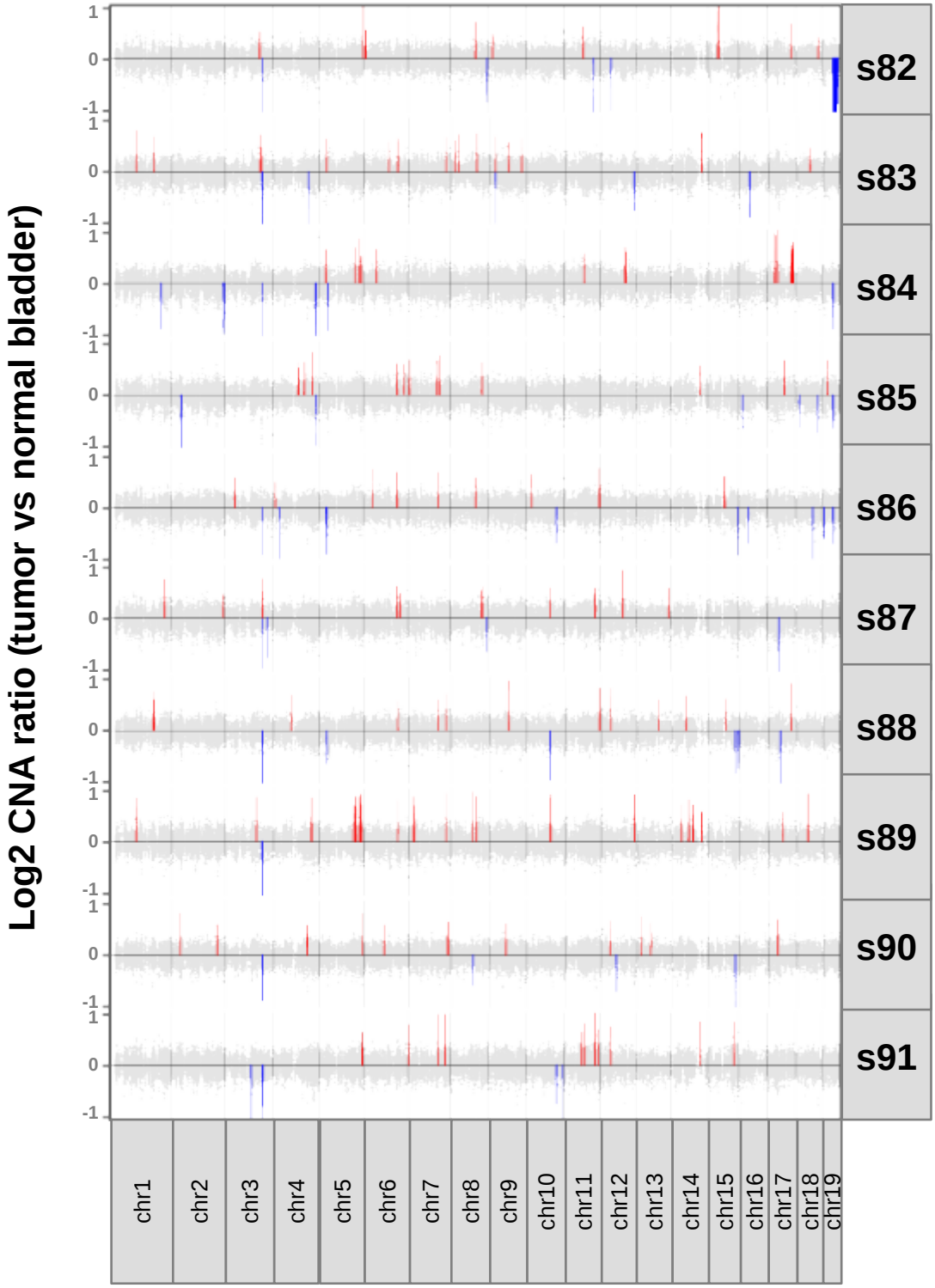
Histology of BBN bladder tumors. (A) Diagram summarizing the BBN treatment protocol, including a representative picture of the mouse bladder at sacrifice after 20 weeks of control or BBN treatment (left). (B) Hematoxylin and eosin stains of a normal mouse bladder and 5 BBN-induced tumors (s85, s87, s88, s89, and s91). BBN tumors show widespread histological abnormalities, including infiltrating cells and cancer cells invading the muscular layer of the bladder. These results are in good agreement with work by Shin K et al (Nature Cell Biology, 2014. 16,469–478), Van Batavia J et al (Nature Cell Biology, 2014. 16, 982–991) and Liang Y et al (Sci Rep, 2016. Jul 5;6:29479).

Supplementary Figure S2



Mutational background of bladder tumors. (A) Heatmap showing expression levels (median-based z-scores) of a list of genes previously found to be differentially expressed in BBN-induced tumors compared to normal bladders by microarray analysis (*Neoplasia. 2004 Sep; 6(5): 569–577*). With the exception of *Atm*, all previosuly identified DE genes were confirmed to be differentially expressed in the BBN tumors by RNA-seq. (B) Extended heatmap summarizing GO terms enriched in the lists of differentially expressed genes in the pre-cancerous bladders against controls or in the BBN-induced tumors against controls. Color intensity tracks with the negative logarithm of the corresponding p-val (classic Fisher's exact test).

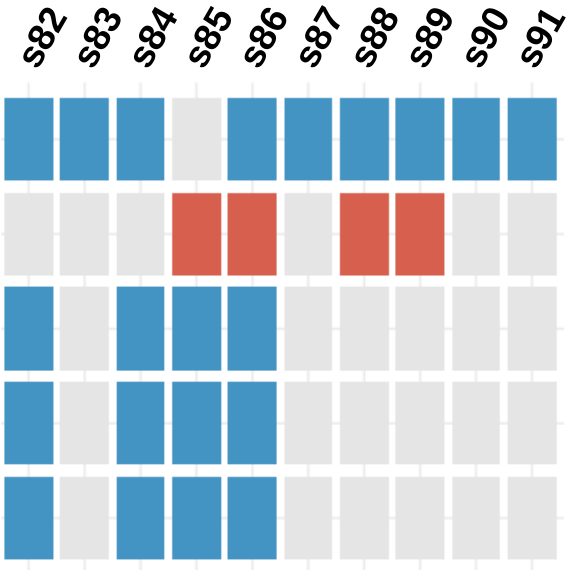
Supplementary Figure S3



Copy number alterations in the BBN tumors. Copy number alterations (CNA) were computed using the control-FREEC software. Each track displays the Log2 CNA ratio between tumor sample and normal bladder controls. Signal was obtained by averaging results obtained running a BBN tumor against each of two control genomes. Significance was assessed by both Wilcoxon and Kolmogorov-Smirnov tests and using the tools included in the control-FREEC package. Potentially altered regions in somatic chromosomes were highlighted in red (potential amplification) or blue (potential deletion).

Supplementary Figure S4

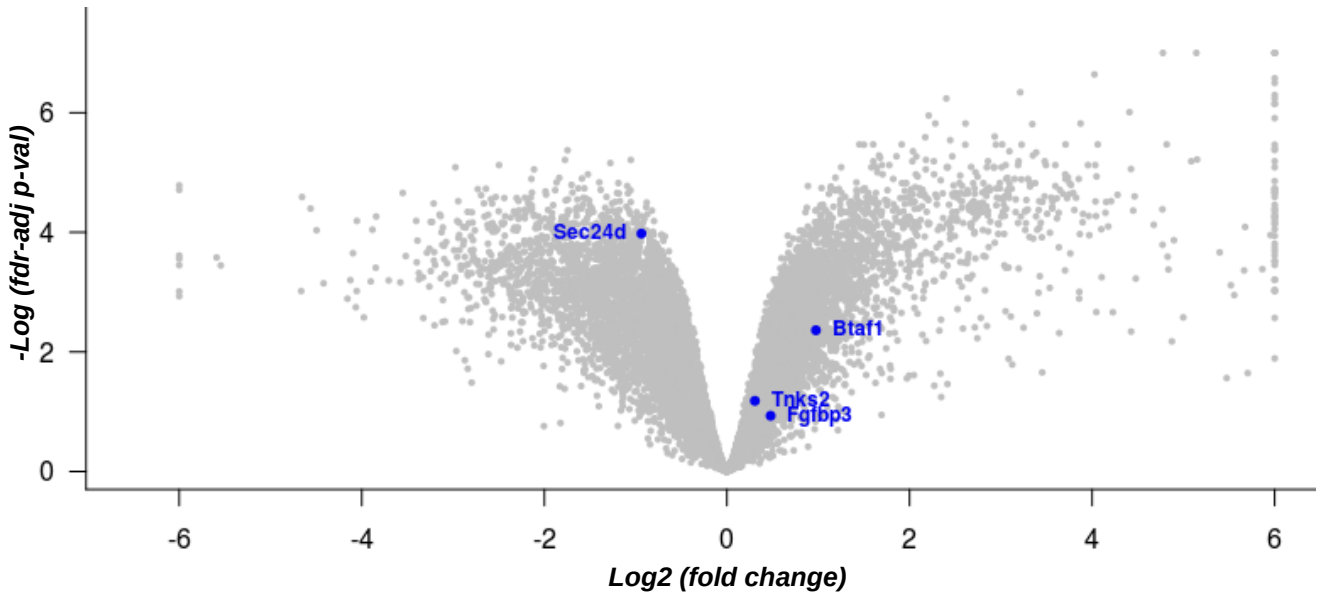
A



CHR	Start	End	Symbol
chr3	*123267496*	*123365636*	**Sec24d**
chr6	*114131241*	*114249886*	***Slc6a11***
chr19	*36834232*	*36893477*	**Tnks2**
chr19	*36917550*	*36919599*	***Fgfbp3***
chr19	*36926079*	*37014057*	***Btaf1***

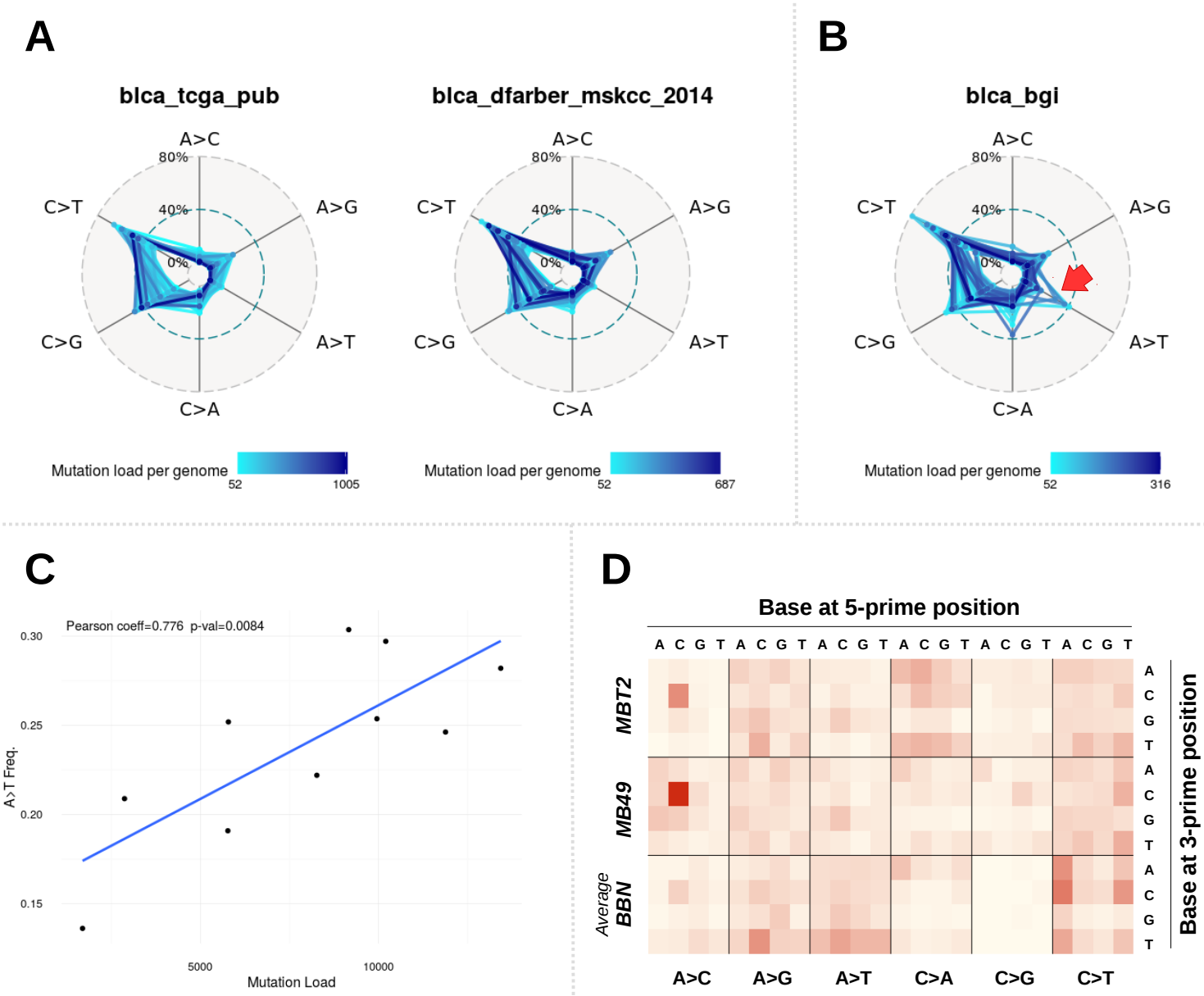
CNA *Deep loss* *Shallow loss* *Low-level gain* *High-level gain*

B



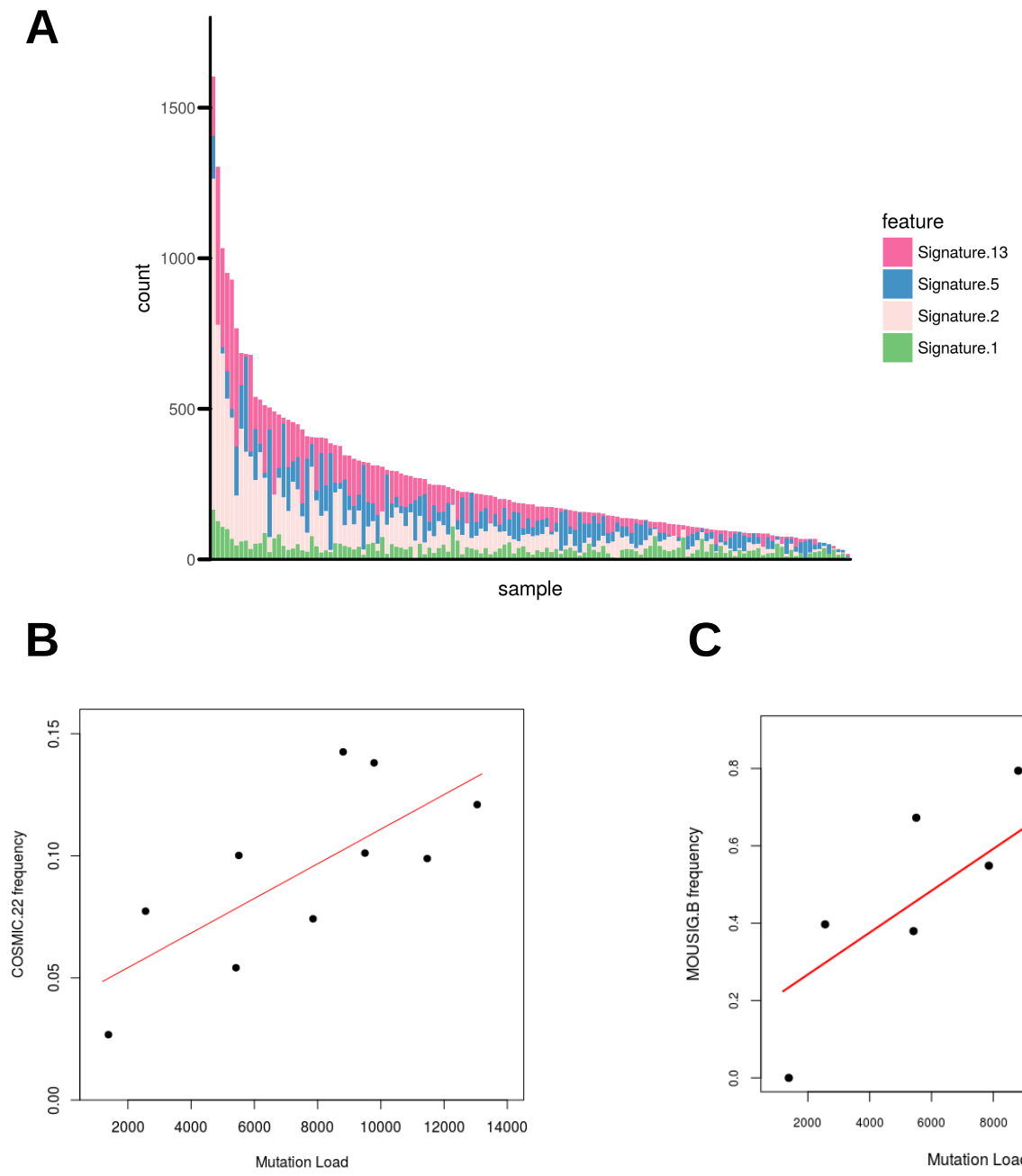
Putative gene amplifications and deletions across BBN tumors. (A) Heatmap summarizing genes with putative CNA by *control-FREEC* in at least four BBN tumor genomes. Gene symbols are displayed as rows and tumor samples as columns. Red squares indicate DNA gain (only low-intensity gains were found), while blue cells represent putative DNA loss (only shallow losses were found). (B) Volcano plot summarizing expression levels of the top five CNA genes (*Sec24d, Slc6a11, Tnks2, Fgfbp3, and Btaf1)* in the BBN tumors compared to control bladders. Blue points correspond to genes with a putative shallow deletion. *Sec24d* was found associated to a shallow deletion in 9/10 BBN tumors. Consistenlty, its expression levels were reduced in tumors compared to controls by RNA-seq (log2 (fold change)= -0.93; fdr-adjusted p-val=0.000105). Despite being often associated to shallow deletions, *Btaf1*, *Tnks2,* and *Fgfbp3* levels were not downregulated in the BBN tumors. *Slc6a11* gene expression was below the detection threshold in both normal bladders and tumors, and hence the gene was not displayed in the plot. To our knowledge, none of these genes has been strongly implicated in cancer development or progression.

Supplementary Figure S5



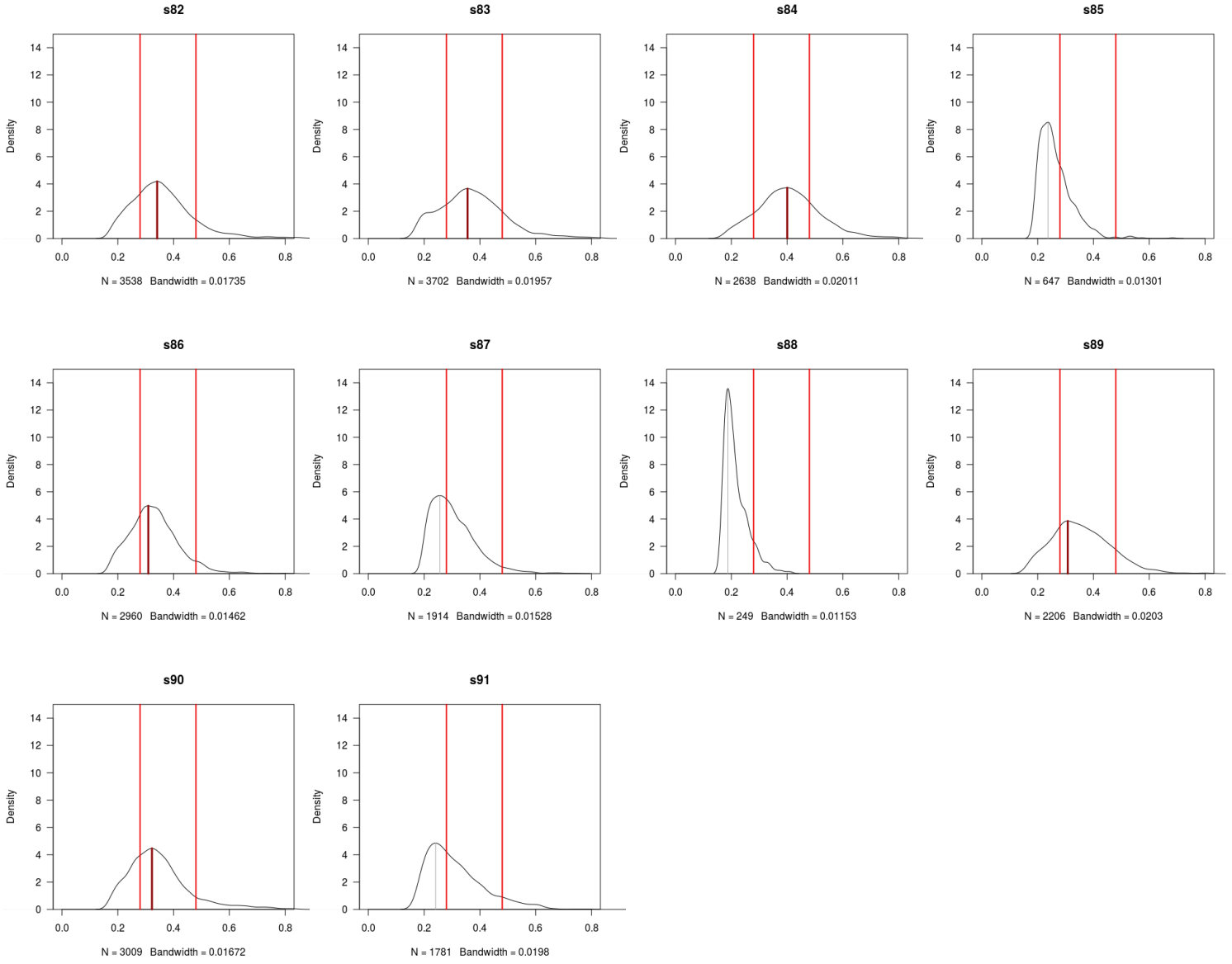
Mutational background of bladder tumors. A,B) Frequency of nucleotide conversion in 30 random genomes from TCGA (pub 2014), MSKCC-DFARBER (2014) (A), and BGI (B) datasets are displayed in the radar charts. Line color corresponds to mutation load. A red arrow indicates those samples with a mutational profile similar to the BBN tumors (BGI dataset). C) Scatter plot showing the relative frequency of A>T transversions (y-axis) plotted against sample mutation loads (x-axis) in the BBN tumor genomes. Pearson Correlation coefficient and p-value were calculated using the cor.test function from the R *stats* package. D) Heat map showing frequencies of nucleotide conversions within the tri-nucleotide context from averaged BBN tumor genomes or mouse bladder cancer cell lines (MBT2 and MB49). Conversions are grouped by mutation type (A>C, A>G, A>T, C>A, C>G, C>T), with different 5' bases organized column-wise and 3' bases organized row-wise.

Supplementary Figure S6



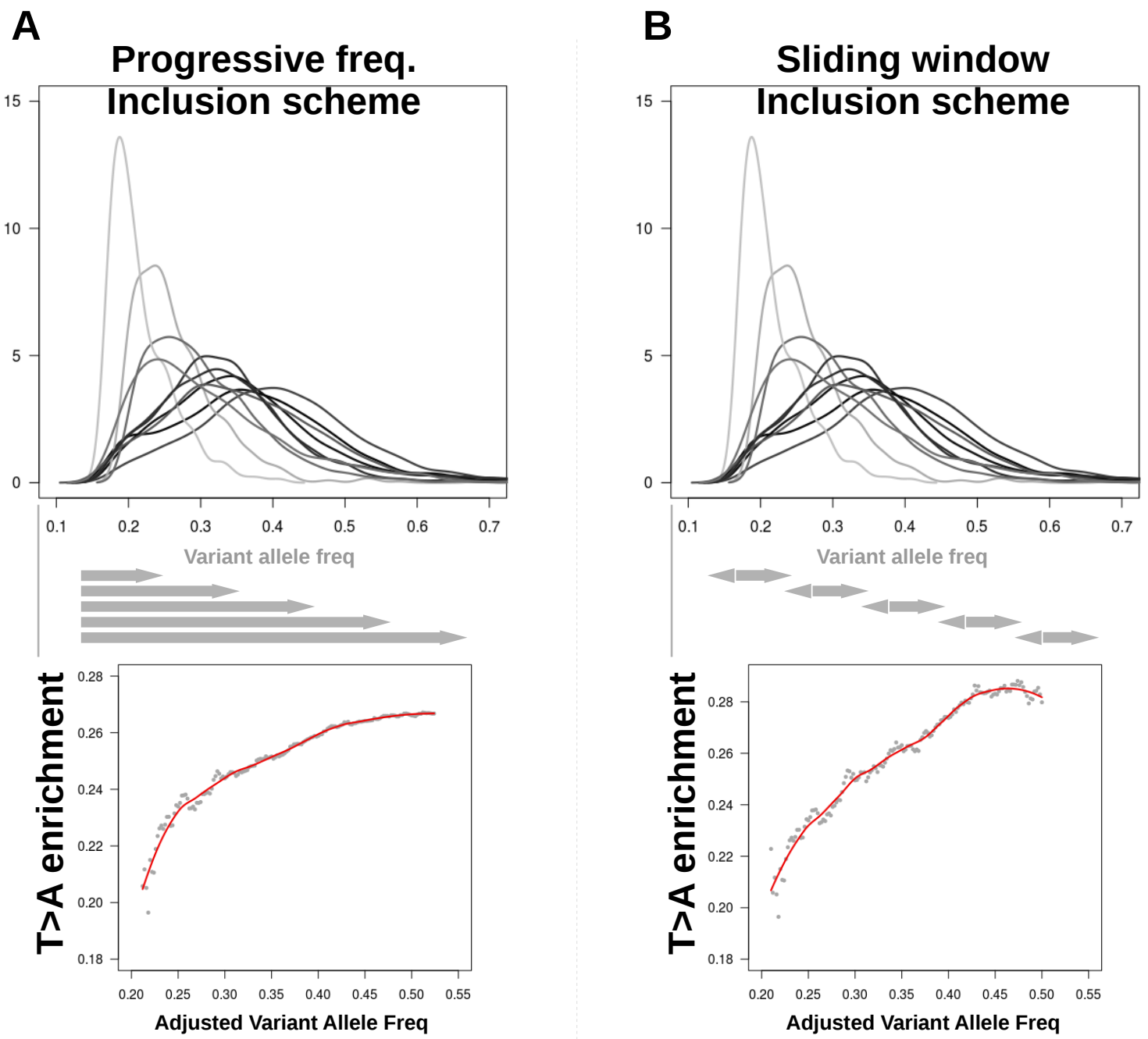
COSMIC mutational signatures contributing to human bladder cancer. (A) The barplot shows the contribution of different COSMIC mutational signatures (COSMIC signatures #1,2,5,13) to the total number of mutation in each human bladder cancer from the TCGA (pub 2014) dataset. (B, C) BBN mouse genomes were deconvoluted against COSMIC signatures 1,2,5,13,22 (B) or the signatures that were *de novo* extracted from BBN genomes (C). Scatter plots show the relative contribution of mutations imputed to [T>A]-rich signatures (B, COSMIC 22; C, MOUSIG.B) plotted against sample mutation loads (x-axis). Each point corresponds to a BBN tumor genome. Positive correlation is highlighted by trendlines (red lines).

Supplementary Figure S7



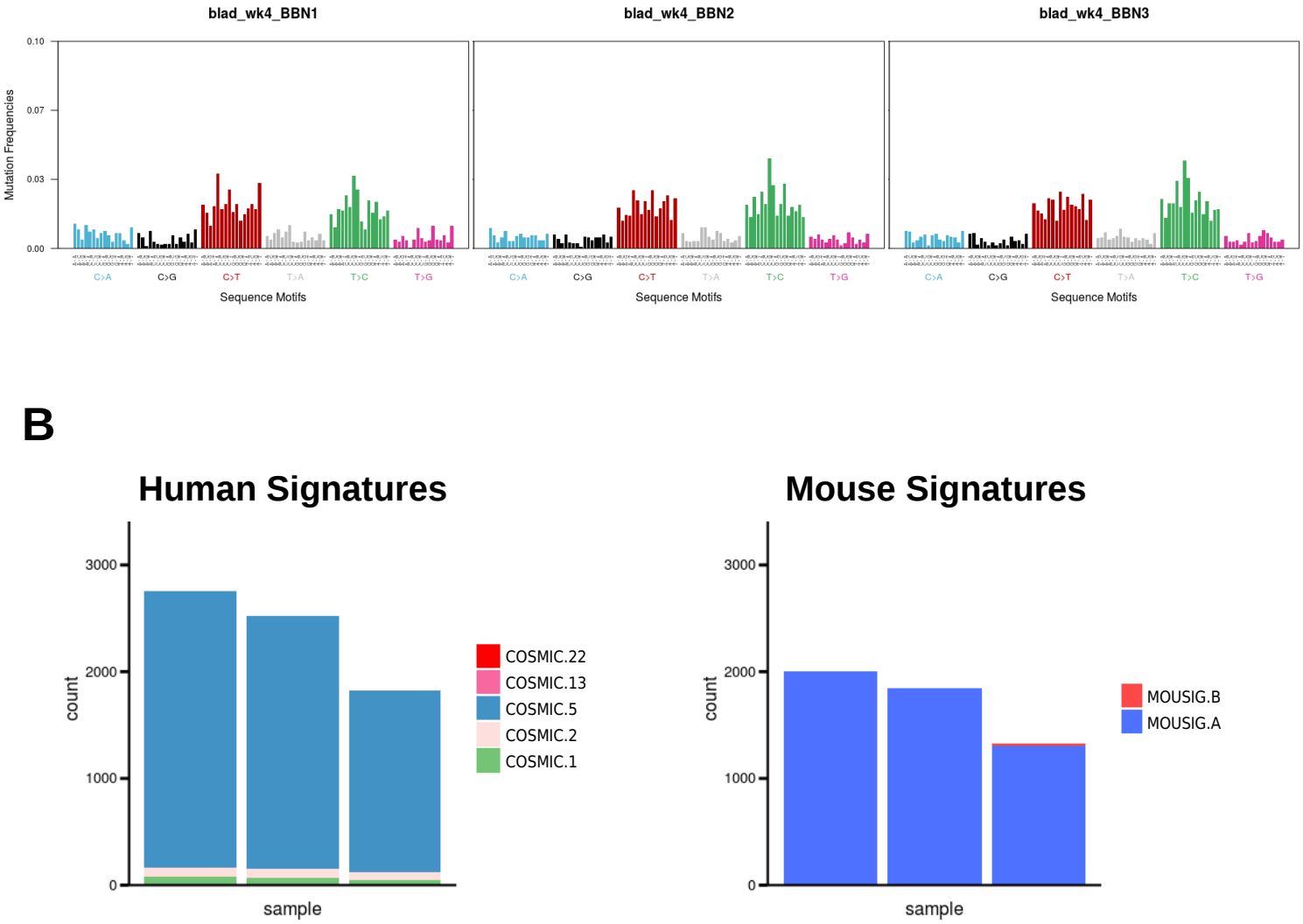
Intra-tumor variant-allele frequencies in BBN tumors. Distributions of intra-tumor variant frequencies across BBN tumors are shown by density plots. Red lines indicate frequency values of 0.28 and 0.48. For each distribution, mode (value that occurs most frequently) is indicated by a magenta line (mode >= 0.28) or a gray line (mode < 0.28). Increases in mode tracks with less skewed variant allele frequency distributions.

Supplementary Figure S8



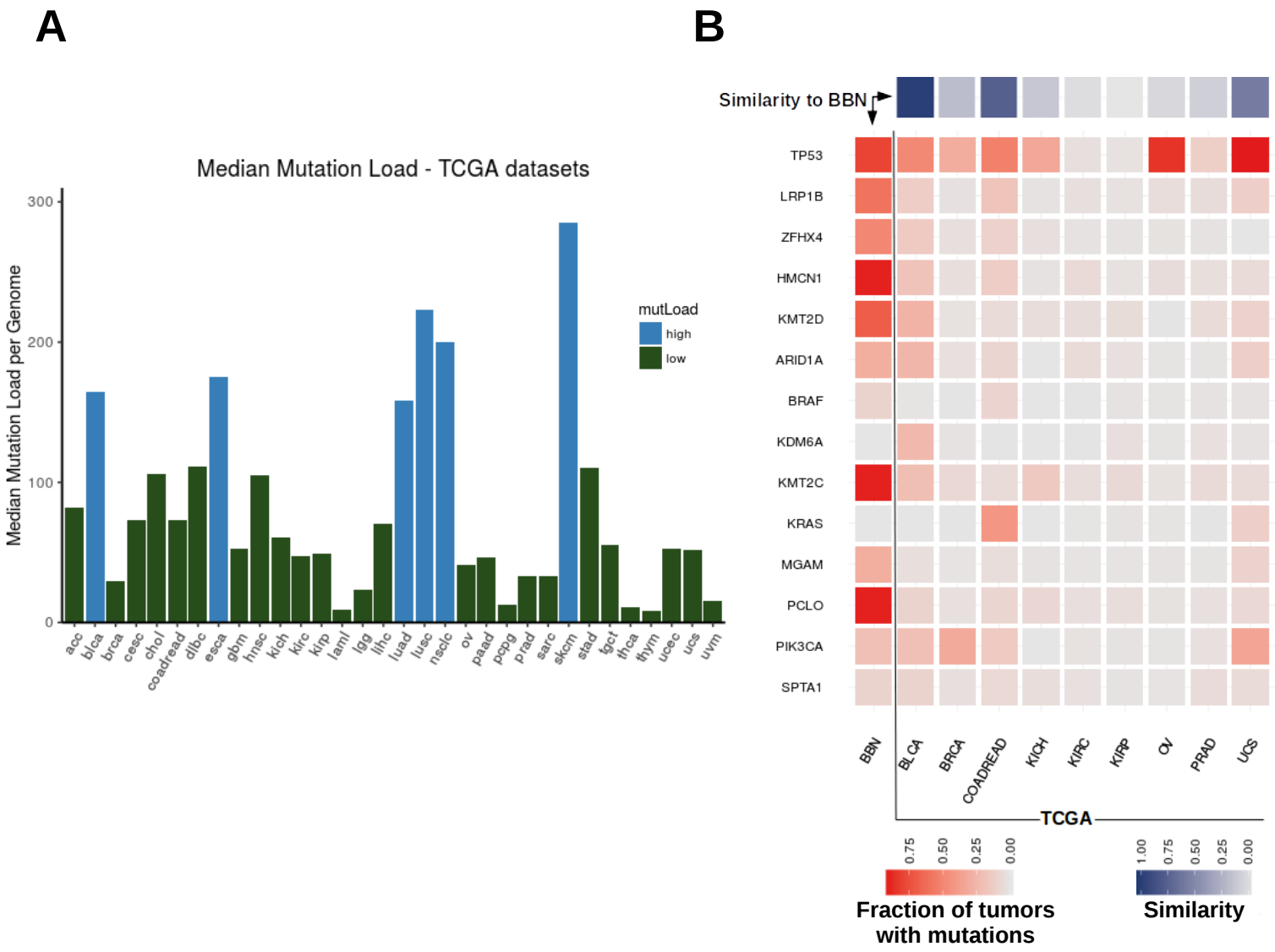
High intra-tumor frequency mutations are enriched in T>A mutations. A, B) Top. Density plot showing the distributions of intra-tumor variant frequencies for each BBN tumor, underlain by a diagram summarizing the data sub-setting scheme. Analyses were performed as described in Material and Metods. Briefly, mutation data from each BBN tumor were subset to include only mutations with adjusted intra-tumor frequencies lower or equal than a progressively sliding threshold (A). Alternatively, the subsetting was performed by including only mutations having intra-tumor frequencies included in a fixed-width sliding frequency window (B). At each iteration, the threshold-selected mutations were counted and relative [T>A] abundance was returned for each dataset. Bottom. Relative median [T>A] ratio values where plotted against the frequency value used as upper threshold in the sub-setting procedure. Trendlines were computed using LOESS (red lines).

Supplementary Figure S9



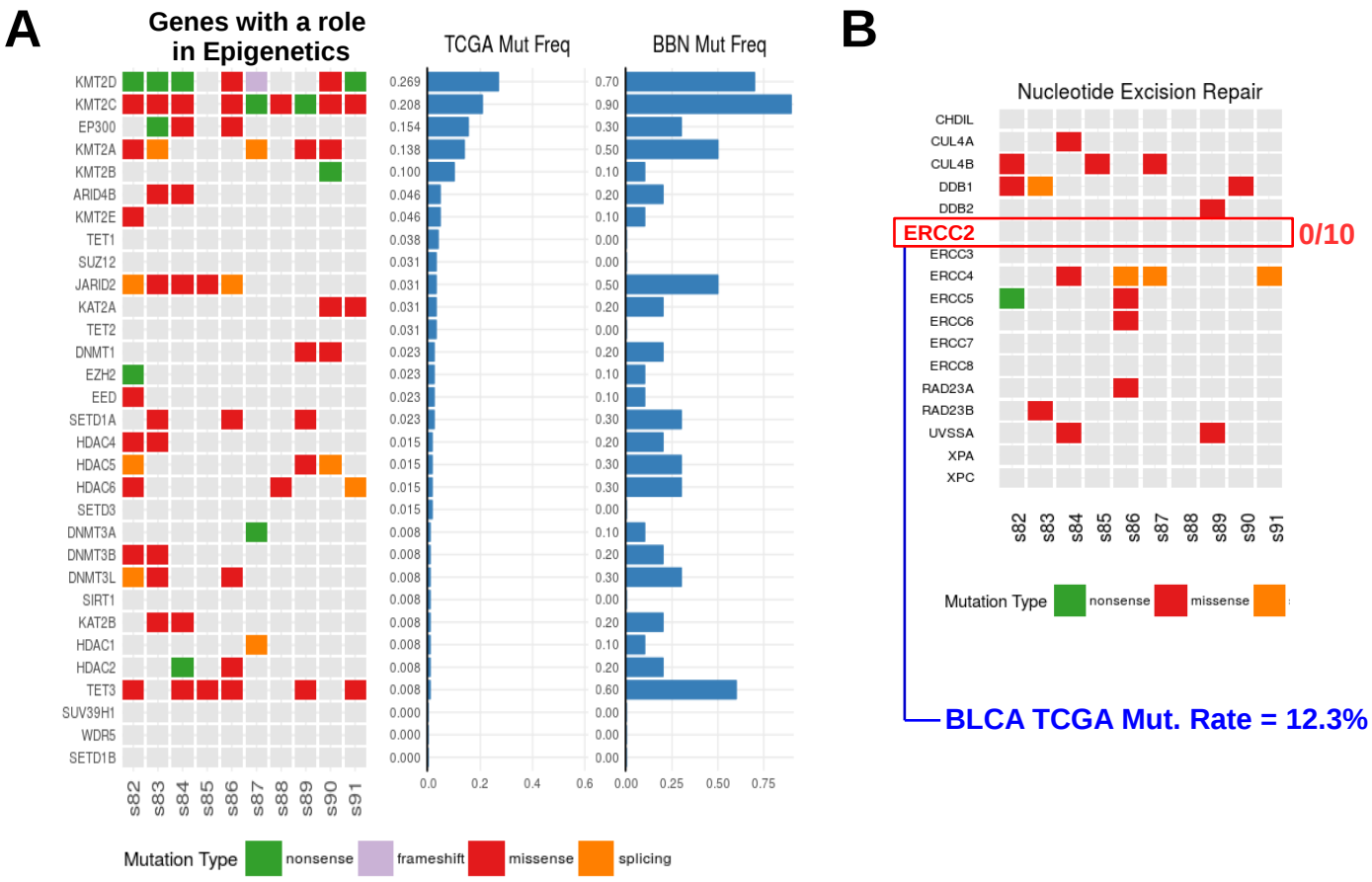
Mutation profiles of mouse bladders at 4 weeks of BBN treatment. A) Mice were treated for 4 weeks with BBN in drinking water and then sacrificed. Bladders were harvested and used for total RNA preparation. After rRNA depletion and library preparation, samples were submitted for RNAseq. Mutations detected in the RNA sequences were cleaned by removing background from RNA preparations from normal bladders as well as the SNP filters used for genomic DNA (SNPdb 146, normal bladder WES data). Genomic regions corresponding to immunoglobulin, T-cell receptor, or other immune-related genes including hyper-variable regions were excluded. The tri-nucleotide profile of the remainder mutations is displayed. (B) Contributions of COSMIC or de novo extracted MOUSIG mutational signatures to mutational catalogs obtained by RNAseq analysis of mouse bladders at 4 weeks of BBN treatment. Each bar represents a BBN tumor and the vertical axis denotes the number of mutations.

Supplementary Figure S10



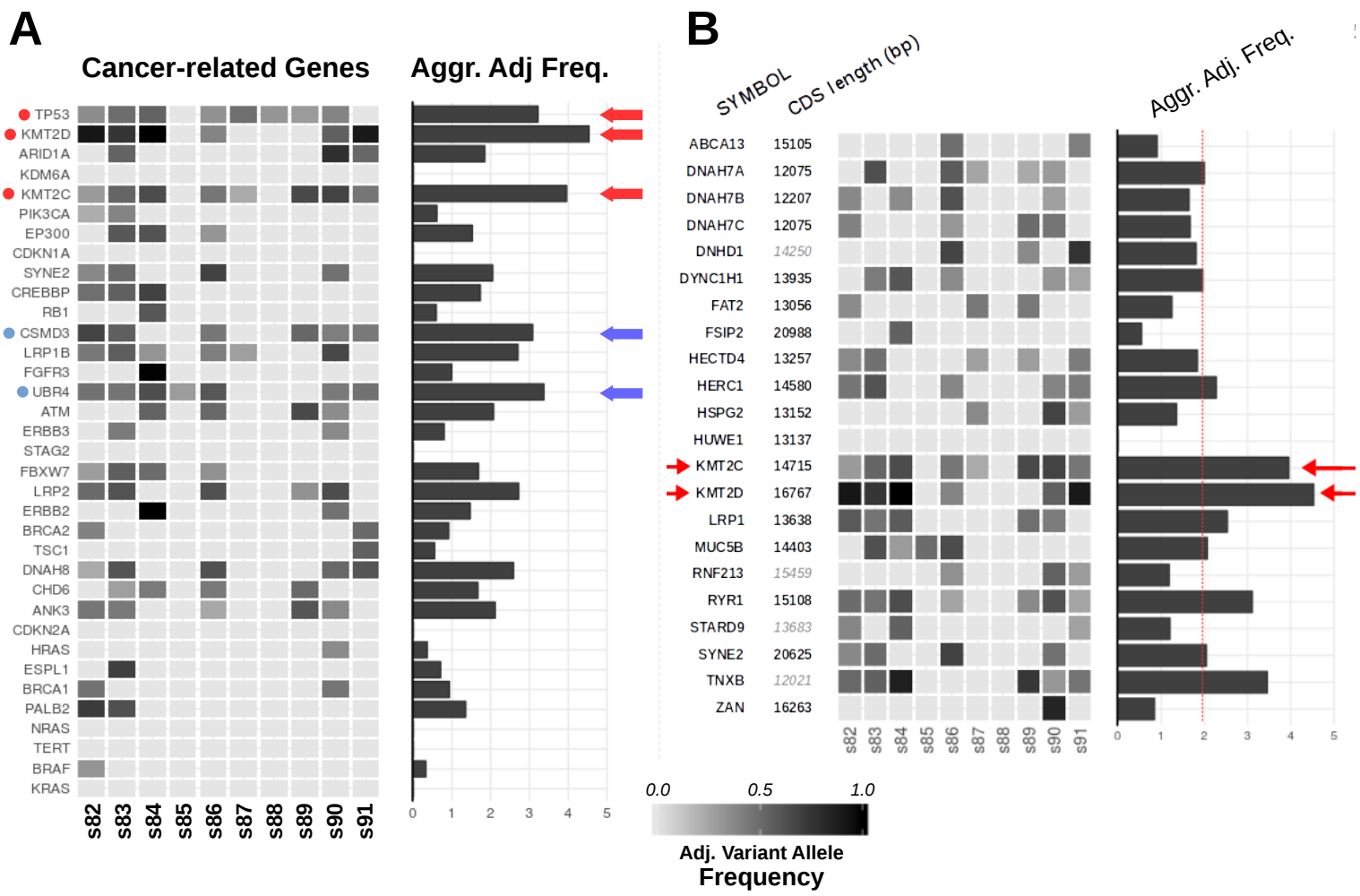
Comparison of mutations found in BBN mouse tumors with human TCGA cancers. A) Mutation from different TCGA datasets were retrieved using TCGAretriever, for each tumor the total number of mutations was computed and the median per cancer type is displayed in the barplot. Cancers with a median mutation load comparable to bladder cancer (+/- 20%) or higher are displayed as blue bars (high mutation load cancers). B) Heatmap displaying mutation frequencies of 15 commonly mutated genes across the BBN dataset, TCGA bladder cancer dataset, and 8 bladder-relaed cancers from TCGA (red). Top. Similarity of each TCGA dataset to the BBN dataset was computed and displayed via a 1-row heatmap (purple). BLCA, bladder cancer; BRCA, breast cancer; COADREAD, colorectal adenocarcinoma; KICH, Kidney Chromophobe; KIRC, Kidney Renal Clear Cell Carcinoma; KIRP, Kidney renal papillary cell carcinoma; OV, ovarian cancer; PRAD, prostate adenocarcinoma; UCS, Uterine Carcinosarcoma.

Supplementary Figure S11



Mutations in genes encoding for Epigenetics and DNA Repair factors. (A, B) Tile charts displaying alterations in genes encoding for epigenetic (A) or nucleotide excision repair (B) factors in 10 BBN tumor genomes (green, nonsense; purple, frameshift; red, missense; orange splicing site; gray, no mutation). (A) Overall fraction of genomes with mutation in each gene is displayed in the barplots for BBN dataset and human TCGA bladder cancer dataset. (B) The row corresponding to ERCC2 gene (0/10 mutations) is highlighted by a red box. The corresponding mutation frequency in genomes from the TCGA bladder cancer dataset is reported (blue).

Supplementary Figure S12

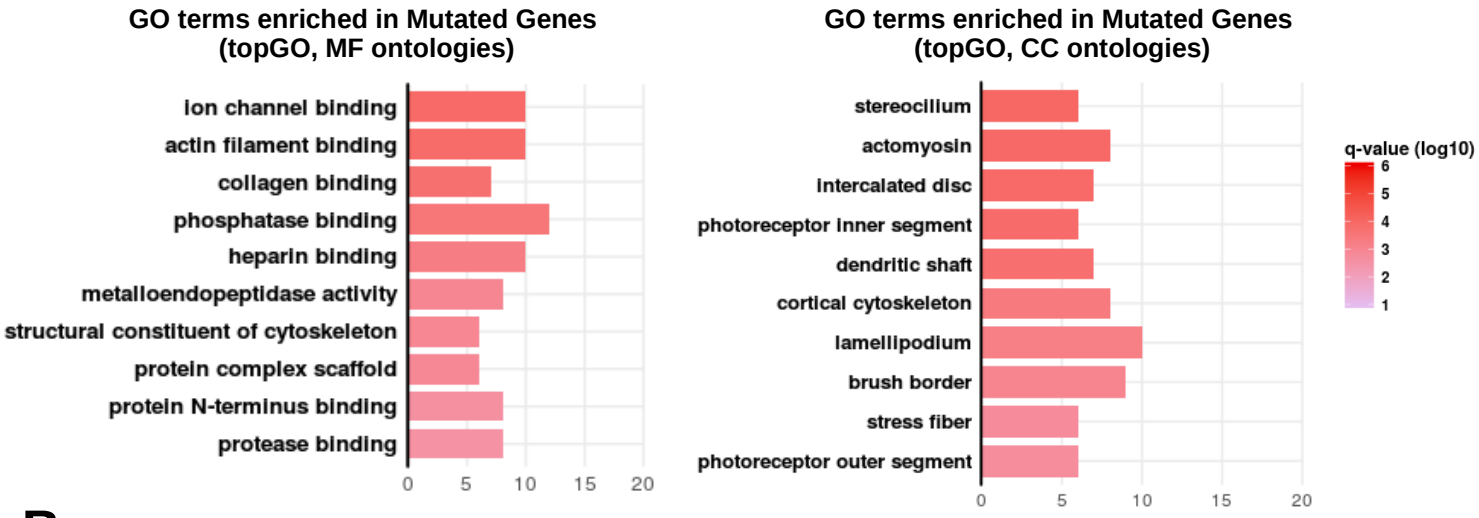


Intra-tumor mutation frequencies of cancer-related and long-CDS genes. (A) Heatmap displaying intra-tumor mutation frequencies of a list of cancer genes in BBN tumor genomes. For each gene and each sample, the variant (insertion/deletion mutations were excluded) with the highest intra-tumor mutation frequency was found (left). A gene-wise aggregate score was computed by summing the intra-tumor frequency values and displayed in the barplot (right). Points and arrows highlight genes with (red) or without (blue) a previously reported causative role in cancer and having an aggregate frequency score similar or higher than *TP53*. (B) A gene set was selected to include mouse genes whose CDS lengths were similar to those of Kmt2c and Kmt2d (range: 12,000 bp to 21,000 bp; *KMT2C*: 14715 bp; *KMT2D*: 16767 bp), in order to test if gene length had an effect on the accumulation of mutations with high intra-tumor frequency. CDS length was retrieved from the NCBI Consensus CDS (black font) or Ensembl (gray, italic font) databases. The heatmap displays intra-tumor mutation frequencies of each gene in the list and in each BBN tumor genome (left). A gene-wise aggregate score was computed by summing intra-tumor frequency values and is displayed in the barplot (right). *Kmt2c* and *Kmt2d* (right arrows) have the highest aggregate scores, which are at least two-fold higher than the majority of other genes included in the set. While it is reasonable to assume that gene length could associate with the likelihood of accumulating passenger mutations in tumors, *Kmt2c* and *Kmt2d* are mutated more often and at higher intra-tumors frequencies than other genes with similar or bigger CDS length in BBN tumors. This observation is consistent with our hypothesis that at least some of the mutations in the *Kmt2c* and *Kmt2d* genes might be oncogenic drivers of tumorigenesis in the BBN model.

Supplementary Figure S13

A

Genes mutated (any mutation) in at least 5 BBN tumors (n = 474)



B

High-impact Mutations *(nonsense/frameshift only)*

ENTREZID	SYMBOL	TUMORS
109032	Sp110	6
212427	A730008H23Rik	6
381280	Hjurp	6
381022	Kmt2d	5
107831	Adgrb1	5
12287	Cacna1b	4
329942	Csmd2	3
110880	Scn4a	3
207304	Hectd1	3
244895	Peak1	3
67016	Tbc1d2b	3
14800	Gria2	3
11426	Macf1	3
59057	Zfp24	3
18189	Nrxn1	3
231051	Kmt2c	2
235626	Setd2	2
18193	Nsd1	2
...	88 more	2

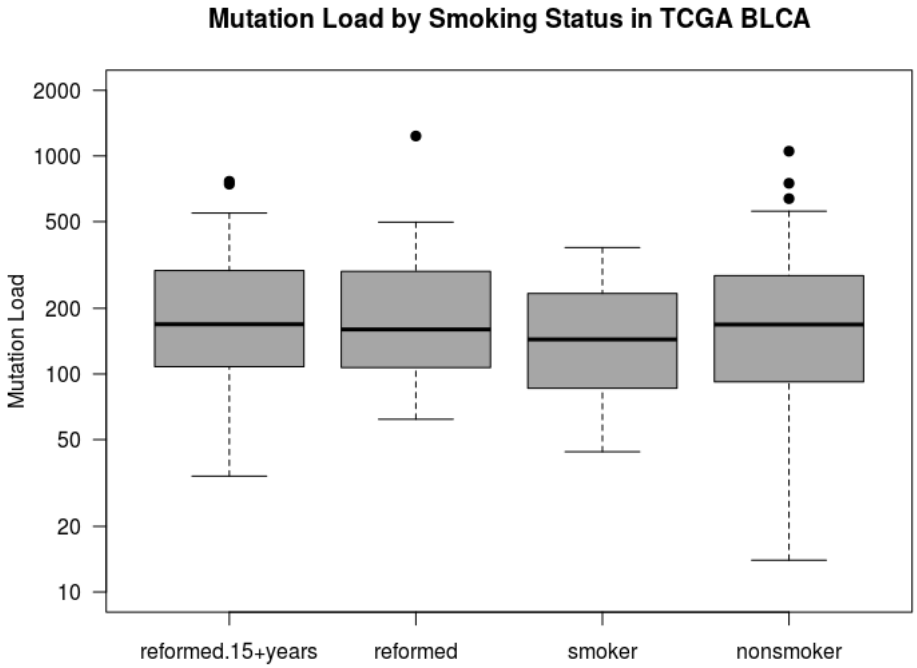
Pathway enrichment analysis of genes frequently mutated in BBN tumors. (A) Pathway enrichment analysis of commonly mutated genes in BBN tumors (mutation in at least 5 BBN tumors) against Gene Ontology (GO) terms using topGO. Terms in the “Molecular Function” and “Cellular Component” ontologies are shown. Bars indicate the number of genes (gene count) identified in the set of commonly mutated genes and belonging to each GO term. Bar color corresponds to -Log10(fdr-adjusted p-values). (B) Table summarizing the genes carrying high-impact (nonsense, frameshift) mutations in at least 2 BBN tumors. Entrez identifier, official symbol identifier and number of tumors carrying an high impact mutation are reported. Genes belonging to the ReactomeDB term “PKMTs methylate histone lysines” are highlighted in red. 88 genes with high-impact mutations in 2 BBN tumors were omitted.

Supplementary Figure S14

A

Sample ID	Mut Load	Staging	Histology
s88	1970	T2	solid
s85	3094	T1	papillary
s87	5972	T3	solid
s91	5977	T4	papillary
s89	8581	T3	papillary
s84	9334	T2	squamous
s86	10150	T2	papillary
s90	10422	T2	solid
s82	12092	T4	solid
s83	13614	T3	papillary

B



Association between Mutation Load and Clinical Features in mouse and human bladder tumors. (A) Table summarizing Mutation Load and clinical features of ten BBN-induced tumors. Sample identifier, total number of variants (mutation load), Staging (T1-T4) and histology (solid, papillary, squamous) class are reported. (B) Barplot showing the number of NSVs found in bladder tumors in four segments of the TCGA bladder cancer patient population. Patients were stratified according to their smoking status: never-smokers (less than 100 cigarettes in the lifetime), current smokers, reforemd smokers (who stopped smoking in the last 15 years), or reformed smokers since 15+ years.