Cell Reports Medicine, Volume *2*

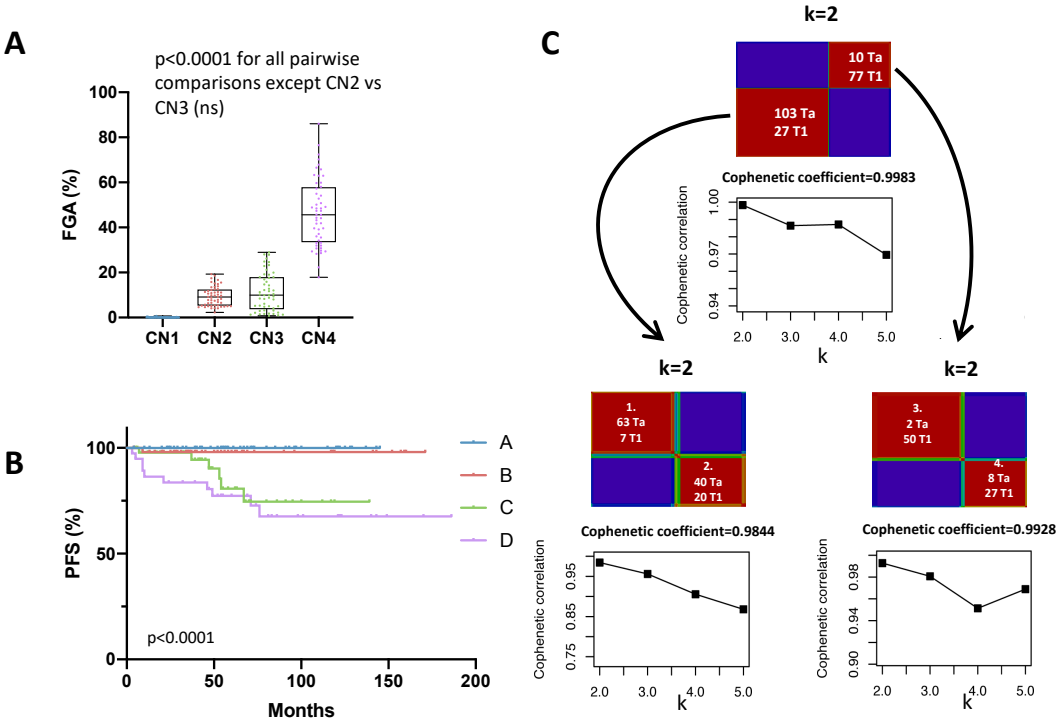
Supplemental information

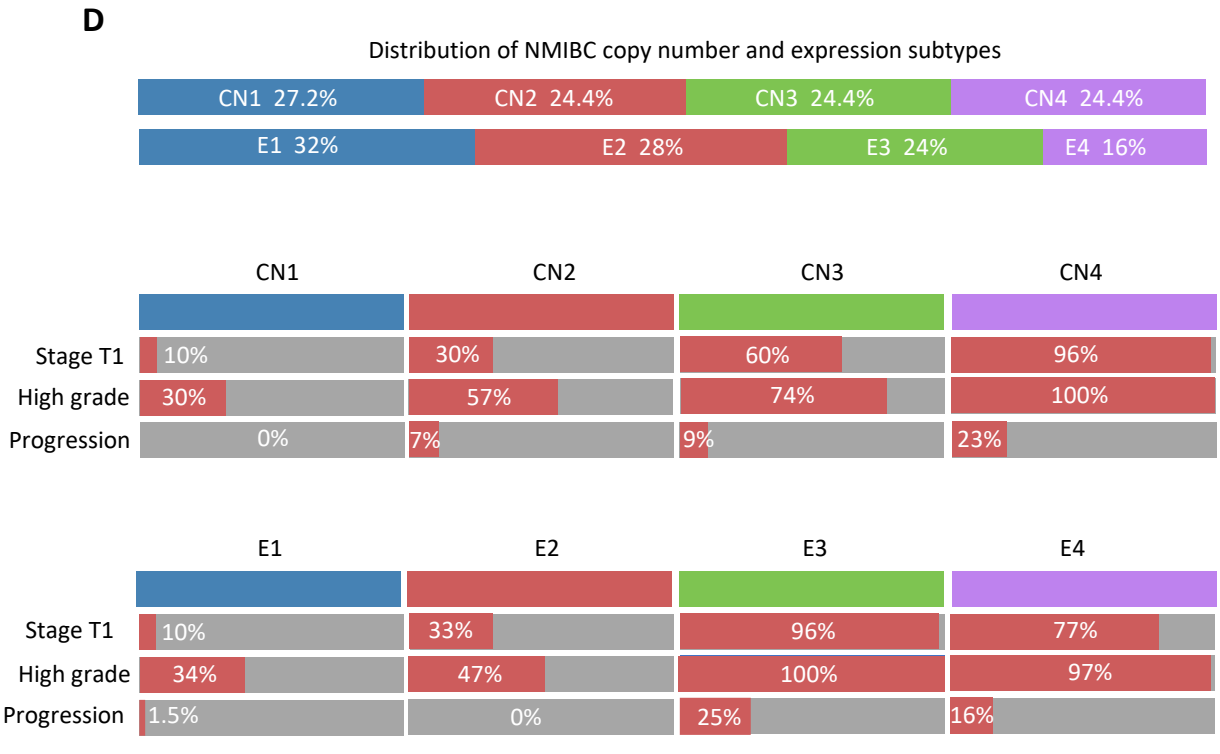
Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight

Carolyn D. Hurst, Guo Cheng, Fiona M. Platt, Mauro A.A. Castro, Nour-al-dain S. Marzouka, Pontus Eriksson, Emma V.I. Black, Olivia Alder, Andrew R.J. Lawson, Sia V. Lindskrog, Julie E. Burns, Sunjay Jain, Jo-An Roulson, Joanne C. Brown, Jan Koster, A. Gordon Robertson, Inigo Martincorena, Lars Dyrskjøt, Mattias Höglund, and Margaret A. Knowles

Table S2. Genes selected for targeted sequencing. Related to Figures 1-3.

ACAN	EGFR	LARP1B	RAD21
AHNAK2	ELF3	LGALS8	RARG
AKT1	EP300	LPHN3	RB1
ARHGAP18	EPG5	LRRC7	RBM10
ARHGEF10	ERBB2	MAGI3	RBM6
ARHGEF3	ERBB3	MAML1	RHOA
ARID1A	ERCC2	MAPK8IP3	RHOB
ARID2	ESPL1	MECOM	RREB1
ARID4A	FANCA	MYCBP2	RXRA
ASH1L	FAT1	MYO5B	RYR2
ASXL2	FAT2	NAT10	SCN1A
ATM	FAT3	NCOR1	SLC25A48
ATP6V1B2	FAT4	NCSTN	SPTAN1
ATP7B	FBXW7	NF1	STAG1
B3GNT9	FGFR3	NFE2L2	STAG2
BRAF	FMN1	NFE2L3	STK38
BRCA2	FOXA1	NOTCH1	SYNE1
BTG2	FOXQ1	NOTCH2	SYNE2
C1ORF173	FREM2	NOTCH3	TET3
CACNA1D	HAUS6	NRAS	TEX15
CCND1	HEPACAM	OSMR	TNC
CCND3	HERC1	PAIP1	TP53
CDKN1A	HMCN1	PALM3	TRAK1
CDKN2A	HRAS	PCDHA9	TSC1
CDKN2B	HRNR	PDZD2	TSC2
CEP290	INADL	PGS1	TXNIP
CHD6	ITK	PHF3	UEVLD
CLTC	KDM3A	PIK3CA	USP47
CLU	KDM6A	PIK3R1	UTY
COL11A1	KIF16B	PIK3R4	VCAN
CPAMD8	KLF5	POLE	WHSC1L1
CREBBP	KMT2A	POLE2	WNK1
DLG4	KMT2C	POTEF	ZFHX3
DOPEY1	KMT2D	PTEN	ZFP36L1
DUX4L4	KRAS	RAB11FIP1	ZFYVE26


A
p<0.0001 for all pairwise comparisons except CN2 vs CN3 (ns)
FGA (%)
CN1
CN2
CN3
CN4
B
PFS (%)
p<0.0001
Months
A
B
C
D
C
k=2
10 Ta 77 T1
103 Ta 27 T1
Cophenetic coefficient=0.9983
Cophenetic correlation
k
1. 63 Ta 7 T1
2. 40 Ta 20 T1
Cophenetic coefficient=0.9844
3. 2 Ta 50 T1
4. 8 Ta 27 T1
Cophenetic coefficient=0.9928


D
Distribution of NMIBC copy number and expression subtypes
CN1 27.2%
CN2 24.4%
CN3 24.4%
CN4 24.4%
E1 32%
E2 28%
E3 24%
E4 16%
CN1
CN2
CN3
CN4
Stage T1
10%
30%
60%
96%
High grade
30%
57%
74%
100%
Progression
0%
7%
9%
23%
E1
E2
E3
E4
Stage T1
10%
33%
96%
77%
High grade
34%
47%
100%
97%
Progression
1.5%
0%
25%
16%

Supplementary Figure 1. Features of copy number and expression subtypes of all 217 NMIBC. Related to Figure 1.

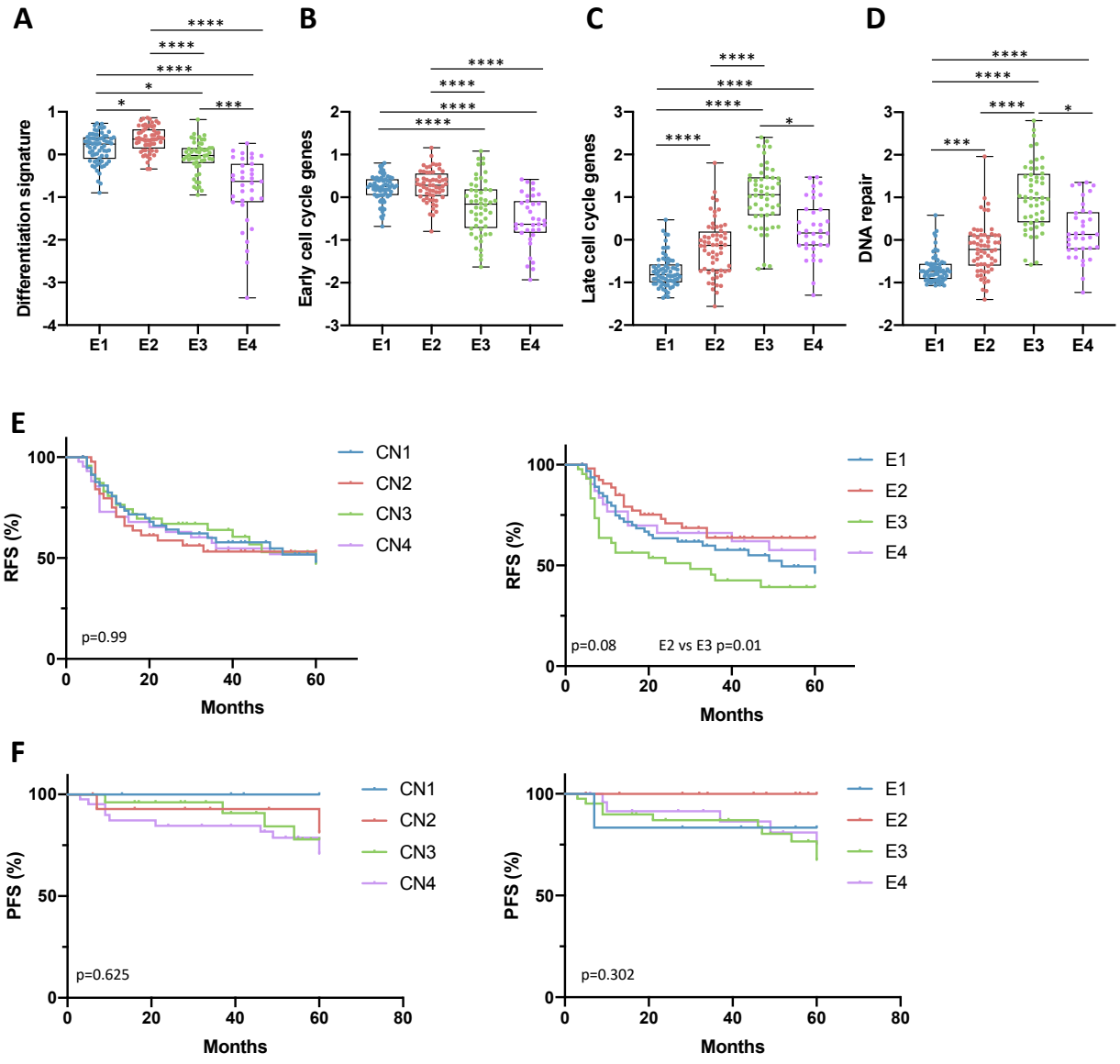
(A) Fraction of genome altered (FGA) in copy number subtypes CN1-CN4. Analysis of copy number data from NGS and array-based platforms was restricted to those genomic regions associated with BAC clones present on the 1Mb resolution CGH array. Each region was assigned a copy number class (0 = no copy number alteration; 1 = gain; 2 = high-level amplification; -1 = loss; -2 = high-level loss). FGA is defined as the percentage of clones reporting significantly altered copy number (gain or loss).

(B) Kaplan-Meier plots of progression-free survival (PFS) of NMIBC patients according to Fraction of Genome Altered (FGA) with tumors assigned to arbitrary groups with <1% FGA (A), 1-<10% FGA (B), 10-<30% FGA (C) and ≥30% FGA (D).

(C) Derivation of four NMIBC expression subtypes by two stage NMF analysis, showing subdivision of tumors of stage Ta and T1 into clusters and plots of cophenetic coefficients among k clusters for each analysis.

(D) Top, distribution of copy number and expression subtypes in 217 NMIBC. Lower panels show tumor stage, grade and progression frequencies according to subtypes.

(A) Kruskal-Wallis test with Dunn's multiple comparison correction. Mean, 25th and 75^{th} percentiles, minimum and maximum values are shown. **(B)** Log-rank analysis.


A
B
C
D
Differentiation signature
Early cell cycle genes
Late cell cycle genes
DNA repair
E1
E2
E3
E4
E
RFS (%)
CN1
CN2
CN3
CN4
p=0.99
Months
p=0.08
E2 vs E3 p=0.01
F
PFS (%)
p=0.625
p=0.302

Supplementary Figure 2. Expression signatures, recurrence-free survival and progression-free survival of NMIBC subtypes. Related to Figure 1.

(A) Urothelial differentiation signature in NMIBC expression subtypes.

(B) Expression of early cell cycle genes in NMIBC expression subtypes.

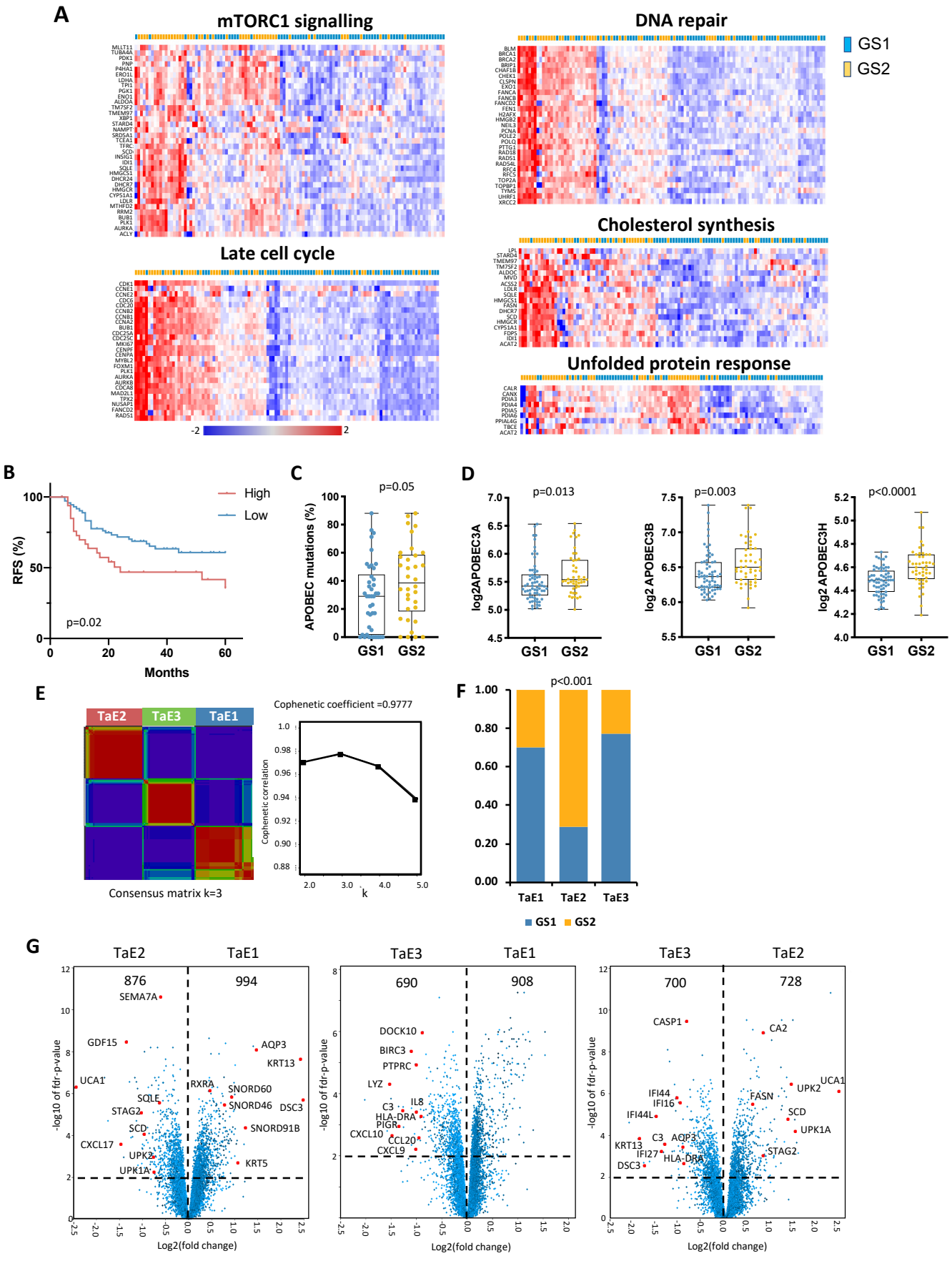
(C) Expression of late cell cycle genes in NMIBC expression subtypes.

(D) Expression of DNA repair genes in NMIBC expression subtypes.

(E) Kaplan-Meier plots of recurrence-free survival (RFS) according to NMIBC copy number subtypes CN1-CN4 (left) and expression subtypes E1-E4 (right).

(F) Kaplan-Meier plots of PFS of patients with T1 tumors only according to NMIBC copy number (left) and expression subtypes (right).

(A-D) Group measurements compared using Kruskal-Wallis test with Dunn's multiple comparison correction. Mean, 25th and 75^{th} percentiles, minimum and maximum values are shown. Adjusted p values; ****$p<0.0001$, ***$p<0.001$, **$p<0.01$, *$p<0.05$. **(E** and **F)** Log-rank analysis.


A
mTORC1 signalling
DNA repair
Cholesterol synthesis
Late cell cycle
Unfolded protein response
GS1
GS2
-2
2
B
High
Low
RFS (%)
p=0.02
Months
C
p=0.05
APOBEC mutations (%)
GS1
GS2
D
p=0.013
log2APOBEC3A
p=0.003
log2 APOBEC3B
p<0.0001
log2 APOBEC3H
E
TaE2
TaE3
TaE1
Consensus matrix k=3
Cophenetic coefficient =0.9777
Cophenetic correlation
k
F
p<0.001
TaE1
TaE2
TaE3
GS1
GS2
G
TaE2
TaE1
876
994
SEMA7A
GDF15
UCA1
SQLE
STAG2
SCD
CXCL17
UPK2
UPK1A
AQP3
KRT13
RXRA
SNORD60
SNORD46
DSC3
SNORD91B
KRT5
-log10 of fdr-p-value
Log2(fold change)
TaE3
TaE1
690
908
DOCK10
BIRC3
PTPRC
LYZ
C3
IL8
HLA-DRA
PIGR
CXCL10
CCL20
CXCL9
TaE3
TaE2
700
728
CASP1
CA2
UCA1
UPK2
IFI44
IFI16
FASN
IFI44L
SCD
UPK1A
KRT13
C3
AQP3
IFI27
HLA-DRA
STAG2
DSC3

Supplementary Figure 3. Features of stage Ta tumors according to Ta-derived genomic and expression subtypes. Related to Figure 2.

(A) Heatmaps of z-scores derived by clustering of expression levels of differentially expressed gene categories in genomic subtypes GS1 and GS2. Top bar in each panel indicates GS1 and GS2 assignment.

(B) Recurrence-free survival (RFS) according to high and low tumor grade. Log-rank analysis.

(C) Percentage of APOBEC-related mutations according to Ta tumor genomic subtype.

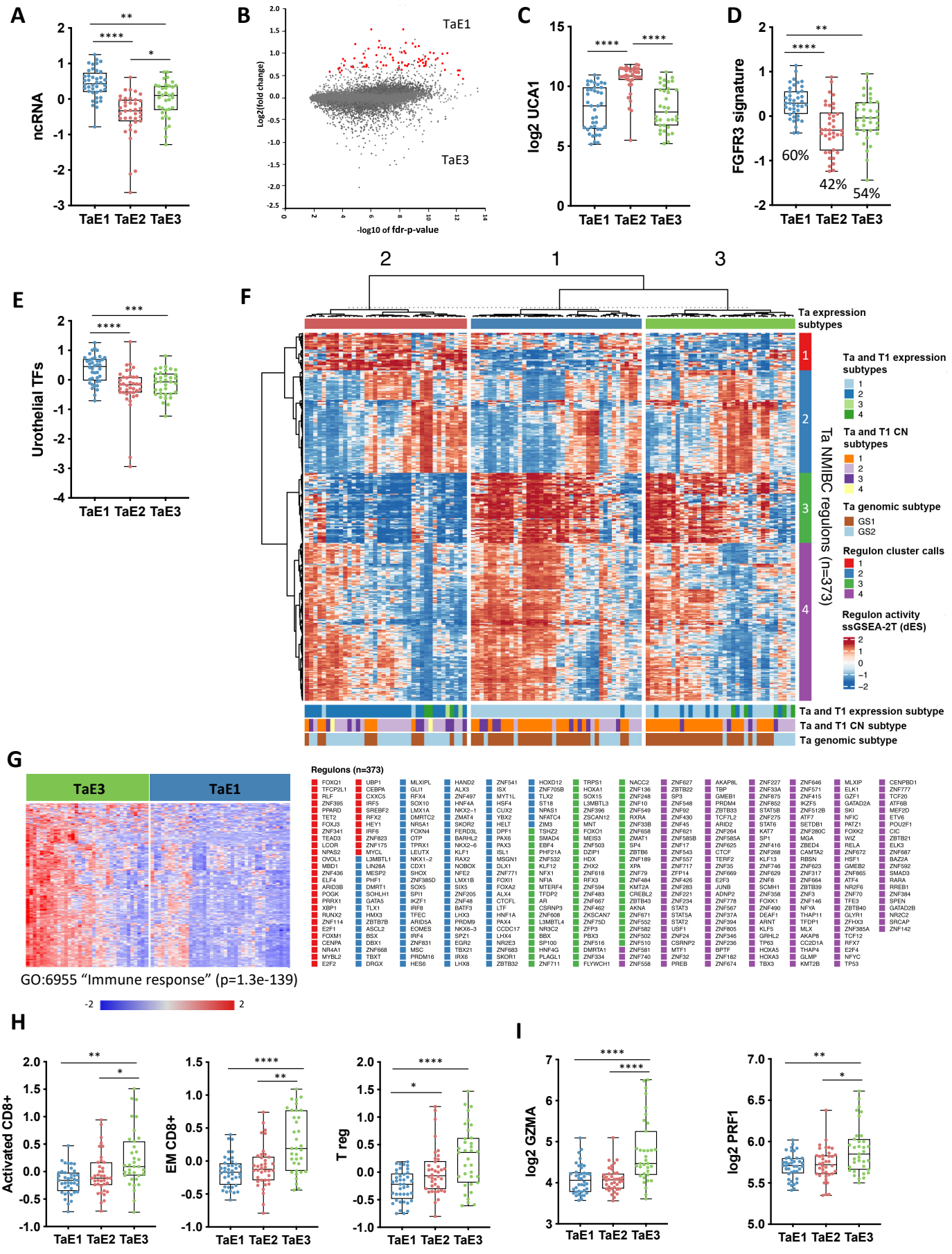
(D) mRNA expression levels of APOBEC3A, 3B and 3H in Ta genomic subtype tumors.

(E) Left, plot of NMF consensus clusters for three Ta expression subtypes TaE1-TaE3. Right, plot showing cophenetic coefficients among k clusters.

(F) Relationships of Ta genomic and expression subtypes.

(G) Volcano plots of genes differentially expressed (LIMMA test) between Ta tumor expression subtypes TaE1-TaE3 with false discovery rate $p<0.01$. Numbers of significantly up-regulated genes in each comparison are indicated at the top of each plot. Selected genes are indicated.

(B) Log-rank analysis. **(C** and **D)** Mann-Whitney test. Mean, 25th and 75^{th} percentiles, minimum and maximum values are shown. **(F)** Chi-square test with Bonferroni correction.


A
ncRNA
TaE1 TaE2 TaE3
B
TaE1
TaE3
Log2(fold change)
-log10 of fdr-p-value
C
log2 UCA1
TaE1 TaE2 TaE3
D
FGFR3 signature
60%
42%
54%
TaE1 TaE2 TaE3
E
Urothelial TFs
TaE1 TaE2 TaE3
F
2
1
3
Ta expression subtypes
Ta NMIBC regulons (n=373)
Ta and T1 expression subtypes
Ta and T1 CN subtypes
Ta genomic subtype
GS1
GS2
Regulon cluster calls
Regulon activity ssGSEA-2T (dES)
Ta and T1 expression subtype
Ta and T1 CN subtype
Ta genomic subtype
G
TaE3
TaE1
Regulons (n=373)
GO:6955 “Immune response” (p=1.3e-139)
H
Activated CD8+
EM CD8+
T reg
TaE1 TaE2 TaE3
I
log2 GZMA
log2 PRF1
TaE1 TaE2 TaE3

Supplementary Figure 4. Expression differences between stage Ta tumor genomic and expression subtypes. Related to Figure 2.

(A) Non-coding RNA signature (GO:16070) according to Ta expression subtypes.

(B) MA plot of genes differentially expressed between TaE1 and T1E3. Small nucleolar RNAs with higher expression in T1E1 are marked in red.

(C) Expression of UCA1 in Ta expression subtypes.

(D) FGFR3-related signature according to Ta expression subtypes. *FGFR3* mutation frequencies are indicated.

(E) Scores for expression of transcription factors (TFs) implicated in urothelial differentiation.

(F) Heatmap of regulon activity for 373 regulons with differential activity in stage Ta tumors supervised by Ta expression subtype. Expression and genomic covariates are shown below.

(G) Heatmap of z-scores for genes in GO:6955 "Immune response" in Ta expression subtypes TaE1 and TaE3.

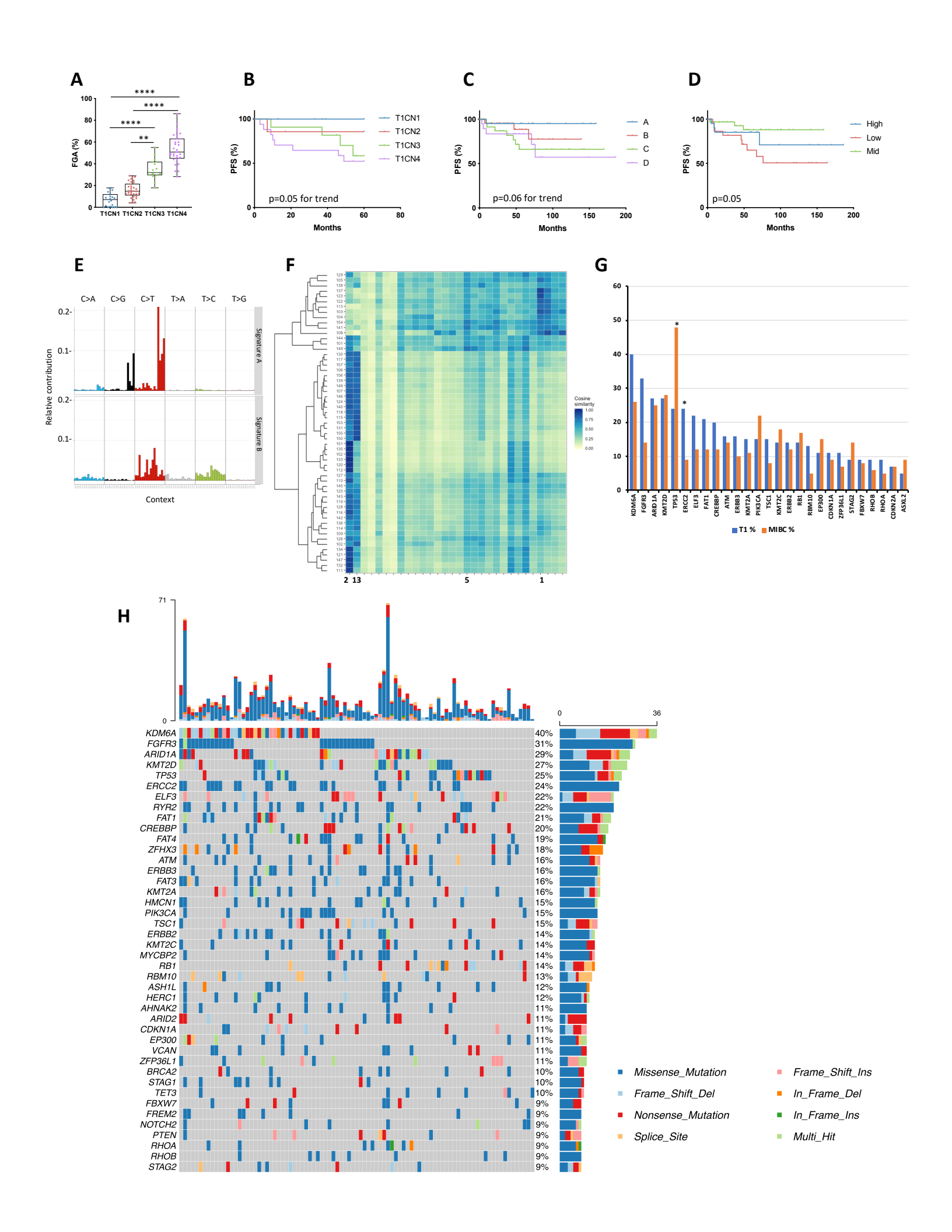
(H) Scores for selected immune signatures [S1] in Ta expression subtypes. Activated DC8+ (left); EM, effector memory (middle); T reg, T regulatory (right).

(I) Expression levels of granzyme (GZMA) (left) and perforin 1 (PRF1) (right) in Ta expression subtypes.

(A, C, D, E, H and **I)** Group measurements compared using Kruskal-Wallis test with Dunn's multiple comparison correction. Mean, 25th and 75th percentiles, minimum and maximum values are shown. Adjusted p values; ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05.

Table S4. snoRNAs differentially expressed in comparisons of Ta expression subgroups. Related to Figure 2.

TaE1 vs 2		TaE1 vs 3	
SNORA11	SNORD119	SNORA11	SNORD15B
SNORA12	SNORD11B	SNORA12	SNORD18A
SNORA14A	SNORD12	SNORA14A	SNORD32B
SNORA20	SNORD126	SNORA20	SNORD35B
SNORA21	SNORD127	SNORA21	SNORD37
SNORA22	SNORD12B	SNORA22	SNORD41
SNORA23	SNORD13	SNORA23	SNORD42A
SNORA24	SNORD14E	SNORA24	SNORD46
SNORA26	SNORD15B	SNORA26	SNORD58A
SNORA30	SNORD18A	SNORA30	SNORD58C
SNORA37	SNORD32B	SNORA37	SNORD59B
SNORA47	SNORD35B	SNORA47	SNORD60
SNORA48	SNORD37	SNORA48	SNORD61
SNORA49	SNORD41	SNORA49	SNORD70
SNORA50	SNORD42A	SNORA50	SNORD71
SNORA52	SNORD46	SNORA52	SNORD78
SNORA54	SNORD58A	SNORA54	SNORD8
SNORA57	SNORD58C	SNORA57	SNORD82
SNORA5A	SNORD59B	SNORA5A	SNORD85
SNORA65	SNORD60	SNORA65	SNORD9
SNORA68	SNORD61	SNORA68	SNORD91B
SNORA71B	SNORD70	SNORA71B	SNORD92
SNORD10	SNORD71	SNORD10	SNORD94
SNORD104	SNORD78	SNORD104	SNORD97
SNORD105	SNORD8	SNORD105	SNORD99
SNORD105B	SNORD82	SNORD105B	SNORA28
SNORD11	SNORD85	SNORD11	SNORA71D
SNORD111	SNORD9	SNORD111	SNORA76
SNORD111B	SNORD91B	SNORD111B	SNORD12C
SNORD116-1	SNORD92	SNORD116-1	SNORD59A
SNORD116-11	SNORD94	SNORD116-11	SNORD63
SNORD116-12	SNORD97	SNORD116-12	SNORA10
SNORD116-14	SNORD99	SNORD116-14	SNORA14B
SNORD116-15	SNORA45	SNORD116-15	SNORA16B
SNORD116-16	SNORD113-5	SNORD116-16	SNORA38B
SNORD116-18	SNORD115-14	SNORD116-18	SNORA5C
SNORD116-2	SNORD53	SNORD116-2	SNORA70C
SNORD116-23	SNORD93	SNORD116-23	SNORA70D
SNORD116-24	SNORA28	SNORD116-24	SNORA70E
SNORD116-25	SNORA71D	SNORD116-25	SNORA71A
SNORD116-26	SNORA76	SNORD116-26	SNORA71C
SNORD116-27	SNORD12C	SNORD116-27	SNORA80B
SNORD116-29	SNORD59A	SNORD116-29	SNORD121A
SNORD116-8	SNORD63	SNORD116-8	SNORD123
SNORD117	SNORD90	SNORD117	SNORD14A
		SNORD119	SNORD15A
		SNORD11B	SNORD16
		SNORD12	SNORD3A
		SNORD126	SNORD3C
		SNORD127	SNORD69
		SNORD12B	SNORD83A
		SNORD13	SNORD91A
		SNORD14E	SNORD116-13


A
FGA (%)
T1CN1 T1CN2 T1CN3 T1CN4
****
****
****
**
B
PFS (%)
T1CN1
T1CN2
T1CN3
T1CN4
p=0.05 for trend
Months
C
PFS (%)
A
B
C
D
p=0.06 for trend
Months
D
PFS (%)
High
Low
Mid
p=0.05
Months
E
C>A C>G C>T T>A T>C T>G
Relative contribution
Signature A
Signature B
Context
F
Cosine similarity
2 13
5
1
G
T1 %
MIBC %
KDM6A FGFR3 ARID1A KMT2D TP53 ERCC2 ELF3 FAT1 CREBBP ATM ERBB3 KMT2A PIK3CA TSC1 KMT2C ERBB2 RB1 RBM10 EP300 CDKN1A ZFP36L1 STAG2 FBXW7 RHOB RHOA CDKN2A ASXL2
H
KDM6A 40%
FGFR3 31%
ARID1A 29%
KMT2D 27%
TP53 25%
ERCC2 24%
ELF3 22%
RYR2 22%
FAT1 21%
CREBBP 20%
FAT4 19%
ZFHX3 18%
ATM 16%
ERBB3 16%
FAT3 16%
KMT2A 16%
HMCN1 15%
PIK3CA 15%
TSC1 15%
ERBB2 14%
KMT2C 14%
MYCBP2 14%
RB1 14%
RBM10 13%
ASH1L 12%
HERC1 12%
AHNAK2 11%
ARID2 11%
CDKN1A 11%
EP300 11%
VCAN 11%
ZFP36L1 11%
BRCA2 10%
STAG1 10%
TET3 10%
FBXW7 9%
FREM2 9%
NOTCH2 9%
PTEN 9%
RHOA 9%
RHOB 9%
STAG2 9%
Missense_Mutation
Frame_Shift_Ins
Frame_Shift_Del
In_Frame_Del
Nonsense_Mutation
In_Frame_Ins
Splice_Site
Multi_Hit

Supplementary Figure 5. Genomic features of stage T1 tumors. Related to Figure 3.

(A) Fraction of Genome Altered (FGA%) according to T1 copy number subtypes T1CN1-T1CN4.

(B) Kaplan-Meier plots of PFS according to T1 copy number subtypes T1CN1-T1CN4.

(C) Kaplan-Meier plots of PFS according to Fraction of Genome Altered (FGA) of patients with T1 tumors assigned to arbitrary groups with <1% FGA (A), 1-<10% FGA (B), 10-<30% FGA (C) and ≥30% FGA (D).

(D) Kaplan-Meier plots of progression-free survival (PFS) according to tumor mutational burden. High, top 25%; Mid, middle 50%; Low, lowest 25%.

(E) Trinucleotide mutational profiles from two signatures extracted from whole exome data using NMF analysis in MutationalPatterns.

(F) Heatmap generated in MutationalPatterns showing identification of COSMIC mutational signatures in T1 tumors. Signatures 2 and 13 which are attributed to the activity of APOBEC cytidine deaminases, SBS1 and SBS5 are indicated.

(G) Comparison of the frequencies of mutations identified $\geq$ 5% of T1 tumors with those reported in muscle-invasive bladder tumors [S2].

(H) Oncoplot showing genes mutated in $\geq$ 9% of T1 tumors. Data for genes included in the 140 gene target capture. See Table S5 for prediction of oncogenic drivers from whole exome data.

(A) Kruskal-Wallis test with Dunn's multiple comparison correction. Mean, 25th and 75th percentiles, minimum and maximum values are shown. Adjusted p values; ****p<0.0001, **p<0.01. **(B-D)** Log-rank analysis. **(G)** Fisher's exact test with Bonferroni correction. Asterisks indicate p<0.01.

Table S6. dNdScv analysis of exome data from T1 tumors. Related to Figure 3.

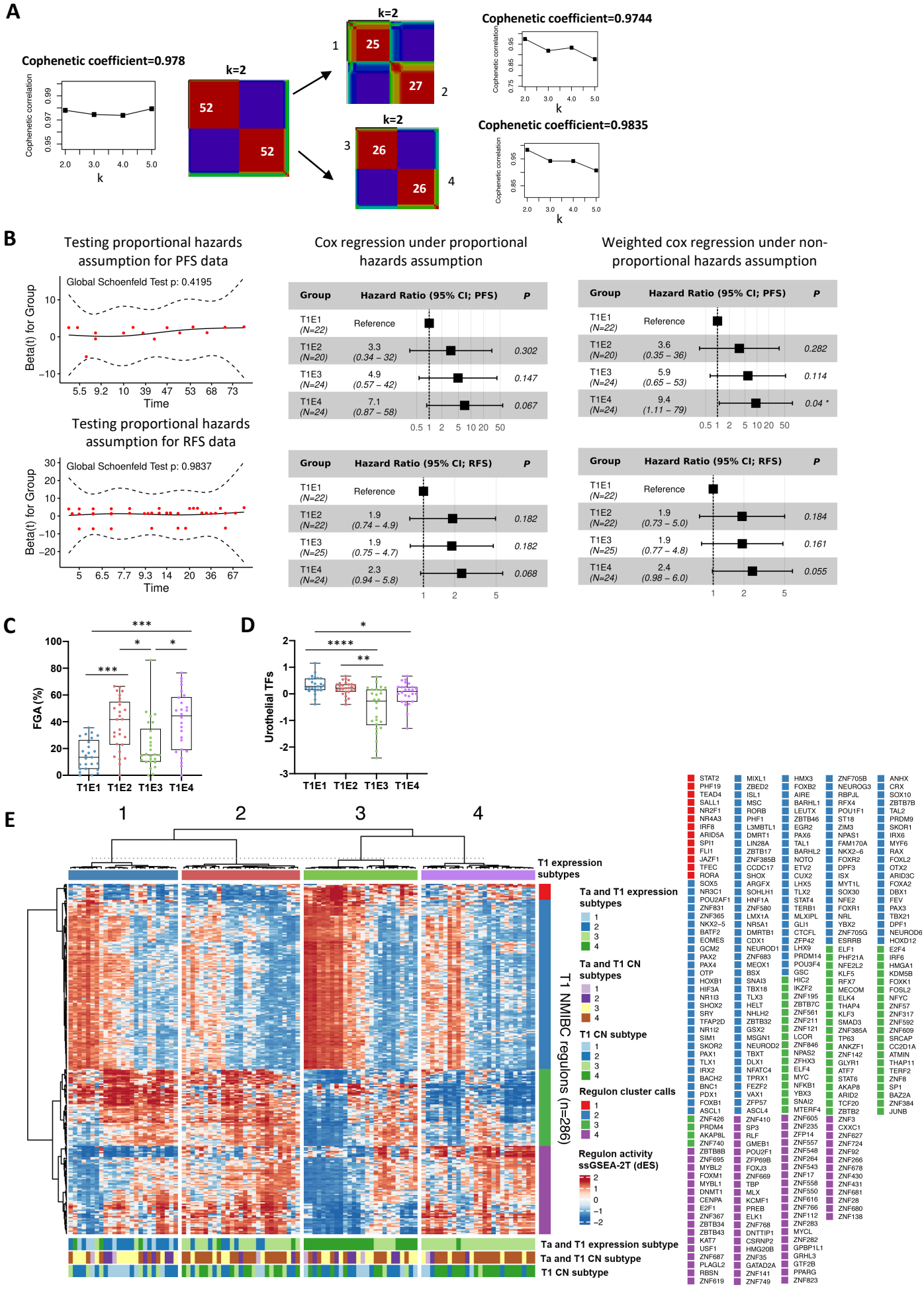
Gene	Synonymous mutations	Missense mutations	Nonsense mutations	Essential splice site	Indels	dN/dS – missense*	dN/dS – nonsense*	dN/dS – indels*	q-value missense mutations	q-value truncating mutations	q-value (all genes)	q-value (known bladder cancer genes)	significance all genes	known bladder cancer genes
KDM6A	2	4	9	2	12	1.485510846	28.06428314	59.51189176	0.792690954	1.84E-06	2.17E-12	6.88E-15	TRUE	TRUE
TP53	0	13	4	0	4	20.25419979	58.93851075	71.99831236	3.16E-05	0.010714376	6.84E-11	2.17E-13	TRUE	TRUE
ELF3	1	4	2	0	9	5.371672988	26.28859173	171.5766234	0.783760701	0.897278091	1.82E-09	5.76E-12	TRUE	TRUE
ARID1A	2	8	11	1	6	1.970017649	30.86657391	18.61373692	0.783760701	2.87E-07	4.48E-09	1.42E-11	TRUE	TRUE
FGFR3	2	21	0	0	2	11.86845137	0	17.53234553	1.77E-05	0.897278091	4.73E-06	1.50E-08	TRUE	TRUE
RB1	0	0	3	1	5	0	19.58206803	38.16918067	0.783760701	0.384810895	0.000228603	7.25E-07	TRUE	TRUE
CDKN1A	0	2	1	0	4	6.352842167	28.29451632	171.9232429	0.783760701	0.897278091	0.00043603	1.38E-06	TRUE	TRUE
KMT2D	2	13	8	2	7	1.785312361	14.09890579	8.964036904	0.783760701	0.002188136	0.00043603	1.38E-06	TRUE	TRUE
ZFP36L1	1	5	1	0	4	7.453908566	30.21378992	83.67945449	0.783760701	0.897278091	0.00043603	1.38E-06	TRUE	TRUE
ARID2	1	4	10	0	0	1.411549215	24.34705699	0	0.817148289	4.00E-05	0.000820136	2.60E-06	TRUE	TRUE
FAM229A	0	3	0	0	3	11.49345738	0	166.2148539	0.783760701	0.897357875	0.003971316	NA	TRUE	FALSE
ERCC2	1	13	0	0	1	9.508835961	0	9.31909825	0.020714612	0.897278091	0.028903276	0.0001057	FALSE	TRUE
SPA17	0	2	0	0	3	7.424679135	0	139.9704033	0.783760701	0.897278091	0.028903276	NA	FALSE	FALSE
TSC1	5	7	3	2	3	1.955453408	13.71759019	18.26223288	0.783760701	0.263760412	0.028903276	0.0001057	FALSE	TRUE
PTEN	0	2	2	0	3	3.11728669	28.404692	52.66213194	0.783760701	0.897278091	0.028903276	0.0001057	FALSE	TRUE
CREBBP	0	12	5	0	3	4.380993841	16.70750866	8.708760255	0.783760701	0.263760412	0.033449633	0.000121161	FALSE	TRUE
EP300	8	13	3	0	7	1.58981765	2.760988549	20.55603991	0.783760701	0.897278091	0.264231139	0.0012841	FALSE	TRUE
ERBB3	1	12	0	0	1	5.501520469	0	5.280591041	0.783760701	0.897278091	0.999804434	0.007835389	FALSE	TRUE
PIK3CA	1	10	0	0	1	5.452910411	0	6.634082103	0.783760701	0.897278091	0.999804434	0.011517583	FALSE	TRUE
STAG2	1	2	1	1	3	0.861942431	6.510764538	16.76556446	0.940156287	0.897278091	0.999804434	0.016367923	FALSE	TRUE
FBXW7	2	7	3	0	0	4.564784934	16.05640084	0	0.783760701	0.897278091	0.999804434	0.016460315	FALSE	TRUE
ERBB2	5	13	0	1	1	3.896005194	3.101184214	5.646364465	0.783760701	0.897278091	0.999804434	0.021392185	FALSE	TRUE
RHOA	1	3	0	0	1	8.248049147	0	36.55584416	0.783760701	0.897278091	0.999804434	0.024896042	FALSE	TRUE
KANSL1	0	0	2	0	1	0	11.5226294	6.412146264	0.783760701	0.897278091	0.999804434	0.025531368	FALSE	TRUE
GNA13	2	5	0	0	1	5.485490625	0	18.76146499	0.783760701	0.897278091	0.999804434	0.028946739	FALSE	TRUE

* Maximum likelihood estimates obtained by dNdScv for missense mutations, nonsense mutations and indels.
Grey shading indicates results from restricted hypothesis testing of known bladder cancer genes (qrht < 0.05).
FAM229A excluded as somatic missense mutations create a rare germline SNP of unknown significance.
SPA17 excluded as all mutations in same sample.

Table S7. *ERBB2* and *ERBB3* mutations and *ERBB2* copy number in T1 tumors. Related to Figure 4.

Sample	*ERBB2* mutation	*ERBB3* mutation	*ERBB2* copy number status	*ERCC2* mutation status	Expression subtype	Copy number subtype
1081	NA	NA	Normal	NA	1	1
1106	p.S310F; p.C311Sfs*10	p.P590S	Normal	p.E606K; p.V604Rfs*103	1	1
1137	Wildtype	Wildtype	Normal	p.T484M	1	1
1315	NA	NA	Normal	NA	1	1
1408	Wildtype	Wildtype	Loss	p.E86Q	1	1
2020	Wildtype	Wildtype	Normal	Wildtype	1	1
2046	Wildtype	Wildtype	Normal	Wildtype	1	1
2390	Wildtype	Wildtype	Normal	p.N238S	1	1
589	NA	NA	Normal	NA	2	1
1062	p.A1243T; c.1737+2T>C	p.G284R	Normal	Wildtype	2	1
571	NA	NA	Normal	NA	3	1
1279	p.D636Pfs*8	Wildtype	Normal	p.N238S	3	1
1318	Wildtype	Wildtype	Normal	p.T76I	3	1
1747	Wildtype	Wildtype	Normal	p.V242F	3	1
1831	p.N259S	p.E150K	Normal	p.S381F; p.T484M	3	1
2165	Wildtype	Wildtype	Normal	Wildtype	3	1
2437	Wildtype	Wildtype	Normal	Wildtype	3	1
1729	Wildtype	Wildtype	Normal	Wildtype	4	1
1766	Wildtype	p.M433I	Normal	p.G665A	4	1
2138	Wildtype	Wildtype	Normal	Wildtype	4	1
94	Wildtype	Wildtype	Normal	Wildtype	1	2
315	NA	NA	Normal	NA	1	2
557	Wildtype	Wildtype	Amp/Gain	Wildtype	1	2
866	NA	NA	Normal	NA	1	2
1094	NA	NA	Normal	NA	1	2
1210	NA	NA	Amp/Gain	NA	1	2
1222	Wildtype	Wildtype	Normal	Wildtype	1	2
1481	Wildtype	Wildtype	Normal	Wildtype	1	2
1968	Wildtype	Wildtype	Normal	Wildtype	1	2
2103	p.S310F	Wildtype	Amp/Gain	Wildtype	1	2
2272	Wildtype	Wildtype	Normal	Wildtype	1	2
1358	Wildtype	p.R679Q	Amp/Gain	Wildtype	2	2
1428	Wildtype	Wildtype	Normal	p.N238S	2	2
1457	Wildtype	Wildtype	Normal	Wildtype	2	2
1629	Wildtype	Wildtype	Normal	Wildtype	2	2
1978	Wildtype	p.V104L	Normal	Wildtype	2	2
2056	Wildtype	Wildtype	Amp/Gain	Wildtype	2	2
2155	Wildtype	Wildtype	Amp/Gain	Wildtype	2	2
555	Wildtype	Wildtype	Normal	Wildtype	3	2
936	Wildtype	Wildtype	Normal	Wildtype	3	2
940	Wildtype	Wildtype	Normal	Wildtype	3	2
1395	Wildtype	Wildtype	Normal	Wildtype	3	2
1443	Wildtype	Wildtype	Amp/Gain	Wildtype	3	2
1521	Wildtype	Wildtype	Normal	Wildtype	3	2
1633	Wildtype	Wildtype	Normal	Wildtype	3	2
1679	Wildtype	Wildtype	Normal	Wildtype	3	2
1810	Wildtype	Wildtype	Normal	p.P463S	3	2
1933	Wildtype	Wildtype	Amp/Gain	Wildtype	3	2
2175	Wildtype	Wildtype	Normal	p.E86Q	3	2
377	Wildtype	Wildtype	Normal	Wildtype	4	2
418	Wildtype	Wildtype	Normal	Wildtype	4	2
534	Wildtype	Wildtype	Normal	Wildtype	4	2
1234	Wildtype	Wildtype	Amp/Gain	Wildtype	4	2
1817	p.R103Q	p.Q298E	Amp/Gain	p.Q662H	4	2
1087	Wildtype	Wildtype	Normal	Wildtype	1	3
1138	NA	NA	Normal	NA	1	3
1145	NA	NA	Normal	NA	1	3
1325	Wildtype	Wildtype	Normal	p.L633V	1	3
1392	Wildtype	Wildtype	Normal	Wildtype	1	3
2429	Wildtype	Wildtype	Normal	Wildtype	1	3
567	Wildtype	Wildtype	Normal	Wildtype	2	3
1023	Wildtype	Wildtype	Normal	Wildtype	2	3

1252	Wildtype	Wildtype	Amp/Gain	Wildtype	2	3
1300	p.E766Q; p.G778D	p.G582E	Normal	Wildtype	2	3
1407	Wildtype	Wildtype	Normal	Wildtype	2	3
1517	Wildtype	Wildtype	Normal	Wildtype	2	3
1735	Wildtype	Wildtype	Normal	Wildtype	2	3
2022	Wildtype	Wildtype	Amp/Gain	Wildtype	2	3
1860	p.V697L	Wildtype	Amp/Gain	p.N238S	3	3
2356	Wildtype	Wildtype	Normal	Wildtype	3	3
2224	p.R103Q	p.Q298E	Normal	p.T76I	4	3
393	Wildtype	Wildtype	Amp/Gain	Wildtype	2	4
544	Wildtype	Wildtype	Amp/Gain	Wildtype	2	4
1211	Wildtype	Wildtype	Loss	Wildtype	2	4
1934	Wildtype	Wildtype	Amp/Gain	Wildtype	2	4
2000	Wildtype	Wildtype	Loss	Wildtype	2	4
2033	p.L651V	Wildtype	Normal	Wildtype	2	4
2139	p.S250F; p.S609Y	Wildtype	Amp/Gain	p.N238S	2	4
2169	Wildtype	Wildtype	Normal	Wildtype	2	4
2178	Wildtype	Wildtype	Normal	Wildtype	2	4
2222	Wildtype	p.K747N	Normal	Wildtype	2	4
132	p.E238Q; p.S310F	p.I600M	Normal	Wildtype	3	4
435	p.E507K	p.M91I; p.E150*	Amp/Gain	p.S246F	3	4
1047	Wildtype	Wildtype	Amp/Gain	Wildtype	3	4
1124	Wildtype	Wildtype	Amp/Gain	Wildtype	3	4
1974	Wildtype	Wildtype	Normal	Wildtype	3	4
2058	Wildtype	Wildtype	Amp/Gain	Wildtype	3	4
411	NA	NA	Amp/Gain	NA	4	4
930	NA	NA	Amp/Gain	NA	4	4
1016	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
1249	Wildtype	Wildtype	Normal	Wildtype	4	4
1419	Wildtype	p.V104L	Normal	p.T484M	4	4
1503	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
1530	Wildtype	Wildtype	Normal	Wildtype	4	4
1559	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
1745	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
1853	p.D277H	Wildtype	Amp/Gain	Wildtype	4	4
1947	Wildtype	Wildtype	Amp/Gain	p.R658G	4	4
2037	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
2064	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4
2214	Wildtype	p.L74F	Amp/Gain	p.Y24C	4	4
2223	Wildtype	p.D863N	Normal	p.E606Q	4	4
2267	Wildtype	Wildtype	Normal	p.N238S	4	4
2423	Wildtype	Wildtype	Amp/Gain	Wildtype	4	4


A
Cophenetic coefficient=0.978
Cophenetic correlation
k
k=2
52
52
1
25
27
2
3
26
26
4
Cophenetic coefficient=0.9744
Cophenetic coefficient=0.9835
B
Testing proportional hazards assumption for PFS data
Global Schoenfeld Test p: 0.4195
Beta(t) for Group
Time
Testing proportional hazards assumption for RFS data
Global Schoenfeld Test p: 0.9837
Cox regression under proportional hazards assumption
Weighted cox regression under non-proportional hazards assumption
Group
Hazard Ratio (95% CI; PFS)
P
T1E1 (N=22) Reference
T1E2 (N=20) 3.3 (0.34 – 32) 0.302
T1E3 (N=24) 4.9 (0.57 – 42) 0.147
T1E4 (N=24) 7.1 (0.87 – 58) 0.067
T1E1 (N=22) Reference
T1E2 (N=20) 3.6 (0.35 – 36) 0.282
T1E3 (N=24) 5.9 (0.65 – 53) 0.114
T1E4 (N=24) 9.4 (1.11 – 79) 0.04 *
Group
Hazard Ratio (95% CI; RFS)
P
T1E1 (N=22) Reference
T1E2 (N=22) 1.9 (0.74 – 4.9) 0.182
T1E3 (N=25) 1.9 (0.75 – 4.7) 0.182
T1E4 (N=24) 2.3 (0.94 – 5.8) 0.068
T1E1 (N=22) Reference
T1E2 (N=22) 1.9 (0.73 – 5.0) 0.184
T1E3 (N=25) 1.9 (0.77 – 4.8) 0.161
T1E4 (N=24) 2.4 (0.98 – 6.0) 0.055
C
FGA (%)
T1E1 T1E2 T1E3 T1E4
D
Urothelial TFs
T1E1 T1E2 T1E3 T1E4
E
1 2 3 4
T1 expression subtypes
Ta and T1 expression subtypes
Ta and T1 CN subtypes
T1 CN subtype
Regulon cluster calls
Regulon activity ssGSEA-2T (dES)
T1 NMIBC regulons (n=286)
Ta and T1 expression subtype
Ta and T1 CN subtype
T1 CN subtype
STAT2 PHF19 TEAD4 SALL1 NR2F1 NR4A3 IRF8 ARID5A SPI1 FLI1 JAZF1 TFEC RORA SOX5 NR3C1 POU2AF1 ZNF831 ZNF365 NKX2-5 BATF2 EOMES GCM2 PAX2 PAX4 OTP HOXB1 HIF3A NR1I3 SHOX2 SRY TFAP2D NR1I2 SIM1 SKOR2 PAX1 TLX1 IRX2 BACH2 BNC1 PDX1 FOXB1 ASCL1 ZNF426 PRDM4 AKAP8L ZNF740 ZBTB8B ZNF695 MYBL2 FOXM1 MYBL1 DNMT1 CENPA E2F1 ZNF367 ZBTB34 ZBTB43 KAT7 USF1 ZNF687 PLAGL2 RBSN ZNF619
MIXL1 ZBED2 ISL1 MSC RORB PHF1 L3MBTL1 DMRT1 LIN28A ZBTB17 ZNF385B CCDC17 SHOX ARGFX SOHLH1 HNF1A ZNF580 LMX1A NR5A1 DMRTB1 CDX1 NEUROD1 ZNF683 MEOX1 BSX SNAI3 TBX18 TLX3 HELT NHLH2 ZBTB32 GSX2 MSGN1 NEUROD2 TBXT DLX1 NFATC4 TPRX1 FEZF2 VAX1 ZFP57 ASCL4 ZNF410 SP3 RLF GMEB1 POU2F1 ZFP69B FOXJ3 ZNF669 TBP MLX KCMF1 PREB ELK1 ZNF768 DNTTIP1 CSRNP2 HMG20B ZNF35 GATAD2A ZNF141 ZNF749
HMX3 FOXB2 AIRE BARHL1 LEUTX ZBTB46 EGR2 PAX6 TAL1 BARHL2 NOTO ETV2 CUX2 LHX5 TLX2 STAT4 TERB1 MLXIPL GLI1 CTCFL ZFP42 LHX9 PRDM14 POU3F4 GSC HIC2 IKZF2 ZNF195 ZBTB7C ZNF561 ZNF211 ZNF121 LCOR ZNF846 NPAS2 ZFHX3 ELF4 MYC NFKB1 YBX3 SNAI2 MTERF4 ZNF605 ZNF235 ZFP14 ZNF557 ZNF548 ZNF264 ZNF543 ZNF17 ZNF558 ZNF550 ZNF616 ZNF766 ZNF112 ZNF283 MYCL ZNF282 GPBP1L1 GRHL3 GTF2B PPARG ZNF823
ZNF705B NEUROG3 RBPJL RFX4 POU1F1 ST18 ZIM3 NPAS1 FAM170A NKX2-6 FOXR2 DPF3 ISX MYT1L SOX30 NFE2 FOXR1 NRL YBX2 ZNF705G ESRRB ELF1 PHF21A NFE2L2 KLF5 RFX7 MECOM ELK4 THAP4 KLF3 SMAD3 ZNF385A TP63 ANKZF1 ZNF142 GLYR1 ATF7 STAT6 AKAP8 ARID2 TCF20 ZBTB2 ZNF3 CXXC1 ZNF627 ZNF724 ZNF92 ZNF266 ZNF678 ZNF430 ZNF431 ZNF681 ZNF28 ZNF680 ZNF138
ANHX CRX SOX10 ZBTB7B TAL2 PRDM9 SKOR1 IRX6 MYF6 RAX FOXL2 OTX2 ARID3C FOXA2 DBX1 FEV PAX3 TBX21 DPF1 NEUROD6 HOXD12 E2F4 IRF6 HMGA1 KDM5B FOXK1 FOSL2 NFYC ZNF57 ZNF317 ZNF592 ZNF609 SRCAP CC2D1A ATMIN THAP11 TERF2 ZNF8 SP1 BAZ2A ZNF384 JUNB

Supplementary Figure 6. Derivation and features of T1 tumor expression subtypes. Related to Figure 5.

(A) Derivation of four stage T1 expression subtypes by two stage NMF analysis, showing numbers of tumors in clusters and plots of cophenetic coefficients among k clusters for each analysis.

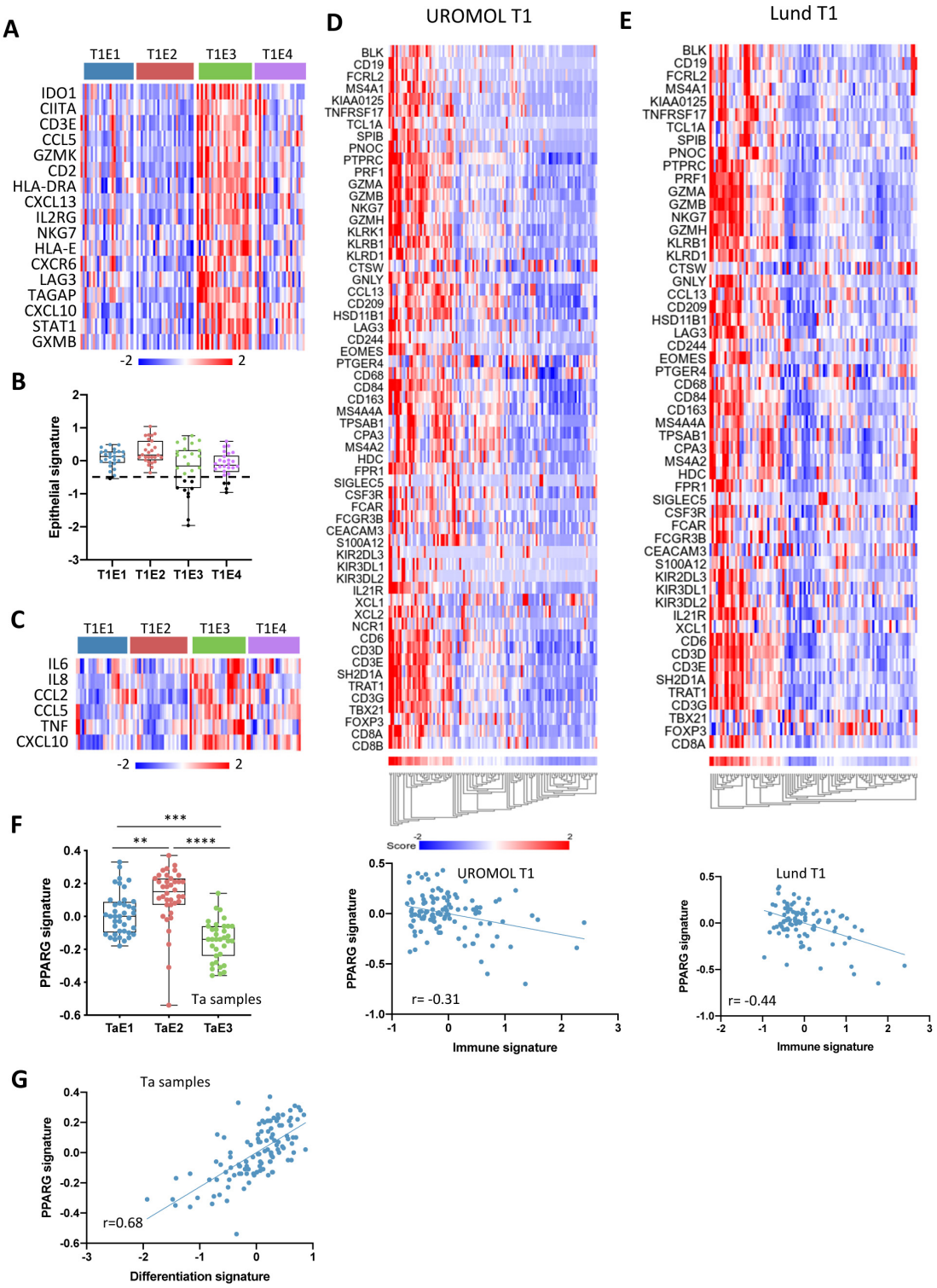
(B) Left, Schoenfeld plots for RFS and PFS data for T1 expression subtypes (see STAR Methods). Middle and right, forest plots showing results of Cox regression analyses under proportional and non-proportional hazards assumptions for RFS and PFS of T1 expression subtypes using the 'coxph' function (Therneau, 2021; R package version 3.2-13, 2021. https://CRAN.R-project.org/package=survival) and 'coxphw' [S3].

(C) Fraction of genome altered (FGA) according to T1 expression subtypes.

(D) Scores for expression of transcription factors (TFs) implicated in urothelial differentiation according to T1 expression subtypes.

(E) Heatmap of regulon activity for 286 regulons with differential activity in stage T1 tumors supervised by T1 expression subtype. Expression and genomic covariates are shown below.

(C and **D)** Kruskal-Wallis test with Dunn's multiple comparison correction. Mean, 25th and 75^{th} percentiles, minimum and maximum values are shown. Adjusted p values; ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05.


A
T1E1
T1E2
T1E3
T1E4
IDO1
CIITA
CD3E
CCL5
GZMK
CD2
HLA-DRA
CXCL13
IL2RG
NKG7
HLA-E
CXCR6
LAG3
TAGAP
CXCL10
STAT1
GXMB
-2
2
B
Epithelial signature
T1E1
T1E2
T1E3
T1E4
C
IL6
IL8
CCL2
CCL5
TNF
CXCL10
D
UROMOL T1
BLK
CD19
FCRL2
MS4A1
KIAA0125
TNFRSF17
TCL1A
SPIB
PNOC
PTPRC
PRF1
GZMA
GZMB
NKG7
GZMH
KLRK1
KLRB1
KLRD1
CTSW
GNLY
CCL13
CD209
HSD11B1
LAG3
CD244
EOMES
PTGER4
CD68
CD84
CD163
MS4A4A
TPSAB1
CPA3
MS4A2
HDC
FPR1
SIGLEC5
CSF3R
FCAR
FCGR3B
CEACAM3
S100A12
KIR2DL3
KIR3DL1
KIR3DL2
IL21R
XCL1
XCL2
NCR1
CD6
CD3D
CD3E
SH2D1A
TRAT1
CD3G
TBX21
FOXP3
CD8A
CD8B
Score
UROMOL T1
PPARG signature
r= -0.31
Immune signature
E
Lund T1
Lund T1
PPARG signature
r= -0.44
Immune signature
F
***
**
****
PPARG signature
Ta samples
TaE1
TaE2
TaE3
G
Ta samples
PPARG signature
r=0.68
Differentiation signature

Supplementary Figure 7. Relationship of PPARG signature to immune infiltration in NMIBC. Related to Figure 6.

(A) Heatmap of z-scores for genes in a T-cell inflamed signature [S4] in T1 expression subtypes.

(B) Expression scores for a panel of epithelial marker genes (see Methods) used to provide an estimate of epithelial tumor cell content in T1 samples. Samples with zscores < -0.05 indicated by black dots were removed from analysis of the relationship of PPARG signature and tumor infiltration. Mean, 25th and 75th percentiles, minimum and maximum values are shown.

(C) Heatmap of z-scores for pro-inflammatory chemokines that attract effector T cells that were shown to be regulated by PPARG in T24 bladder tumor cells [S5] in T1 expression subtypes.

(D) Heatmap of z-scores for 60-gene immune signature [S6] in T1 tumors reported in the first UROMOL study [S7] (ArrayExpress: E-MTAB-4321). Graph below shows correlation of PPARG and immune signatures scores.

(E) Heatmap of z-scores for 60-gene immune signature in T1 tumors from the study of Sjodahl *et al* [S8] (GSE: 32894). Graph below shows correlation of PPARG and immune signatures scores.

(F) PPARG signature score in Ta expression subtypes.

(G) Correlation of PPARG signature and urothelial differentiation signature in Ta samples.

(F) Kruskal-Wallis test test with Dunn's multiple comparison correction. Mean, 25th and 75th percentiles, minimum and maximum values are shown. Adjusted p values; ****p<0.0001, ***p<0.001, **p<0.01. **(D** and **E** lower panels and **G)** Pearson r.

Supplemental references

S1. Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep *18*, 248-262. 10.1016/j.celrep.2016.12.019.

S2. Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue, T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell, 540-556. 10.1016/j.cell.2017.09.007.

S3. Dunkler, D., Ploner, M., Schemper, M., and Heinze, G. (2018). Weighted Cox Regression Using the R Package coxphw. J Statistical Software *84*. doi: 10.18637/jss.v084.i02.

S4. Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. J. Clin. Invest. *127*, 2930-2940. 10.1172/JCI91190.

S5. Korpal, M., Puyang, X., Jeremy Wu, Z., Seiler, R., Furman, C., Oo, H.Z., Seiler, M., Irwin, S., Subramanian, V., Julie Joshi, J., et al. (2017). Evasion of immunosurveillance by genomic alterations of PPARgamma/RXRalpha in bladder cancer. Nat Commun *8*, 103. 10.1038/s41467-017-00147-w.

S6. Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M.L., Geller, M.A., Odunsi, K., Beechem, J., and Fling, S.P. (2017). Gene expression markers of Tumor Infiltrating Leukocytes. Journal for immunotherapy of cancer *5*, 18. 10.1186/s40425-017-0215-8.

S7. Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Hoyer, S., Ulhoi, B.P., Vang, S., Reinert, T., Hermann, G.G., Mogensen, K., et al. (2016). Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell *30*, 27-42. 10.1016/j.ccell.2016.05.004.

S8. Sjodahl, G., Lauss, M., Lovgren, K., Chebil, G., Gudjonsson, S., Veerla, S., Patschan, O., Aine, M., Ferno, M., Ringner, M., et al. (2012). A molecular taxonomy for urothelial carcinoma. Clin. Cancer Res. *18*, 3377-3386. 10.1158/1078-0432.CCR-12-0077-T.