
Article:
Hurst, CD orcid.org/0000-0001-7719-1325, Alder, O orcid.org/0000-0002-7975-9727, Platt, FM et al. (15 more authors) (2017) Genomic Subtypes of Non-Invasive Bladder Cancer with Distinct Metabolic Profile and Female Gender Bias in KDM6A Mutation Frequency. Cancer Cell, 32 (5). pp. 701-715. ISSN 1535-6108

https://doi.org/10.1016/j.ccell.2017.08.005

Table S1 related to Figure 1. Details of patients, tumor samples and analyses carried out.
Provided as an Excel File.

Table S2 related to Figure 2. Genes differentially expressed between tumors belonging to subtypes GS1 and GS2.
Provided as an Excel File.

A

SAMPLE ID	GENOMIC SUBTYPE	LUND KNN1	UROMOL 117 CONSENSUS CLUSTER
66	1	UROA/MS1a	1
239	1	UROA/MS1a	1
672	1	UROA/MS1a	1
705	1	UROA/MS1a	1
811	1	UROA/MS1a	1
894	1	UROA/MS1b	1
1017	1	UROA/MS1a	1
1167	1	UROA/MS1a	1
1177	1	UROA/MS1a	1
1361	1	UROA/MS1a	1
1385	1	UROA/MS1a	1
1436	1	UROA/MS1a	1
1474	1	UROA/MS1a	1
1567	1	UROA/MS1a	1
1719	1	UROA/MS1a	1
1732	1	UROA/MS1b	1
1858	1	UROA/MS1a	1
1992	1	UROA/MS1a	1
2047	1	UROA/MS1b	1
2055	1	UROA/MS1a	1
2074	1	UROA/MS1a	1
2425	1	UROA/MS1a	1
2452	1	UROA/MS1b	1
494	1	UROA/MS1a	2
933	1	UROA/MS1b	2
1038	1	UROA/MS1b	2
1117	1	UROA/MS1a	2
1440	1	UROA/MS1b	2
1566	1	UROA/MS1a	2
1931	1	UROA/MS1a	2
2065	1	UROA/MS1a	2
2100	1	UROA/MS1a	2
2118	1	UROA/MS1a	2
578	1	UROA/MS1a	3
865	1	UROA/MS1a	3
1142	1	UROA/MS1a	3
1144	1	UROA/MS1a	3
1214	1	UROA/MS1a	3
1273	1	UROA/MS1a	3
1930	1	UROA/MS1a	3
419	1	SCC/MS2b2.2	2
390	1	INFIL/MS2b1	1
787	1	INFIL/MS2b1	1
2161	1	INFIL/MS2b1	1
1058	1	INFIL/MS2b1	2
1103	1	INFIL/MS2b1	2
2013	1	INFIL/MS2b1	2
2498	1	GU/MS2a1	2
675	2	UROA/MS1a	1
1083	2	UROA/MS1b	1
1313	2	UROA/MS1a	1
1514	2	UROA/MS1a	1
1971	2	UROA/MS1a	1
2334	2	UROA/MS1b	1
103	2	UROA/MS1a	2
934	2	UROA/MS1a	2
1063	2	UROA/MS1a	2
1067	2	UROA/MS1b	2
1181	2	UROA/MS1a	2
1213	2	UROA/MS1a	2
1283	2	UROA/MS1b	2
1307	2	UROA/MS1b	2
1387	2	UROA/MS1a	2
1390	2	UROA/MS1a	2
1453	2	UROA/MS1b	2
1467	2	UROA/MS1b	2
1798	2	UROA/MS1b	2
2032	2	UROA/MS1a	2
2104	2	UROA/MS1b	2
2220	2	UROA/MS1a	2
2338	2	UROA/MS1b	2
2063	2	UROA/MS1a	3
1335	2	SCC/MS2b2.2	2
1305	2	INFIL/MS2b1	1
437	2	INFIL/MS2b1	2
1289	2	INFIL/MS2b1	2
2086	2	INFIL/MS2b1	2
1932	2	GU/MS2a2	2
2082	2	GU/MS2a1	2

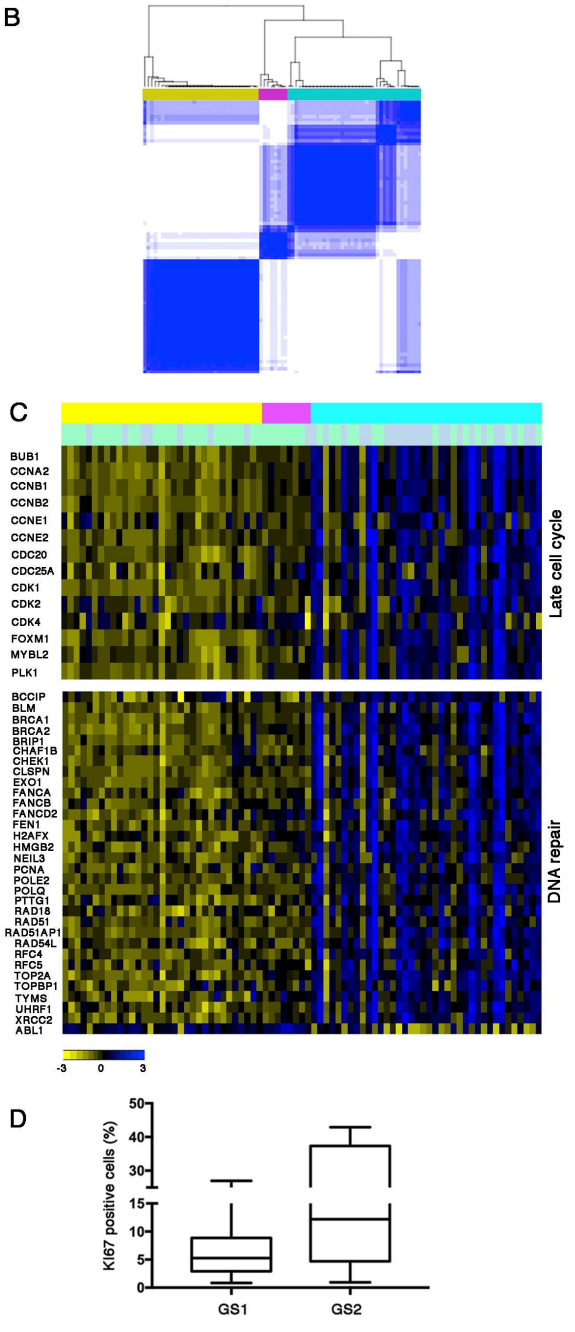


Figure S1 related to Figure 2. Alignment of GS1 and GS2 expression profiles with previously identified NMIBC expression subtypes.

(A) Tabular representation of tumor sample membership assignments based on; hierarchical clustering of copy number data (Genomic Subtype), nearest neighbour analysis results using Lund n=308 dataset as training set, or consensus clustering of microarray data after application of UROMOL classifier.

(B) UROMOL 117 gene classifier was applied to microarray data from Ta tumors (n=79), results are represented as consensus cluster plot (k=3).

(C) Heatmap representation of gene z-scores associated with late cell cycle and DNA repair (Bonferroni adjusted p value < 0.01).

(D) Boxplot representation of KI67 index (percentage of positive cells) according to GS1 and GS2 subtype membership. Each box has lines at the lower quartile, median quartile and upper quartile values. Whiskers represent the minimum and maximum values.

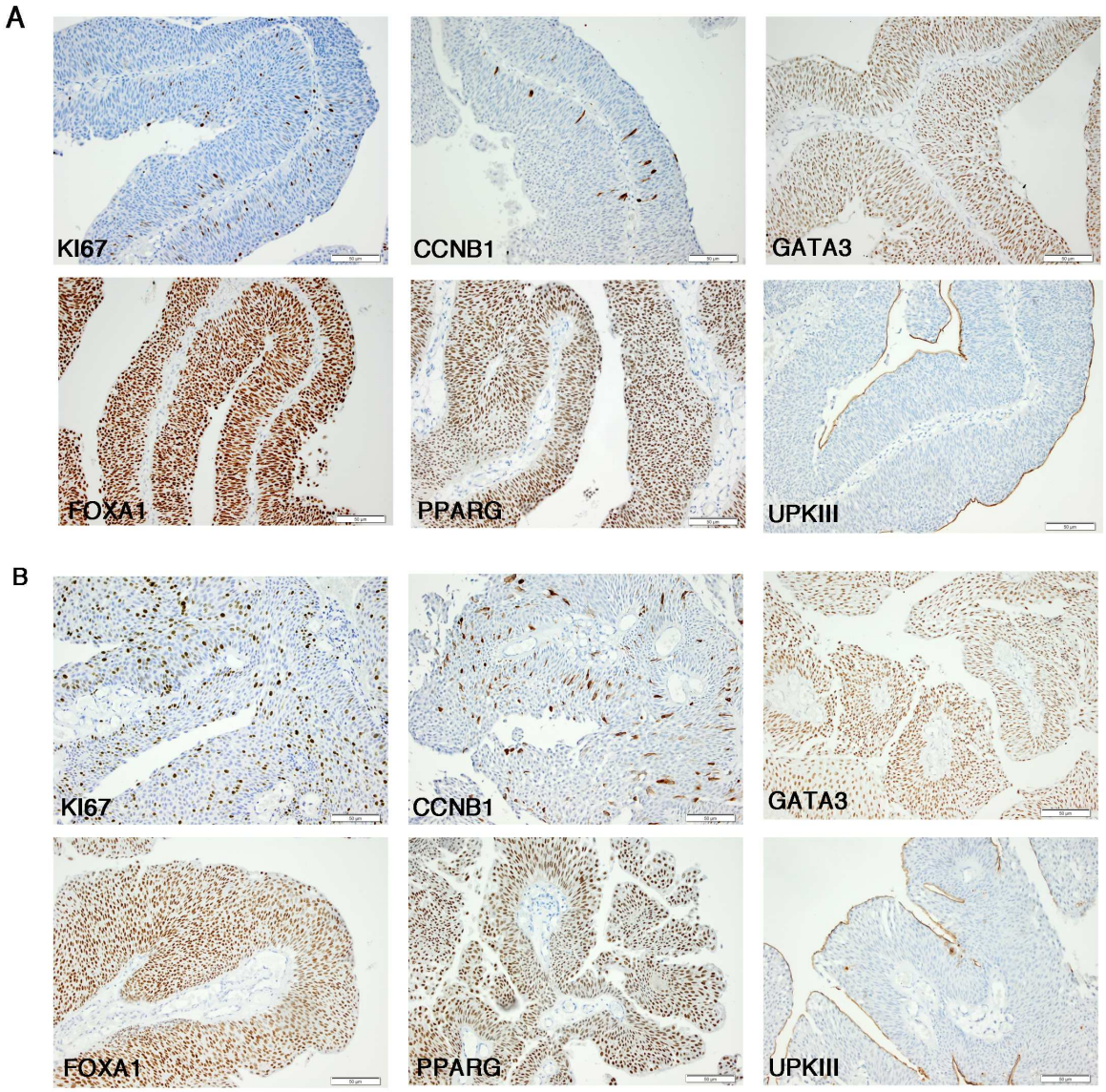


Figure S2, related to Figure 2. Immunohistochemistry for selected proteins in tumors from GS1 and GS2 subtypes.

(A-B) Expression of KI67, CCNB1, GATA3, FOXA1, PPARG and UPKIII in GS1 (A) and GS2 (B) subtype tumors. Scale bars: 50μm.

Table S3 related to Figures 3-6. Variants and APOBEC enrichment values.
Provided as an Excel File.

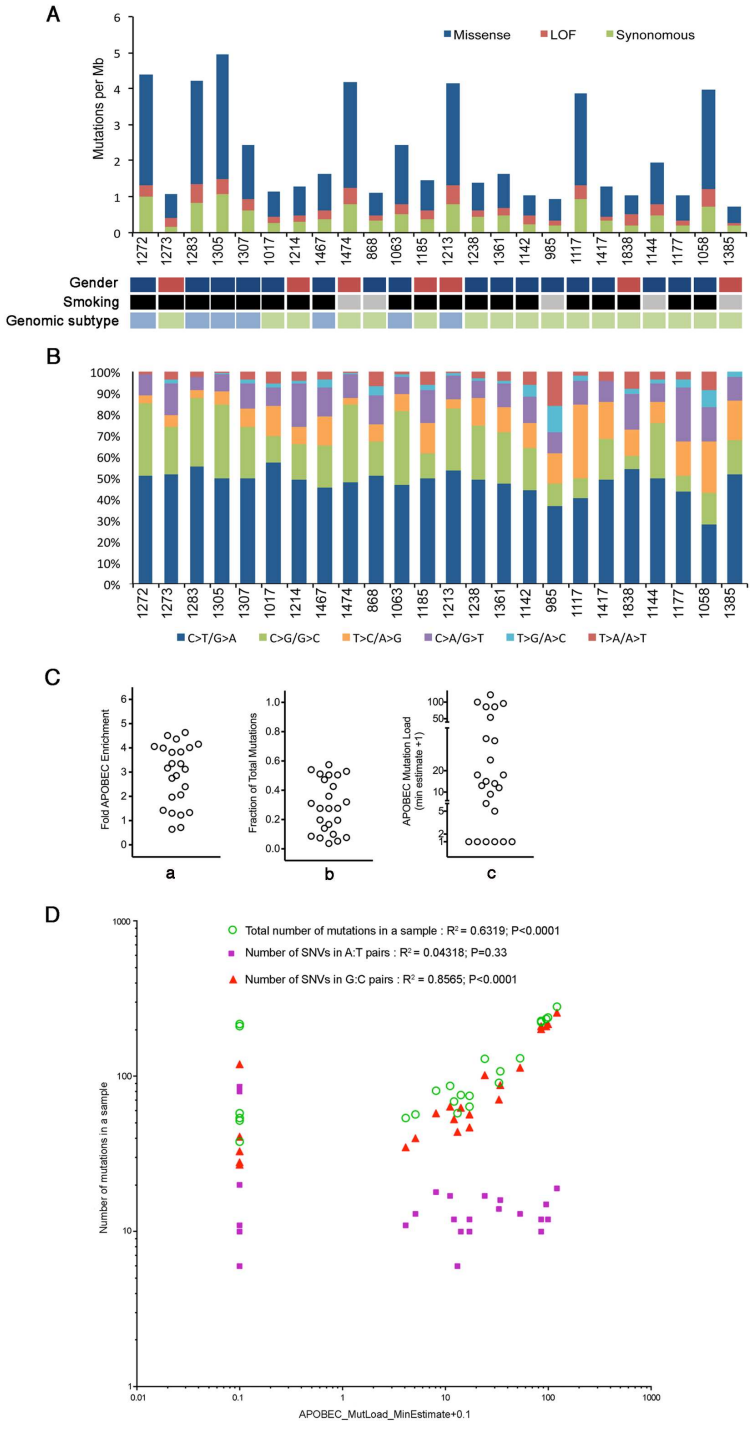


Figure S3, related to Figure 3. Distribution of mutation type and nucleotide substitutions in stage Ta bladder cancer exome sequences.

(A) Frequency and type of mutations in 23 stage Ta tumors and a cell line (LUCC8, patient 1838). Loss of function (LOF) mutations comprise nonsense, frameshift and splicing mutations. Shown below are

information on gender (red, female; blue, male), smoking status (black, ever-smoker; grey, never-smoker) and genomic subtype (green, GS1; blue, GS2).
(B) Nucleotide substitutions identified in bladder cancer exome sequences.
(C) Values characterizing APOBEC mutagenesis in the individual samples (all underlying vaues are in Table S3 sheet S3C).
a. Fold enrichment of the APOBEC mutagenesis signature over the expected occurrence for random mutagenesis in individual samples. Fold enrichment values are taken from the “APOBEC enrichment” column.
b. Relative load of APOBEC signature mutations. The Fraction of Total Mutations values were taken from the “APOBEC fraction of total mutations” column.
c. Minimum estimate of the number of APOBEC induced mutations in a sample. “APOBEC Mutation Load (minimum estimate)” is calculated only for samples passing 0.05 FDR threshold for APOBEC enrichment [”BH_Fisher_p-value_tCw”] =<0.05. Samples with “BH_Fisher_p-value_tCw” value greater than 0.05 receive a value of 0. “APOBEC_MutLoad_MinEstimate” is plotted on a logarithmic scale (with a pseudocount of 1) for better visualization of values in all sections of its range.
(D) APOBEC mutagenesis is the strongest source of mutations in NIBC.
Y- axis shows the log10-transformed total numbers or numbers of categories of mutations in a sample. X-axis shows log10-transformed (with a pseudo count of +0.1 to allow plotting of zero values) values of the “APOBEC Mutation Load (minimum estimate)”.

Table S4 related to Figure 3. Significantly mutated genes identified by Genome MuSic analysis of exome sequence data and genes selected for targeted sequencing

Gene	Chromosome	Indels	SNVs	Total Mutations	Covd Bps	Muts pMbp	p value FCPT	p value LRT	p value CT	FDR FCPT	FDR LRT	FDR CT
Genes with significant mutation frequency from Genome MuSic analysis												
PIK3CA	3	0	16	16	87670	182.5	0	0	0	0	0	0
FGFR3	4	0	11	11	64279	171.13	1.64E-13	0	9.77E-19	1.27E-09	0	7.56E-15
KDM6A	X	5	5	10	225332	44.38	7.21E-10	8.43E-14	4.46E-14	4.18E-06	4.89E-10	2.59E-10
STAG2	X	3	4	7	129461	54.07	1.70E-07	2.77E-09	1.90E-11	0.00078693	1.07E-05	7.34E-08
HRAS	11	1	5	6	44463	134.94	2.40E-07	3.84E-09	1.89E-11	0.0009282	1.13E-05	7.34E-08
EP300	22	1	7	8	176308	45.38	2.92E-05	1.47E-06	3.19E-08	0.09682411	0.0028175	9.26E-05
RHOB	2	0	3	3	14241	210.66	0.00349597	8.80E-07	1.53E-06	1	0.00185793	0.0029518
ARID1A	1	3	1	4	269952	14.82	0.00740156	1.70E-06	3.19E-05	1	0.0028175	0.04627192
Genes selected for SureSelect capture												
ARHGAP18	6											
*ARID1A**	1											
*ARID4A**	14											
*ASH1L**	1											
ATP6V1B2	8											
ATP7B	13											
*BRCA2**	13											
CEP290	12											
CLTC	17											
CLU	8											
COL11A1	1											
*CREBBP**	16											
DLG4	17											
DOPEY1	6											
DYNC1H1	14											
*EP300**	22											
HEPACAM	11											
HERC1	15											
ITK	5											
*KDM3A**	2											
*KDM6A**	X											
KIF16B	20											
LARP1B	4											
MAGI3	1											
*MECOM**	3											
*KMT2A**	11											
*KMT2D**	12											
*KMT2C**	7											
NAT10	11											
*NCOR1**	17											
PALM3	19											
PGS1	17											
PHF3	6											
PIK3CA	3											
RAB11FIP1	8											
*RAD21**	8											
RBM10	X											
RBM6	3											
RHOB	2											
RREB1	6											
*STAG2**	X											
*STK38**	6											
TET3	2											
TSC1	9											
UEVLD	11											
UNC80	2											
USP47	11											
*UTY**	Y											
ZFYVE26	14											

Muts pMbp	The total number of mutations per Mb = (Total Mutations / Covd Bps) * 1e6
p value FCPT	p value for significance of the gene using the Fisher's Combined P value test
p value LRT	p value for significance of the gene using the Likelihood Ratio test
p value CT	p value for significance of the gene using the Convolution test
FDR FCPT	FDR for significance of the gene using the Fisher's Combined P-value test
FDR LRT	FDR for significance of the gene using the Likelihood Ratio test
FDR CT	FDR for significance of the gene using the Convolution test

* Genes with CM function from DAnCER database; http://wodaklab.org/dancer/

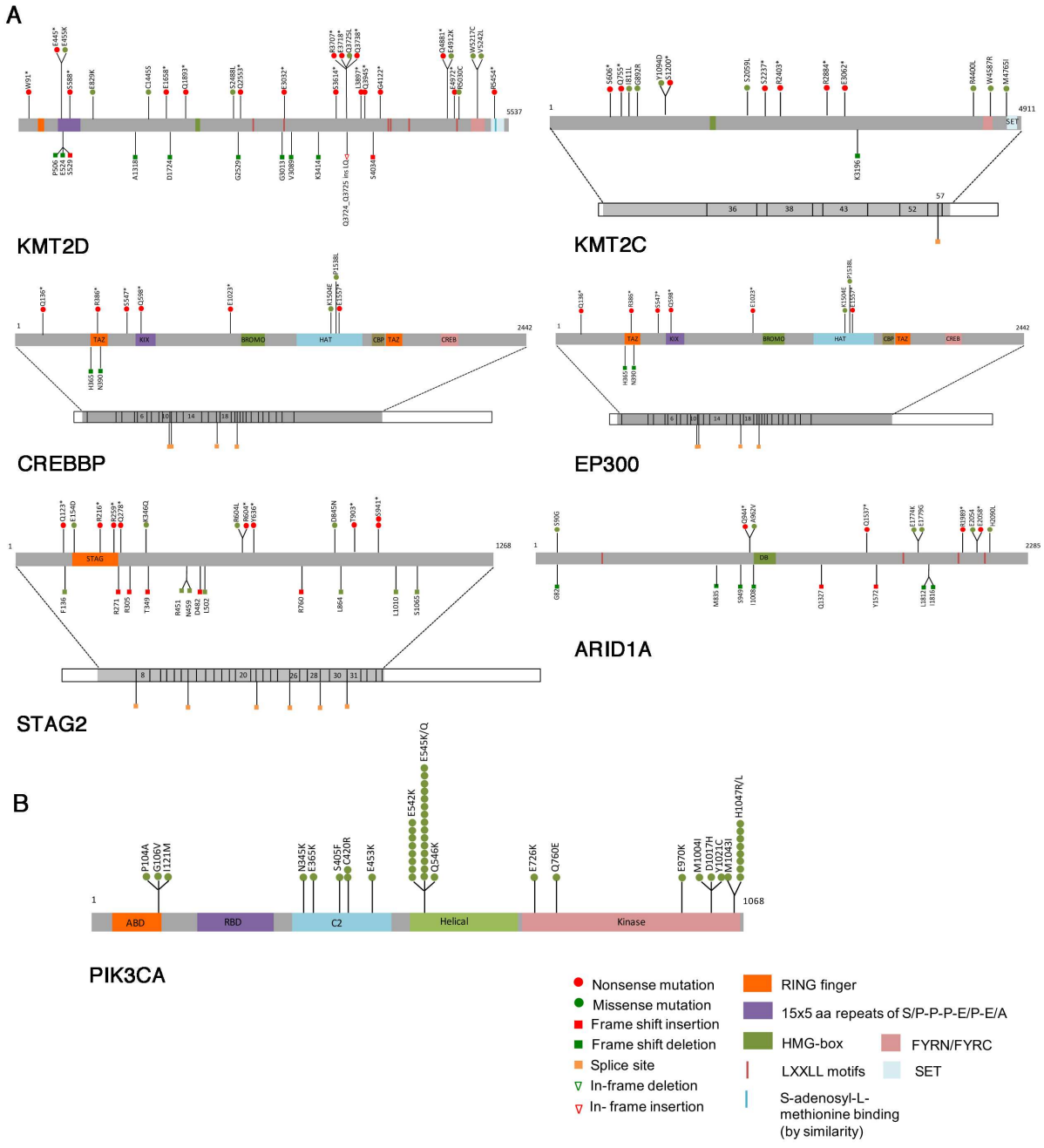
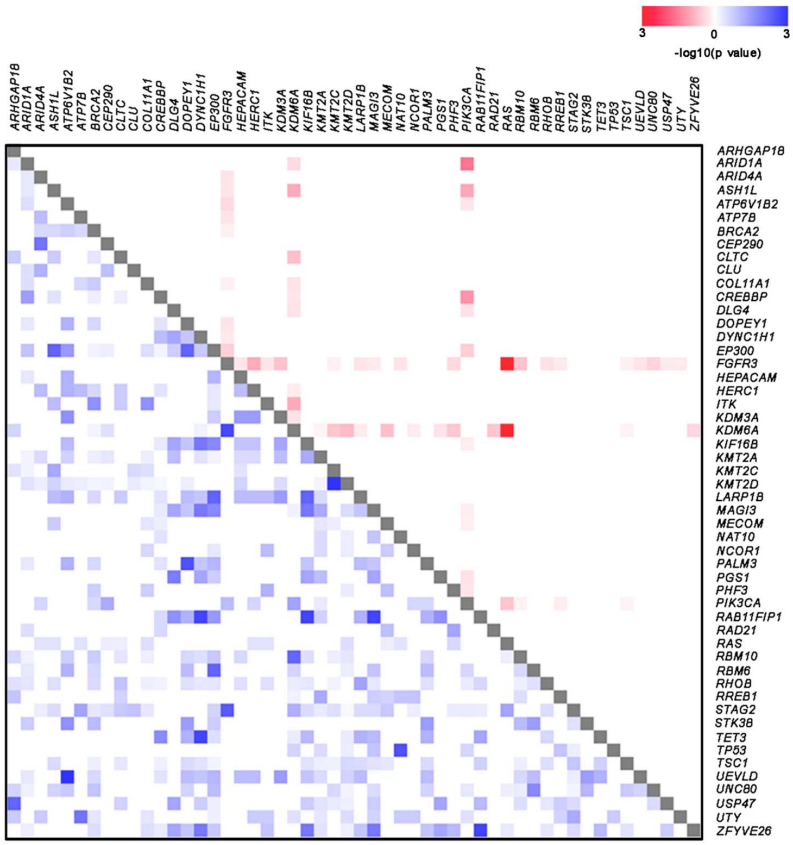


Figure S4, related to Figure 3. Schematics showing the distribution of mutations identified by exome sequencing and targeted re-sequencing in 82 stage Ta bladder tumors.

(A) Mutations in KMT2D, KMT2C, CREBBP, EP300, STAG2 and ARID1A proteins (top bars). Lower bars show cDNA sequence with positions of splice site mutations.
(B) Missense mutations in PIK3CA. ABD, adapter binding domain; RBD, Ras-binding domain; C2, C2 PI3K class I alpha.

A



B

Co-occurence correlations.

Gene 1	Gene 2	p value	FDR
KMT2D	KMT2C	6.61E-05	0.043823723
ATP6V1B2	UEVLD	0.00090334	0.266184884
RAB11FIP1	ZFYVE26	0.00180669	0.266184884
DYNC1H1	RAB11FIP1	0.00180669	0.266184884
DYNC1H1	TET3	0.00180669	0.266184884
KIF16B	RAB11FIP1	0.00180669	0.266184884
MAGI3	RAB11FIP1	0.00180669	0.266184884
KDM6A	FGFR3	0.00205861	0.272971824
DOPEY1	PALM3	0.00268744	0.296962511
NAT10	TP53	0.00268744	0.296962511
STAG2	FGFR3	0.00366252	0.371958351
EP300	LARP1B	0.00513776	0.371958351
EP300	RBM6	0.00513776	0.371958351
ASH1L	EP300	0.00513776	0.371958351
DOPEY1	EP300	0.00513776	0.371958351
KIF16B	LARP1B	0.00532972	0.371958351
ARHGAP18	USP47	0.00532972	0.371958351
KDM6A	RBM10	0.00579472	0.384189789
ARID4A	CEP290	0.00880759	0.496325904
ATP7B	UTY	0.00880759	0.496325904
DYNC1H1	KIF16B	0.01048049	0.496325904
DYNC1H1	MAGI3	0.01048049	0.496325904
DYNC1H1	ZFYVE26	0.01048049	0.496325904
KIF16B	MAGI3	0.01048049	0.496325904
KIF16B	ZFYVE26	0.01048049	0.496325904
MAGI3	ZFYVE26	0.01048049	0.496325904
DLG4	PGS1	0.01219512	0.557611438
RBM6	UNC80	0.01817977	0.754567811
EP300	KIF16B	0.01820978	0.754567811
EP300	MAGI3	0.01820978	0.754567811
CREBBP	TET3	0.01987353	0.798554652
BRCA2	ITK	0.02402891	0.817726605
COL11A1	ITK	0.02402891	0.817726605
ATP6V1B2	KDM3A	0.02439024	0.817726605
ATP6V1B2	STK38	0.02439024	0.817726605
DLG4	RAB11FIP1	0.02439024	0.817726605
PGS1	RAB11FIP1	0.02439024	0.817726605
ATP6V1B2	EP300	0.03161698	0.951219512
EP300	RAB11FIP1	0.03161698	0.951219512
KDM3A	LARP1B	0.03658537	0.951219512
KDM3A	UEVLD	0.03658537	0.951219512
STK38	UEVLD	0.03658537	0.951219512
DOPEY1	STK38	0.03658537	0.951219512
HEPACAM	KDM3A	0.03658537	0.951219512
HERC1	KDM3A	0.03658537	0.951219512
PALM3	STK38	0.03658537	0.951219512
RBM6	STK38	0.03658537	0.951219512
ARID1A	CREBBP	0.03836349	0.951219512
KDM6A	PIK3CA	0.04543883	0.951219512
ATP7B	STAG2	0.04584463	0.951219512
LARP1B	RHOB	0.04595754	0.951219512
STAG2	RAS	0.04782199	0.951219512
PHF3	RAD21	0.04847937	0.951219512
RAB11FIP1	TET3	0.04847937	0.951219512
DLG4	DYNC1H1	0.04878049	0.951219512
DLG4	KIF16B	0.04878049	0.951219512
DLG4	MAGI3	0.04878049	0.951219512
DLG4	USP47	0.04878049	0.951219512
DLG4	ZFYVE26	0.04878049	0.951219512
KDM3A	KIF16B	0.04878049	0.951219512
PGS1	USP47	0.04878049	0.951219512
PGS1	ZFYVE26	0.04878049	0.951219512
DYNC1H1	PGS1	0.04878049	0.951219512
KIF16B	PGS1	0.04878049	0.951219512
MAGI3	PGS1	0.04878049	0.951219512

Mutual exclusivity correlations.

Gene 1	Gene 2	p value	FDR
FGFR3	RAS	7.47E-11	9.90E-08
KDM6A	RAS	0.00110225	0.266184884
ARID1A	PIK3CA	0.02466747	0.817726605

Figure S5, related to Figure 3. Mutual exclusivity and co-occurrence of mutations in stage Ta tumors.

(A) The upper right triangle (red) represents mutually exclusive relationships and lower left triangle (blue) represents co-occurrent relationships between 52 genes with somatic mutations in 82 stage Ta tumors. p value and FDRs were calculated using the Fisher's exact test and heatmap values reflect -log10 (p value). p values (-log10) greater than 3 were assigned the maximum score.
(B) Co-occurrence and mutual exclusivity correlations. Listed are results of Fisher's exact tests for significantly mutated genes to a maximum p value of 0.05. False discovery rate (FDR) is also reported.

Table S5 related to Figure 3. *BRCA2* somatic and germline variants

Ensembl gene ID	Genomic position (hg19)	Variant type	Variant effect	Ensembl transcript ID	cDNA change	Protein change	SIFT(score)	PolyPhen(score)	Variant class*	Sample/ patient ID	Somatic/ germline	MAF§
ENSG00000139618	13:32912678 C>G	SNV	Missense	ENST00000544455	c.4186 C>G	p.Q1396E	tolerated(1)	benign(0)	Class 2	718	Somatic	47
ENSG00000139618	13:32912846 C>T	SNV	Nonsense	ENST00000544455	c.4354 C>T	p.Q1452*	-	-	Class 5	718	Somatic	47
ENSG00000139618	13:32906732 C>T	SNV	Nonsense	ENST00000544455	c.1117 C>T	p.Q373*	-	-	Class 5	1116	Somatic	21
ENSG00000139618	13:32899251 A>G	SNV	Missense	ENST00000544455	c.355 A>G	p.T119A	tolerated(0.11)	possibly damaging(0.469)	Class 2	1117	Somatic	50
ENSG00000139618	13:32893393 G>C	SNV	Missense	ENST00000544455	c.247 G>C	p.E83Q	deleterious(0.01)	benign(0.246)	Class 3	1307	Somatic	46
ENSG00000139618	13:32911598 G>C	SNV	Missense	ENST00000544455	c.3106 G>C	p.E1036Q	tolerated(0.1)	probably damaging(0.943)	Class 3	1474	Somatic	48
ENSG00000139618	13:32912318 G>C	SNV	Missense	ENST00000544455	c.3826 G>C	p.E1276Q	tolerated(0.09)	probably damaging(0.953)	Class 2	1527	Somatic	46
ENSG00000139618	13:32969035 C>G	SNV	Missense	ENST00000544455	c.9466 C>G	p.Q3156E	deleterious(0)	benign(0.303)	Class 3	1527	Somatic	45
ENSG00000139618	13:32910799_32910804 del6	Deletion	In-frame deletion	ENST00000544455	c.2307_2312 del6	p.I770_L771del	-	-	Class 3	1991	Somatic	62
ENSG00000139618	13:32936728 G>C	SNV	Missense	ENST00000544455	c.7874 G>C	p.R2625T	deleterious(0)	probably damaging(0.993)	Class 3	2013	Somatic	34
ENSG00000139618	13:32911181 G/C	SNV	Missense	ENST00000544455	c.2689 G/C	p.E897Q	deleterious(0.04)	possibly_damaging(0.508)	Class 3	712	Germline	N/A
ENSG00000139618	13:32915315 G/A	SNV	Missense	ENST00000544455	c.6823 G/A	p.E2275K	tolerated(0.46)	benign(0.217)	Class 3	1231	Germline	N/A
ENSG00000139618	13:32929284 A/G	SNV	Missense	ENST00000544455	c.7294 A/G	p.R2432G	deleterious(0)	benign(0.001)	Class 2	575	Germline	N/A
ENSG00000139618	13:32937324 C/A	SNV	Missense	ENST00000544455	c.7985 C/A	p.T2662K	deleterious(0)	benign(0.003)	Class 3	712	Germline	N/A
ENSG00000139618	13:32972626 A/T	SNV	Nonsense	ENST00000544455	c.9976 A/T	K3326*	-	-		1474	Germline	N/A

* Classification according to Association for Clinical Genetic Science (ACGS) Guidelines

2 Unlikely to be pathogenic

3 Unknown significance

5 Clearly pathogenic

§ Mutant allele frequency

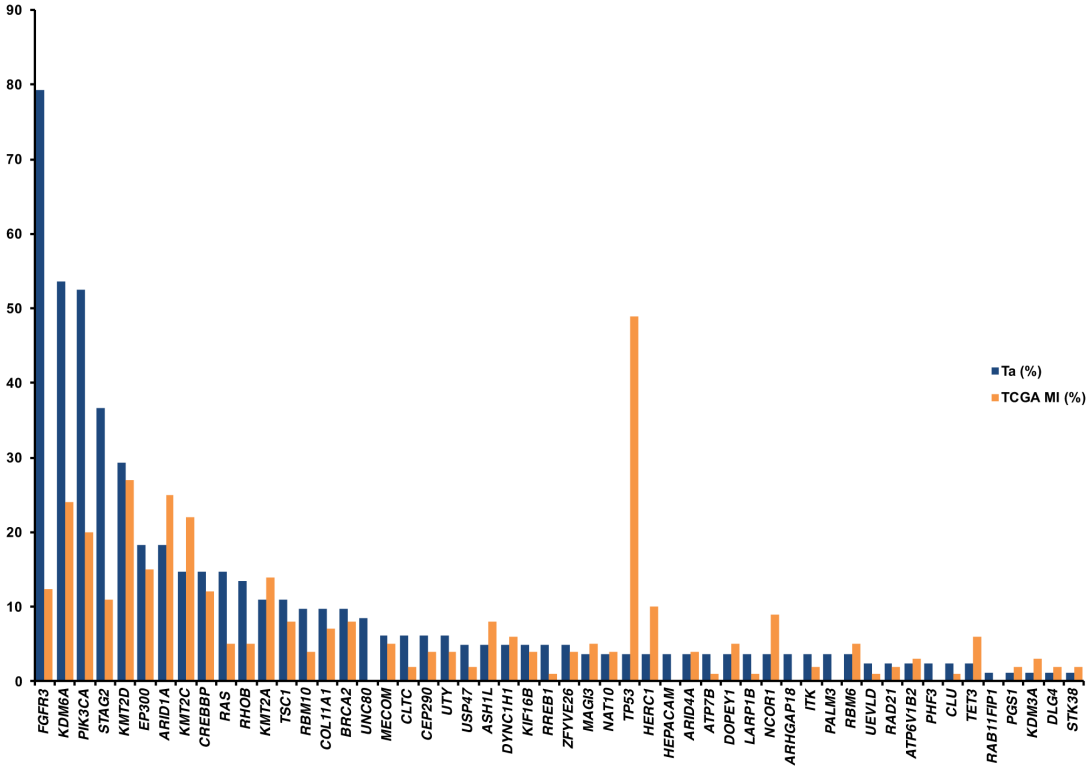


Figure S6, related to Figure 3. Comparison of mutations identified in NIBC and MIBC.

Frequencies of mutations identified by exome and targeted sequencing of 82 stage Ta tumors in this study and data from 130 muscle-invasive bladder tumors analysed by The Cancer Genome Atlas Research Network [Data accessed from cBioPortal (http://cbioportal.org/public-portal/index.do)] are shown.

Table S6 related to Figure 3. Chromatin modification genes with mutation in 24 bladder tumor exomes

Gene*	Total mutations	Inactivating mutations
ACTB	1	0
ACTR6	1	0
ADAR	1	0
APPL1	1	0
ARID1A	4	3
ARID2	2	2
ARID4A	1	1
ASH1L	2	2
ATRX	1	0
AURKB	1	1
BANF1	1	0
BAZ1A	1	0
BAZ2A	1	1
BRCA1	1	0
BRCA2	3	0
CARM1	1	0
CENPK	1	0
CHEK1	1	0
CHEK2	1	0
CREBBP	5	3
DHX30	1	0
DHX9	1	0
EMD	1	0
EP300	9	7
EPC2	1	0
GSG2	2	0
HDAC6	1	0
HDAC9	1	0
HELLS	1	0
HESX1	1	0
HEY2	1	1
HIRA	1	0
HNRNPR	1	0
HSP90AA1	2	0
HSPA8	1	1
HUWE1	2	0
INO80	1	0
JMJD1C	1	0
KDM3A	1	1
KDM6A	11	10
KIF11	1	0
KIF4A	1	0
LEO1	1	0
MECOM	2	2
MGA	1	0
KMT2A	4	2
KMT2D	7	4
KMT2C	3	2
MTA1	1	0
MYSM1	2	0

NASP	5	0
NCOA3	1	0
NCOR1	1	1
NCOR2	1	0
NOC2L	1	0
NPM1	1	0
PABPC1	3	0
PHC3	1	0
PHF17	1	0
POLR2A	1	0
PRDM2	1	0
RAD21	1	1
RREB1	2	1
RSF1	1	0
SETD1A	1	0
SIRT4	1	0
SMC3	1	1
SND1	2	0
SP1	1	0
SRCAP	1	0
STAG2	9	7
STK38	1	1
SUPT6H	2	0
SUV420H1	1	1
SYCP3	1	0
TACC2	1	0
TAF1	1	0
TAF5L	1	0
TBL1XR1	3	1
TLE2	1	0
TOP2B	2	0
UTY	4	2
WHSC1L1	2	0
ZEB1	2	0

* Genes with CM function from DAnCER database; http://wodaklab.org/dancer/