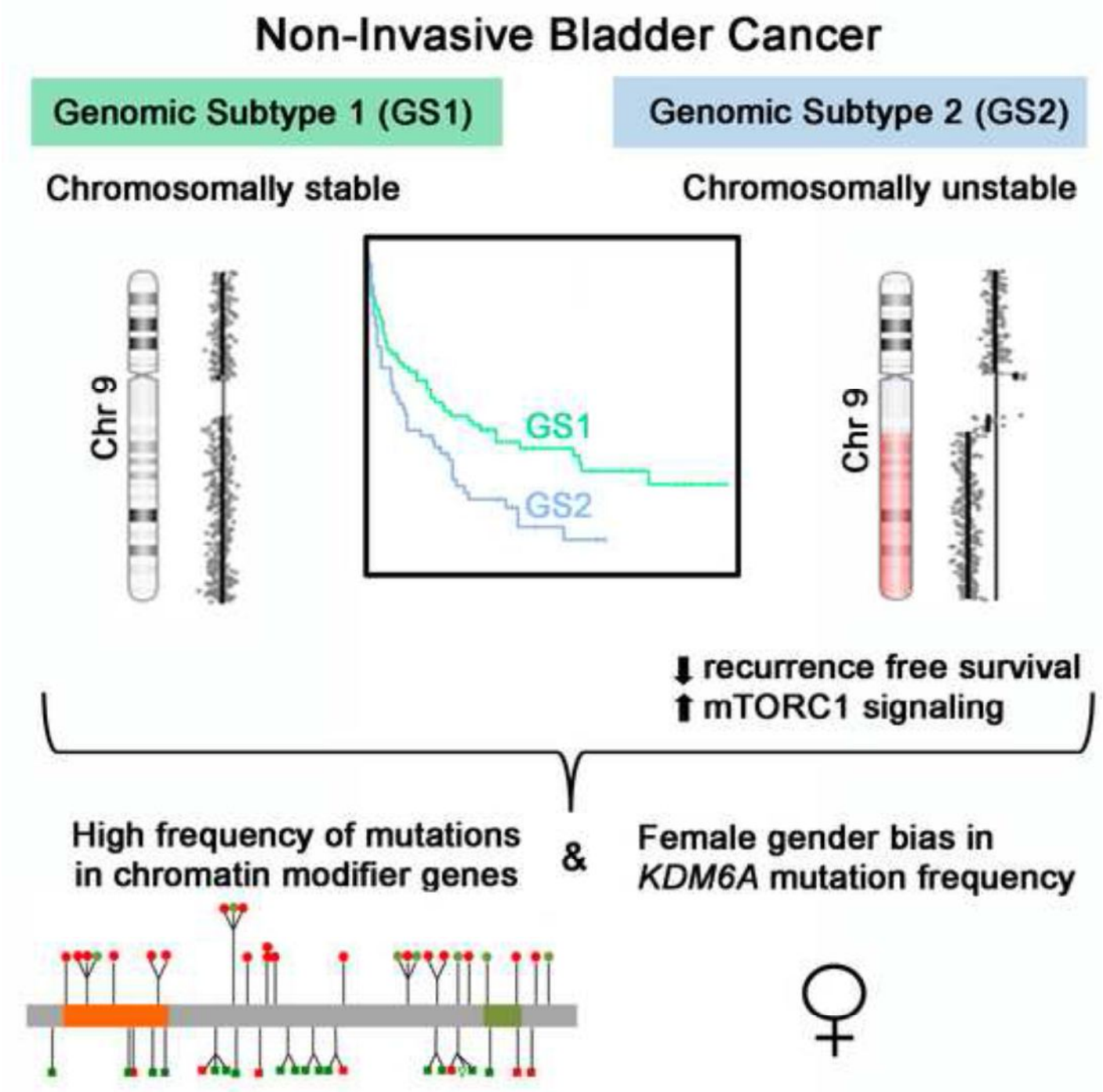
Graphical Abstract

Click here to download Graphical Abstract Graphical abstract.tif



Genomic Subtypes of Non-Invasive Bladder Cancer with Distinct Metabolic Profile, Clinical Outcome and Female Gender Bias in *KDM6A* Mutation Frequency

Carolyn D. Hurst[1,2], Olivia Alder[1,2], Fiona M. Platt[2], Alastair Droop[3], Lucy F. Stead[4], Julie E. Burns[2], George J. Burghel[5], Sunjay Jain[6], Leszek J. Klimczak[7], Helen Lindsay[5], Jo-An Roulson[8], Claire F. Taylor[3], Helene Thygesen[3], Angus J. Cameron[9], Anne J. Ridley[9,10], Helen R. Mott[11], Dmitry A. Gordenin[12] and Margaret A. Knowles[2].

[1]These authors contributed equally
[2]Section of Molecular Oncology, Leeds Institute of Cancer and Pathology, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK.
[3]Cancer Research UK Leeds Centre, Leeds Institute of Cancer and Pathology, St. James's University Hospital, Leeds LS9 7TF, UK.
[4]Section of Oncology and Clinical Research, Leeds Institute of Cancer and Pathology, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK.
[5]DNA Laboratory, Genetics Service, Ashley Wing, St James University Hospital, Leeds, LS9 7TF, UK.
[6]Pyrah Department of Urology, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK.
[7]Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709, USA.
[8]Department of Histopathology, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK.
[9]Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK
[10]Randall Division of Cell and Molecular Biophysics, New Hunt's House, King's College London, Guy's Campus, London SE1 1UL, UK.
[11]Department of Biochemistry, 80, Tennis Court Road, Cambridge, CB2 1GA, UK.
[12]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709, USA.

Correspondence and Lead Contact: m.a.knowles@leeds.ac.uk

SUMMARY

Bladder cancer incurs a higher lifetime treatment cost than other cancers due to frequent recurrence of non-invasive disease. Improved prognostic biomarkers and localised therapy are needed for this large patient group. We defined two major genomic subtypes of primary stage Ta tumors that showed differential risk of recurrence. The higher risk subtype was characterised by loss of 9q including *TSC1,* increased KI67 labelling index, upregulated glycolysis, DNA repair, mTORC1 signaling, features of the unfolded protein response and altered cholesterol homeostasis. Comparison with muscle-invasive bladder cancer mutation profiles revealed lower overall mutation rates and more frequent mutations in *RHOB* and chromatin modifier genes. More mutations in the histone-lysine demethylase *KDM6A* were present in non-invasive tumors from females than males.

SIGNIFICANCE

Non-invasive (stage Ta) bladder tumors are the largest group of bladder tumors and although there is heterogeneity in outcome, these are currently treated similarly. We identified two major genomic subtypes, one of which is associated with distinct DNA copy number and gene mutation profile and expression changes that appear related to loss of expression of one or both copies of *TSC1.* This identifies a metabolic vulnerability with potential for therapeutic intervention in >50 % of cases. We also identified a gender-related molecular difference. The high frequency of mutation of the histone demethylase *KDM6A* in female tumors may indicate a distinct epigenetic landscape in the female bladder.

INTRODUCTION

The majority of bladder tumors are non-muscle-invasive bladder cancers (NMIBC). Of the estimated 429,000 new cases of bladder cancer diagnosed each year worldwide (Ferlay et al., 2015), approximately 75% are NMIBC (Babjuk et al., 2017). NMIBC include high-grade tumors that invade the submucosa (stage T1) and carcinoma *in situ* (CIS), which have high risk of progression to muscle invasion and may require cystectomy. However, the largest group of NMIBC are non-invasive bladder cancers (NIBC) that do not penetrate the epithelial basement membrane (stage Ta). The recommended treatment for low risk Ta tumors is resection followed by a single instillation of a chemotherapy agent, and for intermediate risk tumors, induction and maintenance intravesical chemotherapy or BCG (Babjuk et al., 2017).

Patients with NIBC have long life-expectancy. However local disease recurrence in up to 70% of cases necessitates long-term disease monitoring via cystoscopy. Consequently, bladder cancer has high prevalence (Ploeg et al., 2009) and incurs a higher lifetime cost for treatment than all other cancers (James and Gore, 2013; Svatek et al., 2014). Assessment of risk of recurrence and progression and associated treatment decisions are based on a nomogram that includes tumor size, grade, multifocality and the presence of CIS (Sylvester et al., 2006). This stratification is not yet complemented by molecular biomarkers and no biomarker-guided localised treatments have been developed.

Recent studies have used gene expression analysis to sub-classify bladder tumors and seek prognostic signatures. In muscle-invasive bladder cancer (MIBC), expression subgroups show relationships to outcome and therapeutic response (Choi et al., 2014; Damrauer et al., 2014; Cancer Genome Atlas Research Network, 2014a). Less information is available for NMIBC. However, a study that included both NMIBC and MIBC showed segregation of NMIBC across several expression subgroups (Sjodahl et al., 2012) and recently a large study of the entire spectrum of NMIBC confirmed this molecular heterogeneity and defined three expression classes with prognostic implications (Hedegaard et al., 2016).

Although there has been extensive analysis of the genome of MIBC (Gui et al., 2011; Guo et al., 2013; Nellist et al., 1999; Cancer Genome Atlas Research Network, 2014a), only small numbers of stage Ta NIBC have been analysed by genome sequencing (Balbas-Martinez et al., 2013; Nordentoft et al., 2014). Common features identified by candidate gene analysis include mutations in the *TERT* promoter, activating point mutations in *FGFR3* and *PIK3CA,* and inactivating mutations in *STAG2* (Reviewed in (Knowles and Hurst, 2015)). Our previous genome-wide copy number analysis revealed two major genomic sub-groups of stage Ta tumors (Hurst et al., 2012). However, knowledge of their mutational landscape remains incomplete and definition of subtypes based on mutational signatures has not been attempted. Identification of biomarkers for increased risk of recurrence may allow modification of disease monitoring and improved biological understanding may suggest approaches to localised therapy for these patients. Here we investigate the relationships of genomic copy number profile, expression subtype membership and mutation signature with patient gender and outcome in NIBC.

RESULTS

We analysed 140 primary stage Ta transitional cell carcinomas and matched blood. These were selected based on initial diagnostic assessment as grade 1 or 2 according to the 1973 WHO classification (WHO, 1973). All were re-evaluated for this study by a single pathologist (J-A R). Clinicopathologic data including 1973 and 2004 WHO classifications (Eble et al., 2004; WHO, 1973) are summarised in Table S1. The majority (134/140) were grade 1/2 according to the 1973 classification system; 61% were low and 39% high grade according to the 2004 system. Median age at diagnosis was 71, and the male:female ratio was 2.06. We also analysed a recently-established grade 2 stage Ta tumor-derived cell line (LUCC8, patient 1838) and matched immortalised lymphoblasts (total sample number = 141). Follow-up ranged from 1-200 months (median 69).

Association of Genomic Subtype with Recurrence-Free Survival

Previously, we identified two copy number groups of stage Ta tumors (Hurst et al., 2012). Here, using unsupervised clustering on low pass whole genome sequencing and array CGH data on this extended sample series of 141 (including 17 from the previous study), we confirmed the existence of two major copy number subtypes (Figure 1A). Seventy-eight tumors contained no or few copy number alterations (Genomic subtype 1; GS1). The remaining 63, with more alterations, were distinguished primarily by high frequency of chromosome 9 deletion, with 9q deletion in all but eight samples (GS2) (Table S1). Repeated clustering omitting chromosome 9 showed that potential impact of other common regions of copy number alteration was not masked by loss of 9.

GS2 assignment was strongly associated with the 2004 high grade classification ($p=0.0053$). Disease recurrence occurred in 47.5% of patients (GS1, 45%; GS2, 51%). As recurrence is the major clinical problem in NIBC, we examined the relationship of genomic subtype to recurrence-free survival (RFS). This revealed a significant difference between GS1 and GS2 ($p=0.0059$) (Figure 1B). We also assessed the relationship of RFS to 9q loss and 2004 grade. Both showed associations but these were less significant (Figure 1B). This suggests that genomic subtype has clinical relevance and may provide superior prognostic information to grade.

Subtype-Specific Gene Expression

To determine whether there is divergence between GS1 and GS2 tumors at the mRNA expression level, we conducted microarray analysis on samples from each class (n=79; GS1=48, GS2=31). We identified 363 differentially expressed genes (limma $q<0.01$), of which 21% were upregulated and 79% downregulated in GS2 (Table S2). As expected, the majority of differentially expressed genes (n=248) were located on chromosome 9. The remainder were scattered throughout the genome (Figure 2A).

Gene Set Enrichment Analysis showed significant overrepresentation of genes involved in mTORC1 signaling ($p=0.006$) (Figure 2B). This is compatible with loss of one

copy of 9q including the mTORC1 negative regulator *TSC1* in the majority of GS2 samples. Only three GS1 samples had loss of 9q sequences and all retained the *TSC1* region. Overrepresentation of genes in processes influenced by upregulated mTORC1 signaling included those involved in the unfolded protein response (p=0.031) (Figure 2C), cholesterol homeostasis (p=0.032) (Figure 2C and D) and glycolysis (p=0.044) (Figure 2C). The same processes were identified when chromosome 9 genes were omitted from the analysis (p=0.008, 0.012 and 0.031 respectively).

To examine alignment of the expression profiles of these genomic subtypes with previously identified NMIBC expression subtypes, we conducted nearest neighbour analysis (knn1) using the Lund dataset of 308 NMIBC and MIBC samples (Sjodahl et al., 2012) as a training set. Although our dataset has a distinct composition from this set, the analysis indicated that our cohort aligned primarily with the “Urobasal A” subtype (40/48 GS1 and 24/31 GS2) (Figure S1A) and there was a trend towards membership of the Lund MS1a subgroup of Urobasal A in GS1 samples (33/40 GS1 UroA vs 14/24 GS2 UroA; p=0.04).

Recent RNA-Seq analysis of NMIBC (stage Ta, T1 and CIS) by the UROMOL study identified three expression classes (Hedegaard et al., 2016). Implementation of the UROMOL 117 gene classifier and subsequent consensus clustering indicated the presence of 3+ subtypes in our cohort, and a bias in membership according to GS assignment (Figure S1A and S1B). In particular, GS2 samples showed significant bias towards UROMOL class 2, (p=0.0002). Similarly, nearest neighbour analysis indicated that all samples in this dataset align with the UROMOL class 2 subtype. Examination of differential expression between the three identified consensus groups and subsequent gene ontology analysis revealed enrichment for cell cycle and DNA repair genes (p=2.3e-119 and 3.2e-28 respectively) in the class showing a bias for GS2 membership (Figure S1C). This subtype also showed increased KI67 labelling index (Figure S1D), which was accompanied by expression of CCNB1 in both basal and suprabasal cells of the urothelium (Figure S2).

No significant enrichment of differentiation or cancer stem cell-associated genes was detected. Consistent with this, IHC analysis detected positive staining for the urothelial

differentiation-associated markers GATA3, FOXA1 and PPARG that are characteristic of luminal/Urobasal A subtypes described by others (Choi et al., 2014; Damrauer et al., 2014; Cancer Genome Atlas Research Network, 2014a; Sjodahl et al., 2012) in all GS1 and GS2 tumors tested (Figure S2). In summary, mRNA profiling of GS1 and GS2 tumors confirmed the overall luminal status of the samples and revealed that these subtypes of Ta tumors are distinguished by differential mTORC1 signaling.

Mutational Profile

To determine whether distinct mutational signatures underlie copy number class membership, we carried out whole exome and targeted sequencing analysis. Exome analysis of 24 samples (17 GS1; 7 GS2) identified 2973 somatic mutations [2851 single nucleotide substitutions and 122 insertions or deletions (indels)], with an average of 124 ± 79 mutations per sample (range 40-285), giving mean and median somatic mutation rates of 2.41 and 1.64 per megabase (Mb) (Table S3 and Figure S3A). Samples contained an average of 119 (37-283) single nucleotide variants and 5 (0-11) indels. Missense, synonymous and loss of function mutations (nonsense, frameshift or mutations of the invariant dinucleotide at splice junctions) showed averages of 75 (24-180), 25 (9-55) and 14 (3-27) per sample respectively (Figure S3A).

Overall, 47% of nucleotide substitutions were C>T transitions, followed by C>G transversions (24.8%) (Figure S3B). Enrichment of a mutational signature compatible with the action of the APOBEC family of cytidine deaminases (Burns et al., 2013; Roberts et al., 2013) is prevalent in both MIBC (Cancer Genome Atlas Research Network, 2014a) and NMIBC (Acar et al., 2015; Hedegaard et al., 2016; Nordentoft et al., 2014). Here, 18 samples (75%) had up to 5-fold enrichment of this signature (Figures 3A and S3C). The fraction of these mutations reached 60% in some samples and APOBEC signature mutations represented 35% of all mutations. APOBEC cytidine deaminase(s) can cause mutations only in G:C base pairs. The number of these mutations and of mutations with stringent APOBEC signature strongly correlated with the total number of mutations in each

sample except in two outliers (Figure S3D), one of which contained two POLE missense mutations (1117), the other a missense mutation in POLE2 (1058).

Nonsynonymous mutations were present in 1949 genes, 208 of which were mutated in ≥2 samples. Inactivating mutations were present in 286 genes, 21 of which showed mutation in more than one sample (Table S3). GenomeMuSic analysis (Dees et al., 2012) revealed significant mutation rates in the order *PIK3CA*, *FGFR3, KDM6A, STAG2, HRAS, EP300, RHOB* and *ARID1A* (Table S4). Three samples showed missense mutations in *RHOB*, two of which were identical (E47K). Other genes, including chromatin modifiers *KMT2A, KMT2D* and *UTY*, the RNA-binding protein *RBM10*, and *BRCA2*, showed ≥ 3 mutations but did not reach statistical significance. The majority of mutations in chromatin modifier genes including *KDM6A, KMT2D, KMT2C, CREBBP, EP300, ARID1A* and *STAG2* were inactivating (Figure S4A).

The frequent loss of chromosome 9 in bladder cancer has led to many studies aimed at identifying candidate tumor suppressor genes. Two 9q genes, *TSC1* and *NOTCH1,* are inactivated by mutation in bladder cancer (Platt et al., 2009; Rampias et al., 2014). As the detected mutation frequencies of these genes are significantly lower than the overall frequency of 9q loss, it is possible that other 9q target genes remain to be identified. Eight samples analysed by exome sequencing had copy number loss of 9q. One of these had a nonsense mutation in *TSC1* but no mutations were identified in *NOTCH1*. No other 9q gene was mutated more than once.

To determine if particular gene categories were significantly over-represented, we conducted unbiased Gene Ontology analysis, using genes mutated in two or more samples as input. Two gene ontology networks featuring purine nucleoside binding (including RHOB) (p=1.20 E-06) and chromatin modification (p=9.92 E-07) dominated. Forty-nine genes were selected to determine prevalence in an additional 58 samples using custom target capture. Selection was based on significance, frequency of mutation in the discovery series, type of mutation, known functional roles, and comparison with the Cancer Gene Census (Futreal et al., 2004) (Table S4). *FGFR3, KRAS, NRAS, HRAS* and *TP53* mutation data were acquired

independently. Nonsynonymous mutations in 28 genes were identified in ≥ 5% of samples. Data for these genes in the 82 tumors analysed by exome and targeted sequencing is shown in Figure 3B.

Strikingly, 17 of 48 mutations in *PIK3CA* were not in the major hotspot codons (E542, E545, H1047) identified in bladder and other tumors (Platt et al., 2009). Several of these, including 4 of 5 in the C2 domain, have been reported in other tumor types, two of which (N345K, C420R) have experimentally confirmed gain of function (Gymnopoulos et al., 2007) (Figure S4B).

Mutations in *KMT2C* and *KMT2D* were significantly concordant (p=6.61E-05, FDR<0.05) (Figure S5), illustrating lack of redundancy between these components in this context. As observed previously (Jebar et al., 2005), RAS and *FGFR3* mutations were mutually exclusive (p=7.47E-11, FDR<0.05). We also observed mutual exclusivity between RAS and *KDM6A* mutations (p=1.10E-03), likely reflecting the common co-occurrence of *KDM6A* and *FGFR3* mutations (p=2.06E-03) (Figure S5).

Ten somatic mutations in *BRCA2* were identified in eight samples. Two of these are predicted to be pathogenic (nonsense). Three of the 7 missense variants were predicted as deleterious or potentially deleterious using SIFT (Kumar et al., 2009) and/or PolyPhen (Adzhubei et al., 2013) algorithms, one of which (T2662K) has been reported as pathogenic when found in the germline (Shirts et al., 2016). We also assessed germline sequences and identified one potentially pathogenic variant (rs11571833; K3326*) in a patient with a somatic mutation (Table S5). This variant has previously been associated with risk of breast and ovarian cancer (Meeks et al., 2016). Four missense variants were present in both germline and tumor DNA, of which 3 were predicted to be deleterious; none of these have been reported as pathogenic.

Relationship of Genomic Class to Mutational Landscape

We examined mutation frequency, mutation signatures and affected genes in the context of copy number class membership. Mutation rate per megabase derived from whole

exome sequencing data was higher in tumors assigned to GS2 (Mean 1.85 vs 3.75; p=0.0013) (Figure 4A). Consistent with this, GS2 tumors were also more likely to contain significant APOBEC-related signatures (Figure 4B). APOBEC3H was the only member of this gene family that showed differential transcript levels between the two copy number classes (Figure 4C). We examined the possibility that the detected difference in mutation frequency could in part be due to mutations in DNA repair genes. Mutations in candidates such as *BRCA2* were equally distributed in the two classes. A few genes (e.g. *UEVLD*, *USP47*) were exclusively mutated in one class and not the other, but the total number of mutations in these genes was low (n=3 and 4 of 82 respectively). We assessed the relationship between specific mutations and RFS, and no associations were found ($p>0.05$; $q>0.01$).

FGFR3 and *PIK3CA* mutation frequencies were not significantly different in the two subtypes, though *HRAS* mutations were more frequent in GS1 (20% vs 3%; p=0.021) (Figure 4D). Consistent with mTORC1 pathway gene expression enrichment in GS2, mutations in *TSC1* were more common in GS2 (Figure 4E) (4% in GS1 vs 19% in GS2; p=0.0383). Mutations in many genes associated with chromatin regulation were present in both genomic subtypes (Figure 4F and G). However, *STAG2* mutations were more frequent in GS1 (46% vs 25%) (Figure 4G). This higher frequency of *STAG2* inactivation in GS1 was confirmed by immunohistochemistry with 33% of GS1 tumors showing no positive staining in contrast to 20% of GS2 (not shown).

Comparison of the Mutational Landscape of Non-Invasive and Muscle-Invasive Bladder Cancer

We compared mutation frequencies in NIBC with those found in MIBC (Cancer Genome Atlas Research Network, 2014a) (http://www.cbioportal.org/public-portal/index.do). This revealed both known differences, including absence of *TP53* mutations and high frequency of *FGFR3, PIK3CA* and *STAG2* mutations in stage Ta tumors, and a significantly higher mutation frequency of *KDM6A*, RAS, *RHOB* and *UNC80* in NIBC (Figure S6). Notable

absences from the list of mutated genes in stage Ta tumors apart from *TP53*, are *CDKN1A, RB1, ERCC2, ERBB3* and *FBXW7,* which are mutated in >10% of MIBC (Cancer Genome Atlas Research Network, 2014a).

Structural and Functional Analysis of *RHOB* Mutations

We found mutations in the Ras superfamily GTPase member *RHOB* at significantly higher frequency (13.4% overall; 15% of GS1; 11% of GS2; Figure 3B) than reported in MIBC (5%) (Cancer Genome Atlas Research Network, 2014a). To explore the mutation spectrum further and to determine whether mutation is more common in NMIBC overall, we screened 36 stage T1 tumors by direct sequencing. This identified 6 additional mutations (overall frequency, 12.7%). Whilst missense variants were in several "hotspots" (Figure 5A), there was also a large deletion that includes the "insert region", an alpha-helical structure that in RHOA is essential for Rho kinase activation and cell transformation (Zong et al., 2001). A subset of missense mutations is located in regions of the protein implicated in interaction with PKNs 1-3 (Hutchinson et al., 2013) (Figure 5B and C). Overall, the mutations mapped to two separate regions of the RHOB protein, which broadly correspond to the two contact sites for PKN1 HR1 domains observed in the structure of the RHOA-PKN1 HR1a complex (Maesaki et al., 1999). Glu32 is in the switch 1 region and although it does not directly contact PKN1 it is likely to alter the interaction of Glu29. The finding of a large deletion, together with reports that RHOB has a tumor suppressor function (Huang and Prendergast, 2006), suggested that these may be inactivating mutations. We transiently expressed HA-tagged WT and mutant forms (E32K, E47K and E172K) with PKNs 1, 2 or 3 in HEK293 cells to assess interaction. Co-immunoprecipitation confirmed interaction with WT RHOB and some interaction with all mutants tested (data not shown). However low-level expression of mutant proteins in repeated experiments suggested that these forms may be unstable. This was confirmed by cycloheximide chase assays where the half-life of WT RHOB was measured as 50-60 min and E32K, E47K and E172K showed half-lives of ≤ 20 min (Figure 5D).

Mutations in Chromatin Modifier Genes

Of 633 confirmed chromatin modification (CM) genes in the DAnCER database (http://wodaklab.org/dancer/) (Turinsky et al., 2011), 84 genes showed non-silent mutations in samples analysed by exome sequencing (Table S6). Of these, 25 had inactivating mutations. Overall, 59 of 154 (38%) nonsynonymous mutations in these genes were inactivating (Table S3). All exome-sequenced tumors showed mutation in ≥ 2 CM genes (range 2-16; median 6) and 21 (87%) had inactivating mutations in ≥ 1 CM gene (range 1-5; median 2).

Of 17 CM genes analysed by target capture (Table S4), 11 showed ≥ 5% non-silent mutations (Figure 3B). Ninety-one percent of tumors had one or more mutations in these genes and 68% had 2 or more. Most frequently mutated CM genes were *KDM6A, STAG2, KMT2D, EP300, ARID1A, KMT2C* and *CREBBP* (Figures 6A, 6B and S4B). The H3K4 methyltransferases encoded by *KMT2C/D* and H3K27 demethylase encoded by *KDM6A* are part of a COMPASS-like complex associated with creating permissive chromatin environments at enhancers (Hu et al., 2013). Inactivating mutations in these genes suggest that altered chromatin remodelling and aberrant gene silencing accompanies transformation in these tumors. At least one of these genes was mutated in 70% of NIBC.

Association of *KDM6A* Mutation with Female Gender

Bladder cancer has a male:female incidence ratio of approximately 3:1 (Fajkovic et al., 2011) that is not fully understood (Reviewed in (Dobruch et al., 2015)). Differences in smoking habits are insufficient to account for this and suggested hormonal effects remain unclear. Molecular differences between tumors in males and females have not been explored extensively. Previously, we reported that *STAG2* (Xq25) mutation was associated with female gender in a study of tumors of all grades and stages (Taylor et al., 2013), but this was not confirmed here (p=0.3). An association of male gender with mutation in the RNA splicing gene *RBM10* (Xp11.23) has been reported in lung adenocarcinoma (Cancer

Genome Atlas Research Network, 2014b), but we found no gender-related difference (11% in females; 9% in males). However, we identified a striking association between *KDM6A* (Xp11.2) mutation and female gender (p=0.0092). Such a relationship has not been reported in other cancer types and interrogation of data from MIBC (Cancer Genome Atlas Research Network, 2014a) did not reveal a similar association. Mutations found in *KDM6A* are shown in Figure 6A.

Overall, 20 of 27 samples from female patients (74%) contained at least one *KDM6A* mutation. This gene is not X-inactivated (Greenfield et al., 1998) and, compatible with its predicted tumor suppressor function, 7 female samples had two mutations in the gene. In all but one of these, the mutant allele frequency was in the range 35-45%, suggesting that these may represent mutations in each of the two alleles. In one case the allele frequencies of a missense mutation (R549K) and an essential splice site mutation were ≥ 80%, and examination of sequencing read depth indicated likely deletion of one copy, suggesting that these may be on the same allele (Figure 6C). Thirteen female samples contained a single mutation and examination of read counts in these samples revealed only one additional sample with predicted deletion of the WT allele, suggesting that inactivation of a single allele in many tumors from females may be sufficient for functional effect.

In contrast, only 42% of male tumors (23/55) contained a *KDM6A* mutation. The male paralogue *UTY* (Yq11), which has 83% amino-acid identity with KDM6A (Greenfield et al., 1998) and has also been shown to have some demethylase activity (Walport et al., 2014), may be functionally equivalent to the second *KDM6A* copy in females. Indeed, we found mutation of *UTY* in 5 of 55 male samples, four of which also had a *KDM6A* mutation. Deletion of *UTY* in male samples with *KDM6A* mutation has been found in other cancers (van Haaften et al., 2009) and loss of Y is a common event in bladder cancer (Minner et al., 2010). We examined sequencing read depth in all male samples and identified 14 with predicted deletion (25%), 12 of which also had *KDM6A* mutation (Figure 6C), giving a total of 47% male samples with inactivation of one or both genes.

As KDM6A participates in COMPASS-like complexes with the chromatin methyl transferases KMT2C and KMT2D (Ford and Dingwall, 2015), which were mutated in 15% and 30% of these samples respectively, we examined the relationships of mutations in these genes, *KDM6A* and *UTY* and their combinations in relation to gender. In females, the majority of mutations in these genes occurred in conjunction with *KDM6A* mutation, with only 8% of samples showing independent mutations (Figure 6D). In contrast, 22% of male samples carried *KMT2C* and/or *KMT2D* mutations in the absence of *KDM6A* and/or *UTY* mutation or deletion.

DISCUSSION

NIBC are the most prevalent bladder tumors with urgent needs for improved prognostic biomarkers and approaches to localised therapy. We aimed to obtain more comprehensive genomic information and seek features that might allow these needs to be addressed. All samples initially selected for analysis were stage Ta primary tumors with predominantly papillary architecture, 95% classified as grade 2 (WHO, 1973), 61% as low and 39% as high grade according to the 2004 classification (Eble et al., 2004). No grade 3 tumors were intentionally included, but three were re-graded as grade 3 upon evaluation by a single pathologist. As such, this series represents the major population of stage Ta tumors at diagnosis. All expression profiling studies of bladder cancer identify a subtype with features similar to luminal breast cancer that expresses GATA3, FOXA1, and PPARγ, and all tumors here expressed these genes, confirming them as luminal.

We define two genomic subtypes, one (GS2) with reduced recurrence-free survival. Many GS1 tumors had no detectable copy number alterations, whereas most GS2 tumors showed loss of 9q. Although 9q loss is suggested as an early event in bladder cancer development, this was not a common solitary event, but rather the most frequent alteration in a general pattern of chromosomal instability, within which we did not identify other common features. One confirmed 9q deletion target is *TSC1*, but this does not account for

the loss of the entire chromosome arm in most cases. However, we found no evidence for other mutational targets, suggesting that haploinsufficiency for 9q genes may be important.

Expression profiling has uncovered clinically important bladder tumor subtypes. The tumors analysed here showed strong alignment to the Urobasal A subgroup (Sjodahl et al., 2012) and we suggest that addition of genomic information may provide prognostic information for this large tumor group. Consensus clustering and nearest neighbour analysis showed alignment of GS2 to the UROMOL class 2. Although this UROMOL class showed highest risk for progression, this was related to a large number of stage T1 tumors in class 2 (Hedegaard et al., 2016). Thus we suggest that alignment of GS2 Ta tumors to this class is unlikely to imply higher progression risk. Only one GS2 sample progressed to muscle invasion and one GS1 progressed to stage T1.

We identified expression features that are predicted to contribute to the observed difference in outcome and suggest potential avenues for therapy. Key hallmarks of GS2 were upregulated DNA repair and mTORC1 signaling and downstream effects of mTORC1 signaling, likely due to deletion or mutation of *TSC1.* mTORC1 stimulates anabolic processes including mRNA translation, aerobic glycolysis and *de novo* lipid and nucleotide synthesis (Saxton and Sabatini, 2017). Deregulated protein synthesis resulting from constitutive mTORC1 activity leads to endoplasmic reticulum (ER) stress that triggers the unfolded protein response, as seen in cells from Tsc1/2 deficient mice (Ozcan et al., 2008) and observed here in GS2. mTORC1 is also implicated in upregulation of genes involved in fatty acid synthesis (e.g. *SCD, FASN*), and cholesterol biosynthesis and homeostasis in GS2. Upregulated genes are involved in both cellular uptake of LDL-cholesterol (*LDLR*) and *de novo* cholesterol biosynthesis (*ACAT2, HMGCS, HMGCR, IDI1, FDPS, SQLE, LSS, CYP51A1*). FGFR3 activation also stimulates lipogenesis (Du et al., 2012), but as the majority of both GS1 and GS2 tumors contained activating *FGFR3* mutations, we conclude that the increase in GS2 results from mTORC1 activation. GS2 also showed a hypoxia-related signature, attributed to the effect of mTORC1 on translation of hypoxia-inducible

factor 1α (HIF1α). Glycolysis supplies both energy and precursor molecules including pyruvate, which can be converted to acetyl-coA, the precursor for cholesterol synthesis.

This reprogramming generates a state of metabolic vulnerability. Tsc1/2 deficient mouse cells are vulnerable to ER stress-induced apoptosis due to inability to adapt to additional ER challenge (Di Nardo et al., 2009; Ozcan et al., 2008) and are sensitive to ER stress inducers such as thapsigargin and tunicamycin (Ozcan et al., 2008). Such cells also succumb to additional oxidative stress (Li et al., 2016; Medvetz et al., 2015) and as mTORC1 inhibits autophagy, they are also vulnerable to further autophagy inhibition (Parkhitko et al., 2014). Drugs that target cholesterol synthesis or transport (Kuzu et al., 2016) may also have activity. Whilst some of these drugs may be toxic systemically, use as intravesical agents may be possible in NMIBC and some may have application in combination with rapalogues in systemic treatment of TSC1-deficient MIBC.

Despite similar patient age and smoking status, the overall mutation load in NIBC was approximately three-fold lower than reported in MIBC where an average of 7.7 (median 5.5) mutations per Mb was reported (Cancer Genome Atlas Research Network, 2014a). As APOBEC mutagenesis plays a major role in both groups (Acar et al., 2015; Hedegaard et al., 2016; Lamy et al., 2016; Cancer Genome Atlas Research Network, 2014a; Nordentoft et al., 2014), other factors may contribute to mutational burden in MIBC. Enrichment of APOBEC mutational signatures in NMIBC has been associated with high-risk tumors (Hedegaard et al., 2016; Lamy et al., 2016) and, compatible with this, we found enrichment in GS2, associated with higher APOBEC3H (A3H) expression. A3H is polymorphic, with seven haplotypes (A3H1-7), and in APOBEC3B-null breast cancers the A3H-1 haplotype contributes to the bulk of APOBEC signature mutations (Starrett et al., 2016). Here, we found no evidence for an association between A3H haplotype and APOBEC mutagenesis (data not shown). It is likely that the upregulated expression of DNA repair genes identified here contributes to the mutational burden and genomic instability of GS2 tumors. We also hypothesise that mutations in DNA repair genes play a larger role than we detected. Mutations in *ERCC2* were reported recently in 11 of 32 high-grade Ta samples (Pietzak et

al., 2017). We found one missense mutation in *ERCC2*, one in *ATM*, one in *FANCD2* and two in *FANCM* in exome-sequenced samples. These genes were not included in our custom target capture but should be included in future analyses.

NIBC are characterised by frequent activating mutations in *FGFR3* and *PIK3CA*. Most screens for *PIK3CA* mutations examine only hotspot codons. Our finding of many non-hotspot mutations indicates that the frequency of mutation in bladder cancer is higher than reported to date and that future mutation screening should examine the entire gene. Interestingly, many tumors in both genomic subtypes had *PIK3CA* activating mutations, indicating that this is not directly related to the activated mTORC1 profile in GS2, and as reported previously (Platt et al., 2009), *TSC1* and *PIK3CA* mutations were not mutually exclusive.

Although a limitation of this study is its power to detect infrequently mutated genes, it is clear that these tumors form a more homogeneous group than MIBC, where only *TP53* is mutated in more than 30% of cases and there is a long "tail" of infrequent mutations (Cancer Genome Atlas Research Network, 2014a). Nevertheless, we found differences between the subtypes, including more mutations of *HRAS* and *STAG2* in GS1 and of *TSC1* in GS2. Mutations in chromatin regulators were present at higher frequencies than in MIBC. All tumors had mutations in 2 and often many CM genes, with *KDM6A* and *STAG2* most frequently mutated, both at higher frequency than in MIBC.

The mutation frequency of the small GTPase *RHOB* was also higher than in MIBC (13% vs 5%) (Cancer Genome Atlas Research Network, 2014a). RHOB levels are decreased in many cancers including bladder (Volanis et al., 2011) and ectopic expression can suppress the transformed phenotype (Huang and Prendergast, 2006). The mutant forms of RHOB were less stable, suggesting a tumor suppressor function of WT RHOB in the urothelium. GTP binding drives a conformational change in RHOB that permits interaction with effectors such as the PKN family of serine threonine kinases (Hutchinson et al., 2013) and our structural analysis predicts that the mutant RHOB proteins described might exhibit

reduced interaction with PKN. Thus, in addition to reducing RHOB stability, RHOB mutants may be further impaired in signaling through reduced interaction with their effectors.

Although bladder cancer incidence is 3-4 times higher in males than in females, women with MIBC have worse outcomes than males (Reviewed in (Marks et al., 2016)). There is little data for NIBC, though a preponderance of grade 1 in females has been reported (Thorstenson et al., 2016). Overall, we found a trend towards GS1 membership in female samples in this series (p=0.049).

We speculate that some of the proposed influences on incidence might lead to gender-related molecular differences and consequent differences in tumor phenotype and clinical behaviour. Our finding of more *KDM6A* mutations in female NIBC is the most significant gender-related molecular difference reported to date. We sought evidence for a similar gender relationship in published studies. None is apparent in MIBC (Gui et al., 2011; Guo et al., 2013; Cancer Genome Atlas Research Network, 2014a). Relatively few NIBC have been screened for mutations, but in a series of exome-sequenced recurrent NMIBC there appears to be a skew towards female patients (Lamy et al., 2016) and a similar skew is apparent in a study of tumors of all grades and stages (Ahn et al., 2016). However, a recent study that included 51 stage Ta samples from patients in Singapore and Taiwan showed no gender bias (Ler et al., 2017). The reason is unclear but could relate to geographical or grade differences in the patient populations or the different methods used. A major limitation of all studies is the small number of female samples analysed and this is more acute if there is indeed a relationship to tumor grade and stage. Examination of a large series of tumors of all grades and stages using a standardised mutation screening approach is required to confirm this gender difference and its relationship to tumor histopathology.

Studies of other tumor types report a bias towards mutation of *KDM6A* in males (Van der Meulen et al., 2015; van Haaften et al., 2009), compatible with the presence of a single allele in males and its escape from X-inactivation in females (Greenfield et al., 1998). Thus, our finding is counterintuitive. However, biallelic mutation was not detected in the majority of females, suggesting that *KDM6A* may be haploinsufficient in the female urothelium. The

protein encoded by *UTY* in males has some demethylase activity but not full functional equivalence to KDM6A (Walport et al., 2014). This is compatible with our finding of *UTY* mutation or deletion in several male samples, though the overall frequency of involvement of either of these genes (47%) was lower than that of *KDM6A* mutation in females.

KDM6A is implicated in transcriptional regulation at promoters and enhancers (Van der Meulen et al., 2014) and loss of its demethylase activity on H3K27 and H3K4 is predicted to lead to transcriptional repression. There is also evidence for enzyme-independent functions of KDM6A. Thus the consequences of inactivation may be multifaceted and likely to be urothelial cell context dependent. Whether there are gender-related epigenetic differences in the normal urothelium is unknown, but if KDM6A has a gender-related role in these cells, this could influence the early tumor pathogenesis pathway.

In summary, these data provide an improved view of the genomic landscape of NIBC that points to chromatin modification as a key oncogenic driver for future therapeutic evaluation and to metabolic vulnerability as a therapeutic target in tumors with loss of 9q. Our results also highlight the potential need for gender stratification in an era of precision oncology.

AUTHOR CONTRIBUTIONS

C.D.H. and M.A.K. conceived the study and designed experiments. S.J. and J-A.R. were involved in procurement and histopathological review of samples. F.M.P., C.F.T. and C.D.H. prepared samples and carried out sequencing and mutation analysis. O.A., A.D., L.S., L.J.K., C.F.T., D.A.G, L.J.K., G.J.B. and M.A.K. analysed sequence data. A.D., L.S., O.A. and H.T. performed bioinformatics and statistical analyses. J.E.B. and C.D.H. carried out functional assays. H.L. assessed significance of *BRCA2* mutations. A.J.R., H.R.M. and A.J.C. provided reagents, advice and interpretation of RHOB analyses. H.R.M. evaluated *RHOB* mutations in relation to protein structural data.

M.A.K., C.D.H., O.A. and D.A.G. wrote the manuscript with the assistance and final approval of all authors.

ACKNOWLEDGEMENTS

We thank Joanne Brown for tissue collection and processing, Laura Kettlewell for tissue culture, Sally Harrison and Oxford Gene Technology for technical support for next-generation sequencing and Ben Knowles for help with artwork. We acknowledge the following funding sources: Yorkshire Cancer Research grant L376PA (M.A.K.), Intramural Research Program of the National Institutes of Health (NIH), National Institute of Environmental Health Sciences, project Z1AES103266 (D.A.G.), Wellcome Trust and University of Leeds Personal Fellowship RGCALA101195 (L.F.S) and Cancer Research UK grant C6620/A15961 (A.J.R.).

REFERENCES

Acar, O., Ozkurt, E., Demir, G., Sarac, H., Alkan, C., Esen, T., Somel, M., and Lack, N. A. (2015). Determining the origin of synchronous multifocal bladder cancer by exome sequencing. BMC Cancer *15*, 871.

Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics / editorial board, Jonathan L Haines [et al] *Chapter 7*, Unit7 20.

Ahn, J., Kim, K. H., Park, S., Ahn, Y. H., Kim, H. Y., Yoon, H., Lee, J. H., Bang, D., and Lee, D. H. (2016). Target sequencing and CRISPR/Cas editing reveal simultaneous loss of UTX and UTY in urothelial bladder cancer. Oncotarget *7*, 63252-63260.

Babjuk, M., Bohle, A., Burger, M., Capoun, O., Cohen, D., Comperat, E. M., Hernandez, V., Kaasinen, E., Palou, J., Roupret, M.*, et al.* (2017). EAU Guidelines on Non-Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. Eur Urol *71*, 447-461.

Balbas-Martinez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Marquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayes, M.*, et al.* (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet *45*, 1464-1469.

Burns, M. B., Temiz, N. A., and Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet *45*, 977-983.

Cancer Genome Atlas Research Network. (2014a). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507,* 315-322.

Cancer Genome Atlas Research Network. (2014b). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543-550.

Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., Kim, J., Kwiatkowski, D. J., Fargo, D. C., Mieczkowski, P. A.*, et al.* (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet *47*, 1067-1072.

Choi, W., Porten, S., Kim, S. S., Willis, D., Plimack, E. R., Hoffman-Censits, J., Roth, B., Cheng, T., Tran, M., Lee, I.-L.*, et al.* (2014). Identification of Distinct Basal and Luminal

Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy. Cancer Cell *25*, 152-165.

Damrauer, J. S., Hoadley, K. A., Chism, D. D., Fan, C., Tiganelli, C. J., Wobker, S. E., Yeh, J. J., Milowsky, M. I., Iyer, G., Parker, J. S., and Kim, W. Y. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. Proc Natl Acad Sci USA *111*, 3110-3115.

Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., Mooney, T. B., Callaway, M. B., Dooling, D., Mardis, E. R.*, et al.* (2012). MuSiC: identifying mutational significance in cancer genomes. Genome Res *22*, 1589-1598.

Di Nardo, A., Kramvis, I., Cho, N., Sadowski, A., Meikle, L., Kwiatkowski, D. J., and Sahin, M. (2009). Tuberous sclerosis complex activity is required to control neuronal stress responses in an mTOR-dependent manner. J Neurosci *29*, 5926-5937.

Dobruch, J., Daneshmand, S., Fisch, M., Lotan, Y., Noon, A. P., Resnick, M. J., Shariat, S. F., Zlotta, A. R., and Boorjian, S. A. (2015). Gender and Bladder Cancer: A Collaborative Review of Etiology, Biology, and Outcomes. Eur Urol *69*, 300-310.

Du, X., Wang, Q. R., Chan, E., Merchant, M., Liu, J., French, D., Ashkenazi, A., and Qing, J. (2012). FGFR3 stimulates stearoyl CoA desaturase 1 activity to promote bladder tumor growth. Cancer Res *72*, 5843-5855.

Eble, J. N., Sauter, G., Epstein, J. I., and Sesterhenn, I. A. (2004). World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs.: IARC Press Lyon).

Fajkovic, H., Halpern, J. A., Cha, E. K., Bahadori, A., Chromecki, T. F., Karakiewicz, P. I., Breinl, E., Merseburger, A. S., and Shariat, S. F. (2011). Impact of gender on bladder cancer incidence, staging, and prognosis. World J Urol *29*, 457-463.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer *136*, E359-386.

Ford, D. J., and Dingwall, A. K. (2015). The cancer COMPASS: navigating the functions of MLL complexes in cancer. Cancer Genet *208*, 178-191.

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M. R. (2004). A census of human cancer genes. Nat Rev Cancer *4*, 177-183.

Gartner, J. J., Davis, S., Wei, X., Lin, J. C., Trivedi, N. S., Teer, J. K., Program, N. C. S., Meltzer, P. S., Rosenberg, S. A., and Samuels, Y. (2012). Comparative exome sequencing of metastatic lesions provides insights into the mutational progression of melanoma. BMC Genomics *13*, 505.

Greenfield, A., Carrel, L., Pennisi, D., Philippe, C., Quaderi, N., Siggers, P., Steiner, K., Tam, P. P., Monaco, A. P., Willard, H. F., and Koopman, P. (1998). The UTX gene escapes X inactivation in mice and humans. Hum Mol Genet *7*, 737-742.

Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li, X., Zhou, L.*, et al.* (2011). Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet *43*, 875-878.

Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q.*, et al.* (2013). Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat Genet *45*, 1459-1463.

Gymnopoulos, M., Elsliger, M.-A., and Vogt, P. K. (2007). Rare cancer-specific mutations in PIK3CA show gain of function. Proc Natl Acad Sci USA *104*, 5569-5574.

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.

Hu, D., Gao, X., Morgan, M. A., Herz, H. M., Smith, E. R., and Shilatifard, A. (2013). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol Cell Biol *33*, 4745-4754.

Huang, M., and Prendergast, G. C. (2006). RhoB in cancer suppression. Histol Histopathol *21*, 213-218.

Hurst, C. D., Platt, F. M., Taylor, C. F., and Knowles, M. A. (2012). Novel tumor subgroups of urothelial carcinoma of the bladder defined by integrated genomic analysis. Clin Cancer Res *18*, 5865-5877.

Hutchinson, C. L., Lowe, P. N., McLaughlin, S. H., Mott, H. R., and Owen, D. (2013). Differential binding of RhoA, RhoB, and RhoC to protein kinase C-related kinase (PRK) isoforms PRK1, PRK2, and PRK3: PRKs have the highest affinity for RhoB. Biochemistry (Mosc) *52*, 7999-8011.

James, A. C., and Gore, J. L. (2013). The costs of non-muscle invasive bladder cancer. Urol Clin North Am *40*, 261-269.

Jebar, A. H., Hurst, C. D., Tomlinson, D. C., Johnston, C., Taylor, C. F., and Knowles, M. A. (2005). FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. Oncogene *24*, 5218-5225.

Knowles, M. A., and Hurst, C. D. (2015). Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer *15*, 25-41.

Kompier, L. C., Lurkin, I., van der Aa, M. N. M., van Rhijn, B. W. G., van der Kwast, T. H., and Zwarthoff, E. C. (2010). FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. PloS One *5*, e13821.

Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc *4*, 1073-1081.

Kuzu, O. F., Noory, M. A., and Robertson, G. P. (2016). The Role of Cholesterol in Cancer. Cancer Res *76*, 2063-2070.

Lamy, P., Nordentoft, I., Birkenkamp-Demtroder, K., Borg Houlberg Thomsen, M., Villesen, P., Vang, S., Hedegaard, J., Borre, M., Bjerggaard Jensen, J., Hoyer, S.*, et al.* (2016). Paired exome analysis reveals clonal evolution and potential therapeutic targets in urothelial carcinoma. Cancer Res *76*, 5894-5906.

Ler, L. D., Ghosh, S., Chai, X., Thike, A. A., Heng, H. L., Siew, E. Y., Dey, S., Koh, L. K., Lim, J. Q., Lim, W. K.*, et al.* (2017). Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. Sci Transl Med *9*, eaai8312.

Li, J., Shin, S., Sun, Y., Yoon, S. O., Li, C., Zhang, E., Yu, J., Zhang, J., and Blenis, J. (2016). mTORC1-Driven Tumor Cells Are Highly Sensitive to Therapeutic Targeting by Antagonists of Oxidative Stress. Cancer Res *76*, 4816-4827.

Maesaki, R., Ihara, K., Shimizu, T., Kuroda, S., Kaibuchi, K., and Hakoshima, T. (1999). The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1. Mol Cell *4*, 793-803.

Marks, P., Soave, A., Shariat, S. F., Fajkovic, H., Fisch, M., and Rink, M. (2016). Female with bladder cancer: what and why is there a difference? Transl Androl Urol *5*, 668-682.

Medvetz, D., Sun, Y., Li, C., Khabibullin, D., Balan, M., Parkhitko, A., Priolo, C., Asara, J. M., Pal, S., Yu, J., and Henske, E. P. (2015). High-throughput drug screen identifies chelerythrine as a selective inducer of death in a TSC2-null setting. Mol Cancer Res *13*, 50-62.

Meeks, H. D., Song, H., Michailidou, K., Bolla, M. K., Dennis, J., Wang, Q., Barrowdale, D., Frost, D., Embrace, McGuffog, L.*, et al.* (2016). BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. J Natl Cancer Inst *108*, djv315.

Minner, S., Kilgue, A., Stahl, P., Weikert, S., Rink, M., Dahlem, R., Fisch, M., Hoppner, W., Wagner, W., Bokemeyer, C.*, et al.* (2010). Y chromosome loss is a frequent early event in urothelial bladder cancer. Pathology (Phila) *42*, 356-359.

Nellist, M., van Slegtenhorst, M. A., Goedbloed, M., van den Ouweland, A. M., Halley, D. J., and van der Sluijs, P. (1999). Characterization of the cytosolic tuberin-hamartin complex. Tuberin is a cytosolic chaperone for hamartin. J Biol Chem *274*, 35647-35652.

Ng, P. C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res *31*, 3812-3814.

Nordentoft, I., Lamy, P., Birkenkamp-Demtroder, K., Shumansky, K., Vang, S., Hornshoj, H., Juul, M., Villesen, P., Hedegaard, J., Roth, A.*, et al.* (2014). Mutational context and diverse clonal development in early and late bladder cancer. Cell Rep *7*, 1649-1663.

Ozcan, U., Ozcan, L., Yilmaz, E., Duvel, K., Sahin, M., Manning, B. D., and Hotamisligil, G. S. (2008). Loss of the tuberous sclerosis complex tumor suppressors triggers the unfolded protein response to regulate insulin signaling and apoptosis. Mol Cell *29*, 541-551.

Parkhitko, A. A., Priolo, C., Coloff, J. L., Yun, J., Wu, J. J., Mizumura, K., Xu, W., Malinowska, I. A., Yu, J., Kwiatkowski, D. J.*, et al.* (2014). Autophagy-dependent metabolic reprogramming sensitizes TSC2-deficient cells to the antimetabolite 6-aminonicotinamide. Mol Cancer Res *12*, 48-57.

Pietzak, E. J., Bagrodia, A., Cha, E. K., Drill, E. N., Iyer, G., Isharwal, S., Ostrovnaya, I., Baez, P., Li, Q., Berger, M. F.*, et al.* (2017). Next-generation Sequencing of Nonmuscle Invasive Bladder Cancer Reveals Potential Biomarkers and Rational Therapeutic Targets. Eur Urol.

Platt, F. M., Hurst, C. D., Taylor, C. F., Gregory, W. M., Harnden, P., and Knowles, M. A. (2009). Spectrum of phosphatidylinositol 3-kinase pathway gene alterations in bladder cancer. Clin Cancer Res *15*, 6008-6017.

Ploeg, M., Aben, K. K., and Kiemeney, L. A. (2009). The present and future burden of urinary bladder cancer in the world. World J Urol *27*, 289-293.

Plon, S. E., Eccles, D. M., Easton, D., Foulkes, W. D., Genuardi, M., Greenblatt, M. S., Hogervorst, F. B., Hoogerbrugge, N., Spurdle, A. B., Tavtigian, S. V., and Group, I. U. G. V. W. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat *29*, 1282-1291.

Rampias, T., Vgenopoulou, P., Avgeris, M., Polyzos, A., Stravodimos, K., Valavanis, C., Scorilas, A., and Klinakis, A. (2014). A new tumor suppressor role for the Notch pathway in bladder cancer. Nat Med *20*, 1199-1205.

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G.*, et al.* (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet *45*, 970-976.

Saxton, R. A., and Sabatini, D. M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell *168*, 960-976.

Shirts, B. H., Casadei, S., Jacobson, A. L., Lee, M. K., Gulsuner, S., Bennett, R. L., Miller, M., Hall, S. A., Hampel, H., Hisama, F. M.*, et al.* (2016). Improving performance of multigene panels for genomic analysis of cancer predisposition. Genet Med *18*, 974-981.

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.

Starrett, G. J., Luengas, E. M., McCann, J. L., Ebrahimi, D., Temiz, N. A., Love, R. P., Feng, Y., Adolph, M. B., Chelico, L., Law, E. K.*, et al.* (2016). The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nat Commun *7*, 1-13.

Svatek, R. S., Hollenbeck, B. K., Holmang, S., Lee, R., Kim, S. P., Stenzl, A., and Lotan, Y. (2014). The economics of bladder cancer: costs and considerations of caring for this disease. Eur Urol *66*, 253-262.

Sylvester, R. J., van der Meijden, A. P., Oosterlinck, W., Witjes, J. A., Bouffioux, C., Denis, L., Newling, D. W., and Kurth, K. (2006). Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol *49*, 466-477.

Taylor, C., Platt, F., Hurst, C., Thygensen, H., and Knowles, M. (2013). Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumor grade and stage and inversely related to chromosomal copy number changes. Hum Mol Genet *23*, 1964-1974.

Taylor, C. F. (2009). Mutation scanning using high-resolution melting. Biochem Soc Trans *37*, 433-437.

Thorstenson, A., Hagberg, O., Ljungberg, B., Liedberg, F., Jancke, G., Holmang, S., Malmstrom, P. U., Hosseini, A., and Jahnson, S. (2016). Gender-related differences in urothelial carcinoma of the bladder: a population-based study from the Swedish National Registry of Urinary Bladder Cancer. Scand J Urol *50*, 292-297.

Turinsky, A. L., Turner, B., Borja, R. C., Gleeson, J. A., Heath, M., Pu, S., Switzer, T., Dong, D., Gong, Y., On, T.*, et al.* (2011). DAnCER: disease-annotated chromatin epigenetics resource. Nucleic Acids Res *39*, D889-894.

Van der Meulen, J., Sanghvi, V., Mavrakis, K., Durinck, K., Fang, F., Matthijssens, F., Rondou, P., Rosen, M., Pieters, T., Vandenberghe, P.*, et al.* (2015). The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. Blood *125*, 13-21.

Van der Meulen, J., Speleman, F., and Van Vlierberghe, P. (2014). The H3K27me3 demethylase UTX in normal development and disease. Epigenetics *9*, 658-668.

van Haaften, G., Dalgliesh, G. L., Davies, H., Chen, L., Bignell, G., Greenman, C., Edkins, S., Hardy, C., O'Meara, S., Teague, J.*, et al.* (2009). Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet *41*, 521-523.

van Oers, J. M., Lurkin, I., van Exsel, A. J., Nijsen, Y., van Rhijn, B. W., van der Aa, M. N., and Zwarthoff, E. C. (2005). A simple and fast method for the simultaneous detection of nine fibroblast growth factor receptor 3 mutations in bladder cancer and voided urine. Clin Cancer Res *11*, 7743-7748.

Volanis, D., Zaravinos, A., Kadiyska, T., Delakas, D., Zoumpourlis, V., and Spandidos, D. A. (2011). Expression profile of Rho kinases in urinary bladder cancer. Journal of BUON : official journal of the Balkan Union of Oncology *16*, 511-521.

Walport, L. J., Hopkinson, R. J., Vollmar, M., Madden, S. K., Gileadi, C., Oppermann, U., Schofield, C. J., and Johansson, C. (2014). Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. J Biol Chem *289*, 18302-18313.

WHO (1973). Histological typing of urinary bladder tumours. International Histological Classification of Tumours *10*.

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572-1573.

Zong, H., Kaibuchi, K., and Quilliam, L. A. (2001). The insert region of RhoA is essential for Rho kinase activation and cellular transformation. Mol Cell Biol *21*, 5287-5298.

FIGURE LEGENDS

Figure 1. Copy Number Subtypes of NIBC and Their Relationship to Recurrence-Free Survival

A. Unsupervised hierarchical cluster analysis of copy number data from 140 Ta tumors and LUCC8 cell line. Columns represent samples and rows genomic position. Blue, copy number gain; yellow, copy number loss. Chromosome number is shown on the left-hand side. Two main clusters, GS1 and GS2, are indicated.

B. Kaplan-Meier plot of recurrence-free survival (RFS) in patients stratified according to genomic subtype (GS1 n=78; GS2 n=61), 9q loss (no loss n=82; loss n=57) and 2004 grade (low n=86; high n=53). A log-rank test reveals statistically significant differences.

See also Table S1.

Figure 2. Analysis of mRNA Expression in GS1 and GS2

A. Heatmap representation of gene z-scores detected as differentially expressed (limma; p value and FDR <0.01) between GS1 (n=48) and GS2 (n=31) using expression microarray analysis. Genes located on chromosome 9 are indicated by sidebar.

B. Gene Set Enrichment Analysis results depicting differential expression of genes associated with mTORC1 signaling (nominal p=0.006) between GS1 and GS2.

C. Heatmap representations of gene z-scores associated with Gene Ontology Biological Processes that differ significantly between GS1 and GS2 samples.

D. mRNA expression levels (log_2) of selected cholesterol (SQLE) and lipid biosynthesis (FASN) genes that were significantly differentially expressed (p<0.01) between GS1 (n=48) and GS2 (n=31) samples. Each box has lines at the lower quartile, median quartile and upper quartile values. Whiskers represent the minimum and maximum values.

See also Table S2 and Figures S1 and S2.

Figure 3. APOBEC Signature and Mutations in NIBC

A. Analysis of an APOBEC mutation signature in exome data from NIBC. Left, distribution of the fold of APOBEC mutation pattern enrichment. Enrichment and q values (one-sided Fisher's exact test) were calculated as described previously (Roberts et al., 2013) and in STAR Methods. Pie chart shows the distribution of the 24 samples analysed in each category. Right, distribution of mutation types in each tumor sample. "APOBEC to G" and "APOBEC to T" denote APOBEC induced C to G and C to T substitutions, respectively, in TCW motifs, where W represents A or T flanking nucleotides. "Non-APOBEC C:G" denotes SNVs in non-APOBEC mutated C:G pairs. "A:T" denotes SNVs in A:T pairs.

B. Oncoplot showing distribution of mutations in ≥ 5% NIBC. Data is derived from exome sequence of 24 samples and selective target capture followed by next generation sequencing of 58 additional samples. Chromatin modification genes are shown in bold. See also Tables S3-S6 and Figures S3-S6.

Figure 4. Comparison of Mutations in GS1 and GS2

A. Boxplot representation of mutations per megabase values based on exome sequence analysis of 17 GS1 and 7 GS2 samples plotted according to subtype membership. Mutations per megabase values are calculated from number of somatic mutations detected by VarScan2 divided by size of genome covered by a minimum of 10 reads in whole exome sequencing data. The p values were calculated using Mann-Whitney test. Each box has lines at the lower quartile, median quartile and upper quartile values. Whiskers represent the minimum and maximum values.

B. Boxplot representation of APOBEC mutation pattern detection in whole exome sequence data (GS1 n=17; GS2 n=7). Fold enrichment is calculated as the number of APOBEC signature mutations (TCW to TTW or TGW and the complementary WGA to WAA or WCA) divided by the frequency of cytosines in the sequence TCW. The p values were calculated using Mann-Whitney test. Each box has lines at the lower quartile, median quartile and upper quartile values. Whiskers represent the minimum and maximum values.

C. Dotplot representation of APOBEC3H mRNA expression (log_2) (TC22000311.hg) detected by microarray analysis of 48 GS1 and 31 GS2 samples. The p values were calculated using Mann-Whitney test. Bars represent median, maximum and minimum values.

D-G. Data from combined exome and targeted sequencing of 82 samples was analyzed to generate an oncoplot showing distribution of mutations in GS1 and GS2 subtypes in *FGFR3*, *PIK3CA* and *HRAS* (D), in *TSC1* (E), in histone modifying enzymes (F) and in chromatin associated proteins (G).

See also Table S3.

Figure 5. Structural and Functional Analysis of *RHOB* Mutations

A. Schematic of RHOB protein showing mutations found in this study (82 Ta and 36 T1 high-grade tumors) and in a previous study of MIBC (Cancer Genome Atlas Research Network, 2014a). G1-G5; G boxes involved in guanine nucleotide binding; switch I and switch II, regions that undergo nucleotide-dependent conformational changes.

B. The structure of RHOB in cartoon form with the positions of the mutations identified. Mutated sidechains are shown in a space-filling representation, with carbons in green, oxygens red and nitrogens blue. The insert region is indicated. Figure generated using Pymol.

C. The structure of a model of RHOB in complex with PKN1. The HR1 domain in the contact 1 site is shown in cyan and that in contact 2 is shown in purple.

D. HEK293 cells were transfected with plasmids encoding WT or mutant (E32K, E47K, E172K) RHOB. 24 hr post-transfection, cycloheximide (CHX) (100 μg/ml) was added to inhibit protein synthesis. Left panel, whole-cell protein lysates at indicated intervals and RHOB expression examined by western blot analysis. Right panel, RHOB expression was normalised to α-tubulin and the amount of RHOB protein remaining at each time point plotted as a percentage of the level at time point 0 min.

See also Table S3.

Figure 6. Mutation of Chromatin Modifier Genes and Relationship to Gender

A. Schematic of KDM6A protein showing the distribution and types of mutation identified in 82 stage Ta bladder tumors. Lower bar is the cDNA sequence with splice site mutations. Domain abbreviations: TPR, tetratricopeptide repeat; JmjC, JmjC domain.

B. Schematic of histone 3 modification and chromatin remodelling processes dysregulated by genomic alterations in 82 NIBC. Increased H3K27me3 (through inactivating mutation in *KDM6A*) or decreased H3K4me3 (through inactivation of KMT2A/2D/2C) is predicted to lead to increased chromatin compaction and repression of transcription. These genes are found in the same complex and are suggested to have concerted action in regulating transcription. 76% of tumors in the study showed mutation in at least one of these genes. The transcriptional regulator MECOM (EVI1) is implicated in interaction with EP300/CREBBP. K, lysine; Me, methylation; Ac, acetylation. Percentages indicate overall frequency of mutation in 82 NIBC.

C. Histograms showing mutant allele frequencies and copy number loss of *KDM6A* and *UTY* in male and female tumor samples.

D. Histograms showing distribution of mutations in *KDM6A, UTY, KMT2C* and *KMT2D* and their combinations in all 82 tumors and separately in tumors from male (n=55) and female (n=27) patients. Numbers represent percentages in each group.

See also Table S3.

STAR★METHODS

KEY RESOURCES TABLE

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author Margaret A. Knowles (m.a.knowles@leeds.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

LUCC8 Cell Line

The human bladder cancer cell line LUCC8 was established from a stage Ta grade 2 (TaG2) tumor (1838) in this laboratory and maintained in complete Keratinocyte Growth Medium 2 (Promocell C-20111) supplemented with 0.09 mM [final] $CaCl_2$, 1% FCS and 3 mM glycine on BD Primaria plastic (BD Biosciences) under standard culture conditions of 37°C and 5% CO_2. Its identity was verified by short tandem repeat DNA typing using the Powerplex 16 System, which confirmed identity to the original tumor and lack of identity with any other cell lines maintained in the laboratory.

Human Tissues and Subject Follow-up Data
Clinical samples and associated clinical data were sourced from the Leeds Multidisciplinary Research Tissue Bank (Leeds East Research Ethics Committee reference: 10/H1306/7). All patients gave written informed consent. Cold cup biopsies were collected, snap-frozen and stored in liquid nitrogen. The remainder of the sample was embedded in paraffin for diagnostic assessment. Samples were graded and staged by a consultant urological pathologist (JR) using the 1973 and 2004 WHO and TNM criteria, respectively (Eble et al., 2004; WHO, 1973). All available clinical information was collected (Table S1). Median follow-up time was 69 months (range 1-200 months). Copy number analysis was carried out on 140 grade 1 or grade 2 stage Ta tumors and the LUCC8 (1838) cell line. Twenty-three of these tumors and the LUCC8 cell line were used for whole-exome sequencing, 58 tumors were used for targeted re-sequencing and seventy-nine tumors were gene expression profiled. Details of the platforms each sample was analysed on are given in Table S1. Thirty six stage T1 high grade tumors were included in analysis of *RHOB* mutation only. No other clinical information was collected for these samples.

HEK293 cells
HEK293 cells were maintained in DMEM (high glucose) containing 10% fetal calf serum under standard culture conditions of 37°C and 5% CO_2.

METHOD DETAILS

DNA Extraction
Genomic DNA was isolated from frozen tissue sections comprising at least 70% tumor cells using a QIAamp DNA Mini Kit or a Gentra PureGene Tissue Kit. DNA was extracted from Venous blood samples using a Nucleon BACC DNA Extraction Kit or by salt precipitation.

Copy Number Analysis
Low pass whole genome sequencing or array-CGH was used to assess copy number alterations in 141 grade 1 or grade 2 stage Ta primary tumors and the LUCC8 cell line. Next-generation sequencing libraries were constructed using NEBNext® reagents according to the manufacturer's instructions. Raw sequencing data was processed as described in the "Read mapping and genotype calling" section. After BAM file generation, ngCGH was used to compare number of read counts between tumor and matched blood samples using a window size of 1000 reads (Gartner et al., 2012). GC correction and copy number calling using the FASST2 Segmentation Algorithm, a Hidden Markov Model (HMM) based approach, were carried out within the Nexus Copy Number software package. The significance threshold for segmentation was set at 1.0E-5 also requiring a minimum of 3 probes per segment and a maximum probe spacing of 1000 between adjacent probes before breaking a segment. The log ratio thresholds for single copy gain and single copy loss were set at 0.2 and -0.2, respectively. The log ratio thresholds for two or more copy gain and homozygous loss were set at 1 and -1 respectively. Array-CGH and copy number calling was carried out as previously described (Hurst et al., 2012). To conduct cluster analysis, each individual region associated with a BAC array clone was assigned a copy number class (0 = no copy number alteration; 1 = gain; 2 = high-level amplification; -1 = loss; -2 = high-level loss). One-way unsupervised hierarchical cluster analysis was conducted using Euclidean distance and the Ward method of linkage.

RNA Extraction and Gene Expression Profiling
Seventy-nine tumors were expression profiled (Table S1). Total RNA was isolated from frozen tissue sections using a RNeasy Plus Micro Kit and amplified using the Affymetrix GeneChip® WT PLUS Reagent Kit according to manufacturer's instructions. The resulting cDNA was quantified using OD (NanoDrop). The cDNA was normalised and hybridised onto Affymetrix Human Transcriptome 2.0 microarrays for 16 hr at 45°C. Microarrays were washed and stained using the Affymetrix GeneChip® Hybridization, Wash, and Stain Kit according to manufacturer's instructions using the Affymetrix GeneChip® Fluidics Station 450. Microarrays were scanned using an Affymetrix GeneChip® 7G microarray scanner. Quality control checks, gene level normalization (using SST-RMA) and signal summarisation was conducted using Affymetrix® Expression Console™ Software. R2 was used for routine data visualization. The UROMOL classifier was implemented by selection and median-centering of each of the 117 genes, then 1000 iterations of ConsensusClusterPlus (CCP) (Wilkerson and Hayes, 2010) v1.36.0 were run using a Pearson distance metric and a random gene fraction of 0.8 (Hedegaard et al., 2016). For comparison to previously published LUND subtypes we used 1-Nearest Neighbour Classification test from the R package "class" (v 7.3-14); 956 genes were shared between array technologies. Data from all 308 tumors (GSE32894) was used as a training set against data described here. GSEA v2.2.1 was carried out using all genes; genomic subtypes were assigned as phenotypes and permuted 1000 times, the test dataset was collapsed to gene symbols and run against gene sets in the Hallmarks database.

Immunohistochemistry
For immunohistochemistry, fixed paraffin embedded tissues were sectioned at a thickness of 5 μM and mounted onto slides. Slides were heated for 20 min at 70°C then deparaffinised in 100% xylene (three times for 3 min), rehydrated in 100% ethanol (four times for 1 min), and finally washed once in water to remove excess ethanol. Antigen retrieval was carried out using the Menarini antigen retrieval unit (Menarini Diagnostics). Slides were heated in citrate buffer (pH6) to 125°C for 2 min, then briefly washed in 1x tris-buffered saline (TBS) automation wash buffer (Menarini Diagnostics). Slides were then rinsed under running water before washing in TBS for 30 s and Menarini wash buffer (Menarini Diagnostics) for 30 s. Endogenous peroxidase and alkaline-phosphatase activity was blocked with 2 drops of Bloxall blocking solution (Vector Laboratories) for 10 min. Non-specific antigen binding was blocked by incubation in a 10% casein solution in antibody diluent (Invitrogen) for 20 min. Sections were stained for CD45, GATA3, FOXA1, KI67, CCNB1, PPARG, STAG2 and UPKIII. All runs included a no primary antibody control. Primary antibodies were diluted in antibody diluent (Invitrogen) and applied to slides for 1 hr at RT. Slides were washed twice in 1x TBS automation wash buffer (Menarini Diagnostics). Post-primary HRP conjugated antibodies were added using a Novocastra post-primary blocking kit (Leica Biosystems). Slides were incubated with 2 drops of post primary reagent (Leica Biosystems) for 30 min, then washed twice in 1x TBS automation wash buffer (Menarini Diagnostics) for 5 min before 2 drops of polymer reagent (Leica Biosystems) was added. Slides were washed twice in TBS for 5 min. Detection was performed using DAB peroxidase substrate (Vector Laboratories). Two hundred microliters of ImmPACT™DAB substrate was applied to each slide for 5 min. Each slide was transferred to a water reservoir for 5 min before counterstaining in haematoxylin for 30 s. Slides were then washed in running water for 1 min, Schotts water for 1 min, 100% ethanol once for 15 s, 100% ethanol once for 2 min, and 100% ethanol twice for 3 min, and immersed in 100% xylene (three times for 3 min). Coverslips were mounted on the slides with DEPEX (Fisher Scientific UK Ltd) and slides were visualised using an Olympus Bx50 microscope (Olympus Optical Co.).

Exome Sequencing
Twenty-three tumors and the LUCC8 cell line and their matched blood samples were analysed by whole-exome sequencing. Libraries were generated using 3 μg of DNA and

enriched for exonic regions using the SureSelect Human All Exon 50Mb kit (20 tumor/blood pairs) or the SureSelect Human All Exon V4 Kit (4 tumor/blood pairs) according to the manufacturer's protocols. Sequencing was performed by Oxford Gene Technology (Oxford, UK) on an Illumina HiSeq 2000 in paired-end mode with 100bp read length and eight samples (four tumor/blood pairs) per lane. Base calling and quality control was performed using Illumina's Real Time Analysis software version 1.6 with standard settings. Sequence files were QC checked using FastQC (v0.10.0) before preprocessing. Adapter contamination and low-quality read ends (< 20) were trimmed using cutadapt 1.3. Any read in which either pair had a length less than 19 was removed from subsequent analysis.

Read Mapping and Genotype Calling
Alignment was performed using BWA 0.7 to hs37D5, the GRch37/hg19-based reference used by the 1000 Genomes Project. Duplicate reads were removed using the Picard v1.56 MarkDuplicates program. Local realignment around indels was performed using the GATK v1.3 RealignerTargetCreator and IndelRealigner in Smith-Waterman mode with reference to dbSNP v132. The Picard v1.56 FixMateInformation program was used to ensure that all mate-pair information was in sync between each read and its mate following local realignment. Base quality scores were recalibrated using GATK v1.3 CountCovariates and TableRecalibration with reference to dbSNP v132. BAM files were sorted and then indexed using SAMtools index. Sequencing metrics were generated using the Picard v1.56 CalculateHsMetrics programme with intervals files relevant to capture regions for either SureSelect Human All Exon 50 Mb or SureSelect Human All Exon V4.

Somatic Variant Analysis
Pileup files (created using SAMtools mpileup, with parameters – d 5000 and q- 20) were used as input to VarScan2 v2.3.5. This was used in somatic mode, with a strand filter, to identify somatic single nucleotide variations and small insertions and deletions (indels). Results were then processed using processSomatic to extract the somatic mutations with the specification that zero reads were to support the variant in the normal (matched blood) sample. High confidence somatic variant and indel calls were converted to mutation annotation format (MAF) using a bespoke script that accesses Ensembl version 75 via the application programming interface.

Validation of Mutation Calls
Primers were designed using BatchPrimer3 **(**http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi?). PCR was performed in a volume of 15 µl containing 1 x PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.2 µM each primer, 1U AmpliTaq Gold and 10 ng of genomic DNA. Cycle parameters were 95°C for 5 min then 35 cycles of 95°C for 45 s, 55°C for 45 s and 72°C for 45 s followed by 72°C for 10 min. 5 µl of PCR products were mixed with 2 µl of illustra™ ExoProStar and incubated at 37°C for 15 min, then 80°C for 15 min. Sequencing was performed using 1 µl of ExoProStar-treated PCR product and a Big Dye Terminator Ready Reaction Mix v1.1 kit. A total of 193 somatic single nucleotide variants identified by whole-exome sequencing were examined using this approach. Of these 178 variants (92%) were confirmed, 1 variant was not detected and 14 variants could not be confirmed due to technical issues with the PCR step.

Mutation Significance Analysis
Significantly mutated genes were identified using the Genome MuSiC package. The genome music bmr calc-covg script was run to count the bases with sufficient coverage in the regions of interest (ROIs) of each gene for each pair of tumor-normal BAM files and to categorise these base changes into AT, CG (non-CpG) and CpG counts. The script also generates total base-counts across all ROIs of each gene for each sample. The MAF file generated during the somatic variant calling step and the per-gene coverage data generated by the Genome MuSiC bmr calc-covg script were used with the Genome MuSiC bmr calc-bmr script to

calculate the overall Background Mutation Rate (BMR) and BMRs of different mutation categories (AT/CG/CpG transitions, AT/CG/CpG transversions, and indels). The output file generated by the Genome MuSiC bmr calc-bmr script contains per-gene mutation rates and this was used with the Genome MuSiC smg tool to test for significantly mutated genes. Using the per-gene mutation rates categorized by mutation type and the overall background mutation rate for each category, Genome MuSiC smg calculates p values and false discovery rates (FDR) for each gene using three tests (Fisher's Combined p value Test (FCPT), Likelihood Ratio Test (LRT), and the Convolution test (CT)). If the FDR of a gene was less than the specified maximum FDR for at least two of the tests, then the gene was considered significantly mutated.

Analysis of APOBEC Mutagenesis Signature

The exome-wide prevalence of the APOBEC mutagenesis signature and the enrichment of this signature over its presence expected for random mutagenesis was evaluated as described in (Roberts et al., 2013) and in (Chan et al., 2015). Analysis results are summarised in Table S3. Briefly, analysis is based on previous findings that APOBECs deaminate cytidines predominantly in a tCw motif and that the APOBEC mutagenesis signature is composed of approximately equal numbers of two kinds of changes in this motif – tCw→tTw and tCw→tGw mutations (complements included; flanking nucleotides shown in small letters; w=A or T). We calculated on a per sample basis, the enrichment of the APOBEC mutation signature among all mutated cytosines in comparison to the fraction of cytosines that occur in the tCw motif among the +/- 20 nucleotides surrounding each mutated cytosine ("APOBEC enrichment" column). In addition, several other parameters that characterize the prevalence of the APOBEC mutagenesis pattern in a sample and/or that are useful for downstream analyses and comparisons were calculated. The most useful parameter for studying correlations is the minimum estimate of the number of APOBEC induced mutations in a sample - "APOBEC Mutation Load (minimum estimate)" column. It was calculated using the formula: ["APOBEC, number of mutations"]x[("APOBEC enrichment"-1)/["APOBEC enrichment"], which allows estimating the number of APOBEC signature mutations in excess of what would be expected by random mutagenesis. Calculated values are rounded to the nearest whole number. "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. The "APOBEC Mutation Load (minimum estimate)" value shows high correlation (0.9-0.95) with all other parameters used to characterize APOBEC mutagenesis pattern, such as APOBEC enrichment, and absolute and relative APOBEC mutation loads.

Targeted Re-Sequencing

Agilent's SureDesign tool was used to design a 296 kb SureSelect custom capture for all coding exons of the 49 selected candidate genes (Table S4). The design included a 10 base pair extension to the 3' and 5' ends of each region. Libraries were generated for a validation sample set consisting of 58 tumors (Table S1) using 1.2 µg of DNA and enriched for targeted exonic regions according to the manufacturer's protocols. The samples were run in a single lane on an Illumina HiSeq 2000 in paired-end mode with 100 bp read length. Raw data handling up to FASTQ file generation was as described above. Paired-end files in FASTQ format were converted to FASTA format in NextGENe v.2.3.4 (SoftGenetics LLC). FASTA reads were quality filtered using the following criteria: median score threshold ≥ 20, maximum number of uncalled bases ≤ 3, called base number of each read ≥ 25, and trim or reject read when ≥ 3 bases with score ≤ 30. Filtered reads were aligned to a GRch37/hg19-based human genome reference sequence provided by NextGENe and variants identified using mutation calling filters with standard settings (details available upon request). Further details relating to file conversion, alignment, and mutation filtering are available at http://www.softgenetics.com/PDF/NextGene_UsersManual_web.pdf. The NextGENe Viewer (Softgenetics) was used to filter variants for inclusion, based on target capture ROI only and

all synonymous base changes were removed. A mutation report containing details of all remaining variants (noncoding, missense, nonsense, indels, splice site) was exported for further filtering. Variants present in the 1000 Genomes database were removed. Variants present in the dbSNP database were removed unless they were assigned a status of 'probably-pathogenic' or 'pathogenic' or had a COSMIC ID. Intronic variants up to 8bp from splice junctions were assessed for predicted effects on splicing using the SPNN splice site prediction tool (http://www.fruitfly.org/seq_tools/splice.html). Only variants with predicted effects were retained. The somatic status of post-filtering variants was confirmed by PCR and Sanger sequencing of the corresponding matched blood DNA samples as described above.

Mutation Analysis of *FGFR3*, RAS and *TP53*

Whole genome amplification (WGA) was performed using the REPLI-g Mini Kit on 25 ng of DNA extracted from fresh-frozen tumor samples. WGA amplified DNA samples were used in single gene mutation detection assays. SNaPshot assays were used to screen for hotspot mutations in *FGFR*, *HRAS*, *KRAS* and *NRAS* as described previously (Kompier et al., 2010; van Oers et al., 2005). High-resolution melt (HRM) analysis (Taylor, 2009) was used to screen for mutations in exons 4-10 of *TP53*. Primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) (primer sequences available upon request). Two sets of primers were designed to cover exon 4 (4a and 4b). The predicted melting profile of the PCR products was determined using DHPLCMelt (http://insertion.stanford.edu/melt.html) and in some cases primers were redesigned or had short GC clamps added so that each amplicon's melting temperature was made as simple as possible. PCR for HRM was carried out in 10 µl reactions overlaid with 20 µl of mineral oil. Each reaction contained 1x Acqua MasterMix (Acquascience), 1x LCGreen+ (Acquacsience), 4 pmol of each primer, and 2 µl of WGA amplified DNA diluted 1/100 with TE. Cycle parameters were 95°C for 15 min then 45 cycles of 95°C for 10 s, annealing for 15 s (56-57°C depending on primer pair) and 72°C for 15 s; after a final denaturation at 94°C for 30 s, reactions were cooled to 25°C at 0.1°C/s to promote heteroduplex formation. HRM was carried out using a HR96 LightScanner (Acquascience). All samples with melting profiles that differed from WT samples were bi-directionally sequenced.

Prediction of Effect of Missense Variants

The web-based Variant Effect Predictor (VEP) tool available through Ensembl was used to obtain PolyPhen-2 (Adzhubei et al., 2013) and SIFT (Ng and Henikoff, 2003) pathogenicity predictions and scores for the protein products of all missense mutations in the discovery and prevalence sets.

Analysis of *BRCA2* Mutations

In silico analysis of BRCA2 variants of uncertain significance was performed with the assistance of Alamut Visual 2.7 software (Interactive Biosoftware), and variants were classified on a five class system as described by Plon *et al.* (Plon et al., 2008). Classification was based on: the frequency of the variant in normal population datasets (Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) [accessed June 2015]); evidence from the current literature if previously described; *in silico* analysis to predict the impact of missense changes on protein function based on sequence conservation [Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org), Align-Grantham Variation with Grantham Deviation (A-GVGD) (http://agvgd.iarc.fr), PolyPhen-2 (genetics.bwh.harvard.edu/pph2)]; and *in silico* prediction of any effect on splicing using splice tools integrated into Alamut Visual.

Gene Ontology Analysis

Gene ontology analysis was conducted using "Database for Annotation, Visualization and Integrated Discovery (DAVID)" v6.7 on a gene list featuring components mutated in two or

more patients (261) using whole genome as background. A p value cut off of $p < 0.001$ and a q value (Benjamini) cut off of $q<0.05$ were implemented.

Mutation Analysis of *RHOB*

RHOB was screened for mutations in 36 stage T1 grade 3 samples by direct sequencing. Primer sequences are available upon request.

Structural Models of RHOB

The RHOB-PKN1 model was generated using Modeller based on the coordinates (1cxz) of RHOA with two separate HR1a domains (available at PDBePISA).

RHOB Plasmids and Site-Directed Mutagenesis

Mutant RHOB E32K and E47K expression constructs were generated by site-directed mutagenesis of pcDNA3.1(+)-HA-RhoB WT, which contains 3 copies of HA-tag fused to the N-terminus of human RHOB. E172K sequence was amplified directly from tumor cDNA. Mutants were cloned into KpnI-XbaI-cut pcDNA3.1 by homologous recombination cloning (InFusion) according to the manufacturer's protocol. Primer sequences are available upon request. Non-dimerising EGFP (A206K; mEGFP) was amplified from pmEGFP-1 (Addgene) and cloned into pcDNA3.1 by InFusion method. WT and mutant RhoB sequences were cloned by InFusion method downstream of mEGFP to generate pcDNA-mEGFP-RhoB. All inserts and junctions were fully sequenced.

RHOB Functional Studies in HEK293 cells

For transfection, 2.5×10^6 cells were plated in 10 cm dishes and transfected the following day using *Trans*IT-293 with 6.7 µg DNA and 25 µl transfection agent per dish. Rates of turnover of WT and mutant forms of RhoB were investigated following cycloheximide (CHX) treatment. Twenty-three hr after transfection of HEK293 cells medium was replaced with medium containing 100 µg/ml CHX in DMSO. Cells were lysed after 0, 10, 20, 40 and 60 min in 200 µl lysis buffer (50 mM Tris-HCl pH7.5, 150 mM NaCl, 10% glycerol, 1% Tween-20, 0.1% NP-40) containing protease inhibitors and phosphatase inhibitors, debris was pelleted and lysates stored at -80°C.

Western Blot Analysis

Protein (5 µg) was separated under denaturing conditions in Any kD™ Mini-PROTEAN® TGX™ precast polyacrylamide gels (Bio-Rad) and transferred to Amersham Protran 0.45 µM pore size nitrocellulose membrane (GE Healthcare Life Sciences). Blots were incubated with primary antibody specific to RHOB in PBS containing 2% skimmed milk powder overnight at 4°C. Horseradish peroxidase-conjugated secondary antibodies and Luminata Forte Western HRP substrate (Millipore) were used for chemiluminescent detection of bound RHOB antibody. Blots were reprobed with alpha-tubulin antibody for assessment of protein loading.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was done in RStudio for Mac (Version 0.99.486). The statistical test LIMMA and a cut off of 0.01 for p value and FDR were applied to identify genes differentially expressed between samples belonging to GS1 and GS2. The association between overall mutation frequency and each of the two variables gender and smoking status (coded as binary: ever smoked yes/no) were assessed using t-tests. For each gene analysed in the full set of 82 samples, the association between mutation status and gender were assessed using Fisher's exact test. The significance of association between mutation status and recurrence free survival was assessed using log rank test and "survival" package version 2.39-5. Patients were censored at time of last follow up and events were classified as recurrence (>=3 months) post-surgery or death. All p values were adjusted for multiple testing using the Benjamini-Hochberg method.

DATA AND SOFTWARE AVAILABILITY

All microarray data have been deposited in GEO: GSE93527.

Supplemental Tables provided as Excel files

Table S1 related to Figure 1. Details of patients, tumor samples and analyses carried out.

Table S2 related to Figure 2. Genes differentially expressed between tumors belonging to subtypes GS1 and GS2.

Table S3 related to Figures 3-6. Variants and APOBEC enrichment values.

KEY RESOURCES TABLE

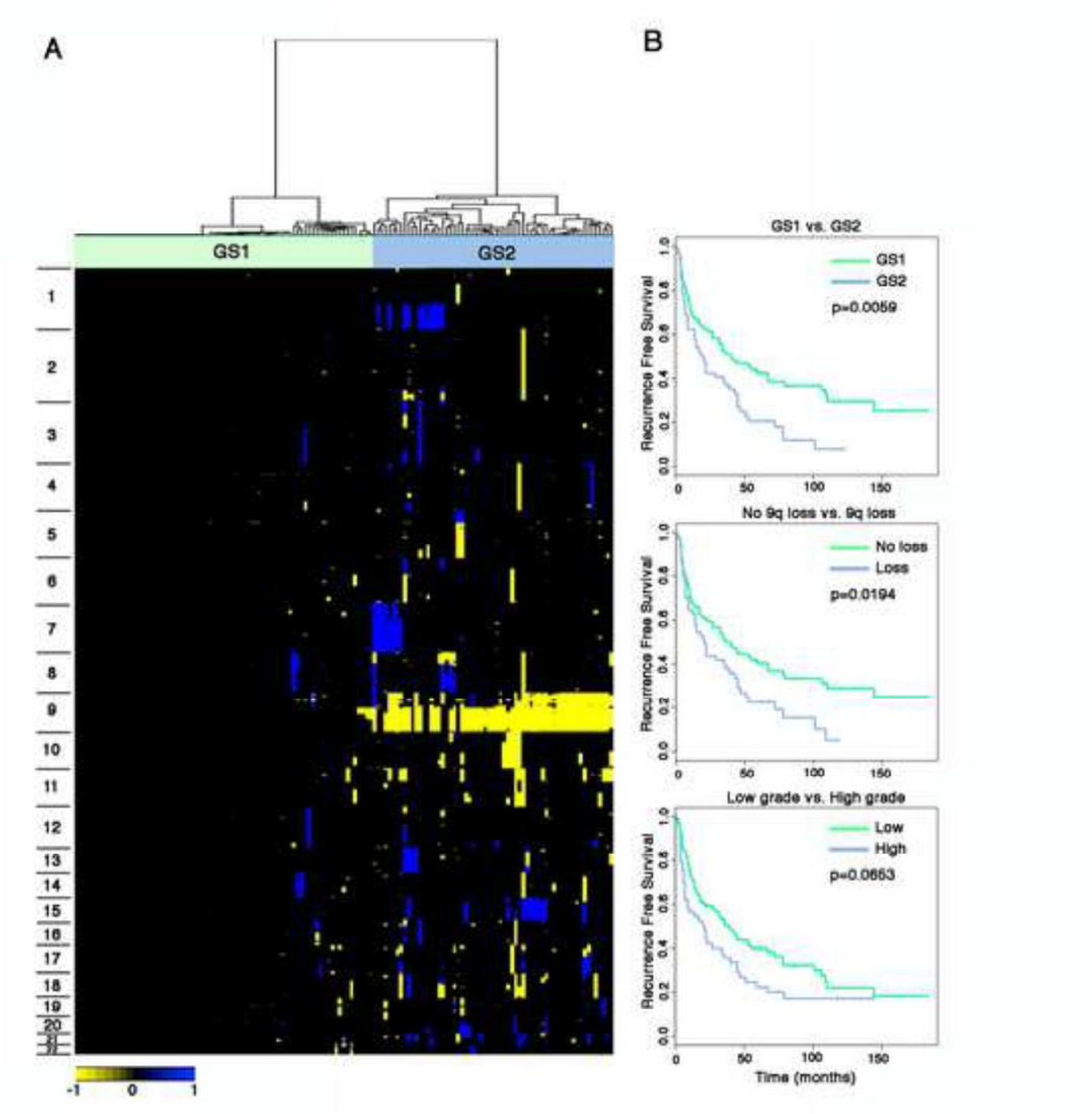
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-CD45	Dako	Cat# M0701, RRID: AB_2314143
Rabbit monoclonal anti-Cyclin B1	Abcam	Cat# ab32053, RRID: AB_731779
Mouse monoclonal anti-FOXA1	Abcam	Cat# ab40868, RRID: AB_732410
Rabbit monoclonal anti-GATA3	Cell Signaling Technology	Cat# 5852S, RRID: AB_10835690
Rabbit monoclonal anti-Ki67	Thermo Fisher Scientific	Cat# RM-9106-S0, RRID: AB_2341197
Rabbit monoclonal anti-PPARgamma	Cell Signaling Technology	Cat# 2435S, RRID: AB_2166051
Mouse-monoclonal anti-RHOB	Santa Cruz	Cat# sc-8048, RRID: AB_628219
Mouse monoclonal anti-STAG2	Santa Cruz	Cat# sc-81852, RRID: AB_2199948
Rat monoclonal anti-alpha-tubulin	AbD Serotec	Cat# MCA77G, RRID: AB_325003
Mouse monoclonal anti-Uroplaklin III	Abcam	Cat# ab78196, RRID: AB_1566851
Biological Samples		
Fresh-frozen primary bladder tumor tissues	Leeds Multidisciplinary Research Tissue Bank (University of Leeds)	http://multirtb.leeds.ac.uk
Formalin fixed paraffin embedded primary bladder tumor tissues	Leeds Multidisciplinary Research Tissue Bank (University of Leeds)	http://multirtb.leeds.ac.uk
Venous blood samples from bladder cancer patients	Leeds Multidisciplinary Research Tissue Bank (University of Leeds)	http://multirtb.leeds.ac.uk
Chemicals, Peptides, and Recombinant Proteins		
Cycloheximide	Sigma	Cat# C7698
Protease inhibitors	Sigma	Cat# P8340
Phosphatase inhibitors	Sigma	Cat# P5726
Critical Commercial Assays		
Affymetrix GeneChip Human Transcriptome Array 2.0 and Genechip WT PLUS Reagent Kit	Affymetrix	Cat# 902310
AmpliTaq Gold DNA Polymerase	Thermo Fisher Scientific	Cat# 4311814
Big Dye Terminator Ready Reaction Mix v1.1 Kit	Thermo Fisher Scientific	Cat# 4336774
GeneChip Hybridisation, Wash, and Stain Kit	Affymetrix	Cat# 900720
Gentra Puregene Tissue Kit	Qiagen	Cat# 158667
HiSeq 2000	Illumina	N/A
illustra Nucleon BACC3 Genomic DNA Extraction Kit	GE Healthcare	Cat# RPN8512
illustra ExoProStar	GE Healthcare	Cat# US77705
InFusion HD Cloning Kit	Clontech Laboratories	Cat# 639650
NEBNext DNA Library Prep Master Mix Set for Illumina	New England BioLabs	Cat# E6040L
NEBNext Singleplex Oligos for Illumina	New England BioLabs	Cat# E7350L
NEBNext High-Fidelity 2X PCR Master Mix	New England BioLabs	Cat# M0541L

PowerPlex 16 System	Promega	Cat# DC6531
QIAamp DNA Mini Kit	Qiagen	Cat# 51306
REPLI-g Mini Kit	Qiagen	Cat# 150025
RNeasy Plus Micro Kit	Qiagen	Cat# 74034
SureSelect^{XT} Human All Exon 50Mb Kit	Agilent Technologies, Inc	Cat# G3370C
SureSelect^{XT} Human All Exon v4 Kit	Agilent Technologies, Inc	Cat# 5190-4631
*Trans*IT-293	Mirus	Cat# MIR 2700
Deposited Data		
Raw and processed Affymetrix microarray data	This paper	GEO: GSE93527
A Molecular Taxonomy of Urothelial Carcinoma	Sjödahl et al., 2012	GEO: GSE32894
Experimental Models: Cell Lines		
Human bladder cancer cell line LUCC8	Derived in this laboratory	N/A
HEK293 cell line	Cancer Research UK, London, UK	N/A
Recombinant DNA		
pcDNA3.1(+)-HA-RhoB WT	cDNA Resource Center, Bloomsburg University	Clone ID: RHO0B0tN00
pmEGFP-1	Gift from Benjamin Glick	Addgene Cat# 36409
Software and Algorithms		
Affymetrix Expression Console Software	Affymetrix	http://www.Affymetrix.com/
BWA	Li and Durbin, 2009	https://sourceforge.net/projects/bio-bwa/
ConsensusClusterPlus (CCP)	Wilkerson and Hayes, 2010	http://bioconductor.org/packages/ConsensusClusterPlus/
Cytoscape	Shannon et al., 2003	www.cytoscape.org/
DAVID	Dennis et al., 2003	https://david.ncifcrf.gov/
EnrichmentMap	Merico et al., 2010	http://baderlab.org/Software/EnrichmentMap
GATK	Broad Institute	https://broadinstitute.org/gatk/
Genome MuSic Package	Dees et al., 2012	http://gmt.genome.wustl.edu/packages/genome-music/
GSEA	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/
Modeller	Eswar et al., 2006	https://salilab.org/modeller/
Nexus Copy Number	Biodiscovery, Inc	http://www.biodiscovery.com/nexus-copy-number/
ngCGH	Gartner et al., 2012	https://github.com/seandavi/ngCGH
PDBePISA	European Bioinformatics Institute	www.ebi.ac.uk/pdbe/pisa/
Picard	Broad Institute	https://broadinstitute.github.io/picard/
R2: Genomics Analysis and Visualization Platform	Jan Koster, Academic Medical Center, Netherlands	http://r2.amc.nl/

R-Studio for Mac	R-Studio	https://www.rstudio.com/products/rstudio/download/
SAMtools	Li et al., 2009	http://samtools.sourceforge.net
Variant Effect Predictor (VEP)	Ensembl	http://www.ensembl.org/Homo_sapiens/Tools/VEP
VarScan2	Koboldt et al., 2012	http://dkoboldt.github.io/varscan

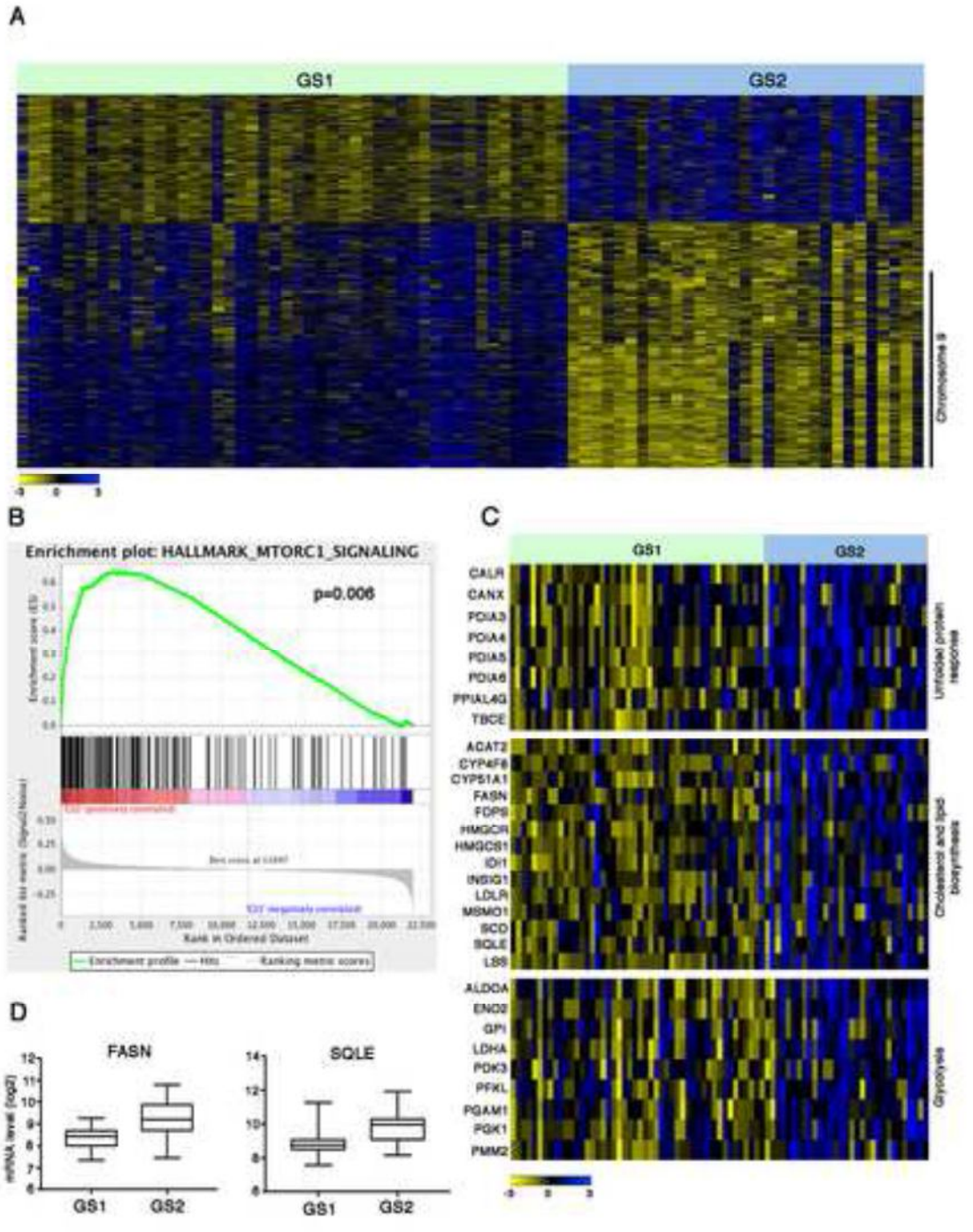
Figure

Click here to download Figure Figure 1 revised 25_7_17.tif



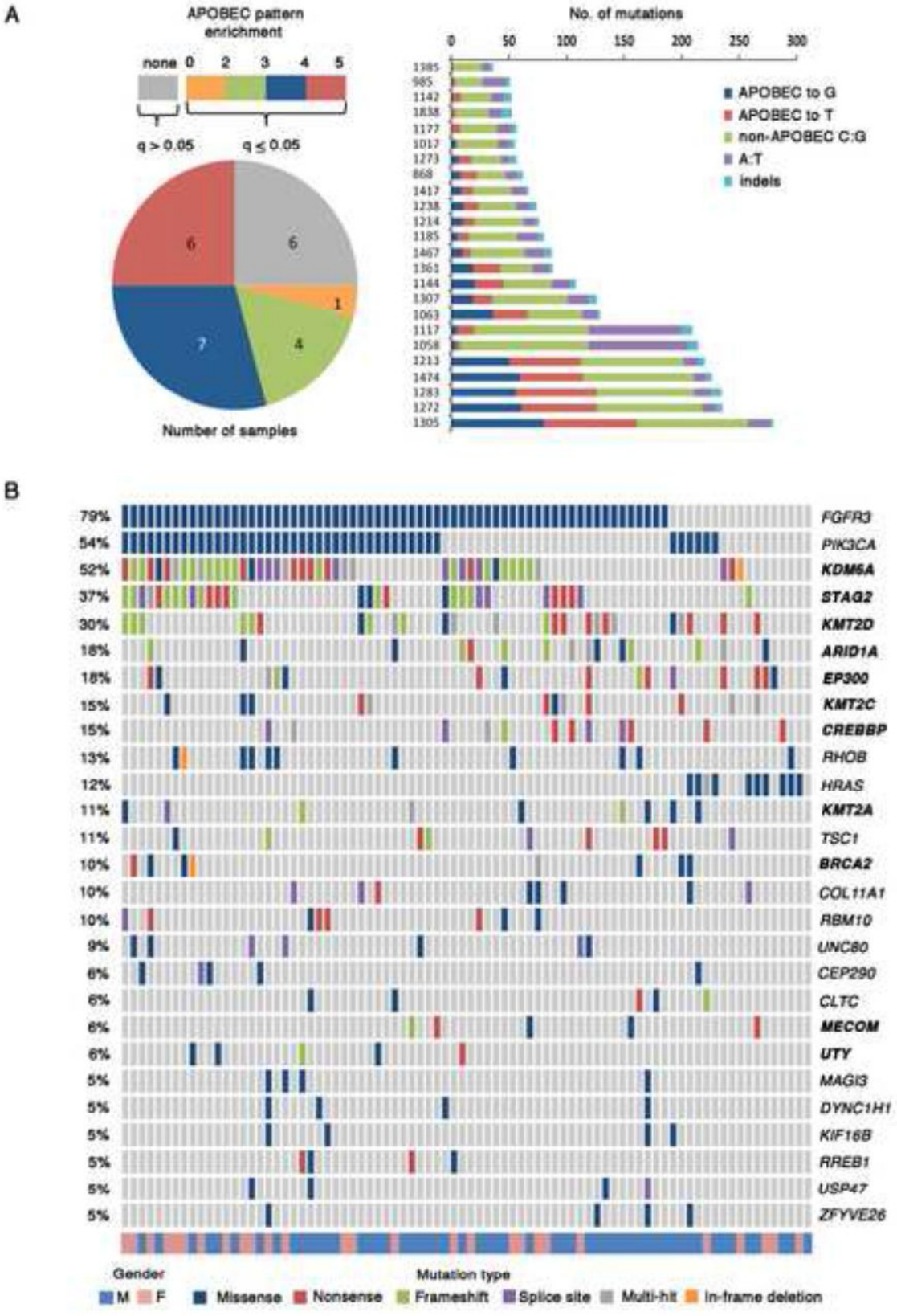
Figure

Click here to download Figure Figure 2 revised 25_7_17.tif



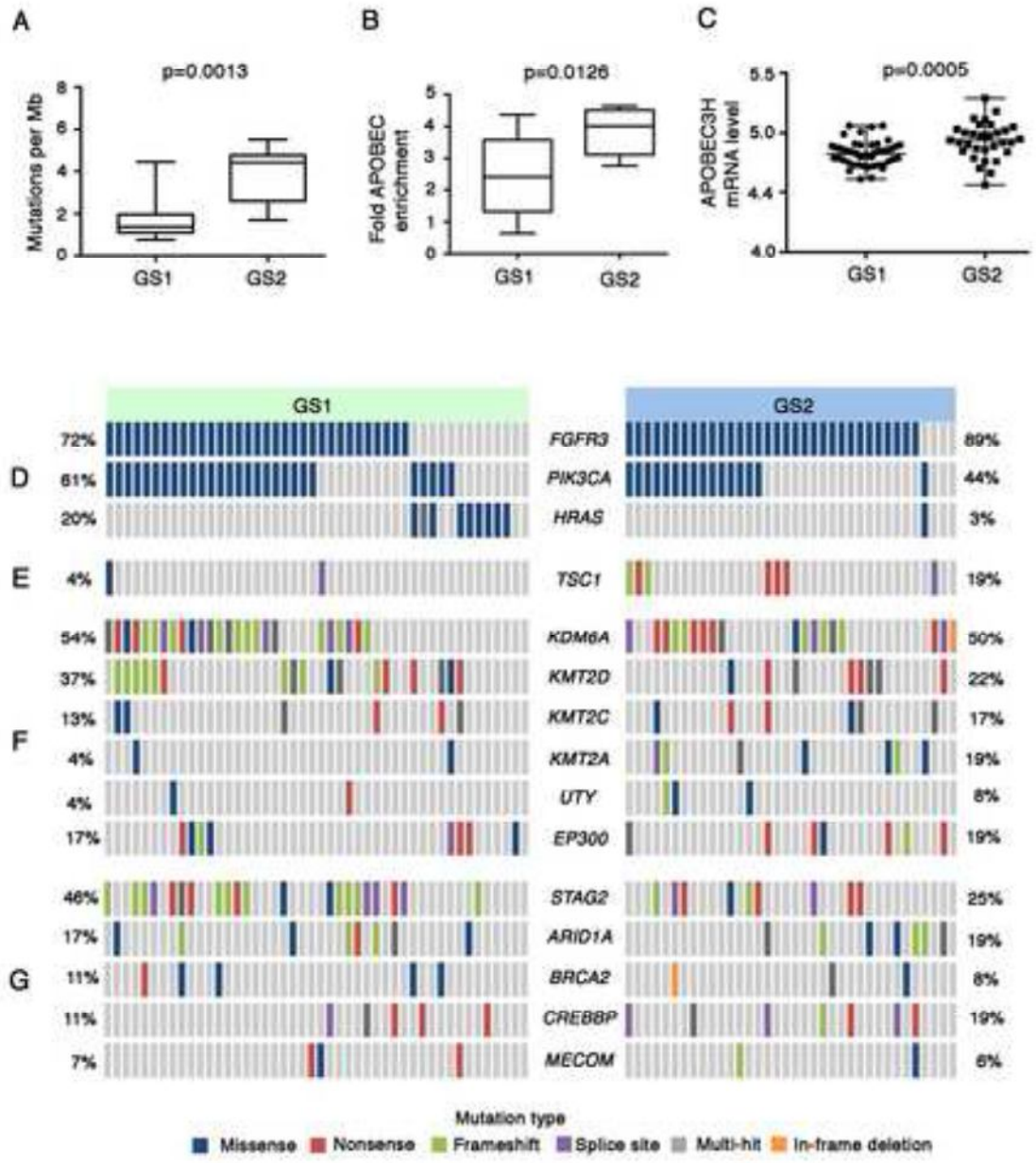
Figure

Click here to download Figure Figure 3 revised 25_7_17.tif



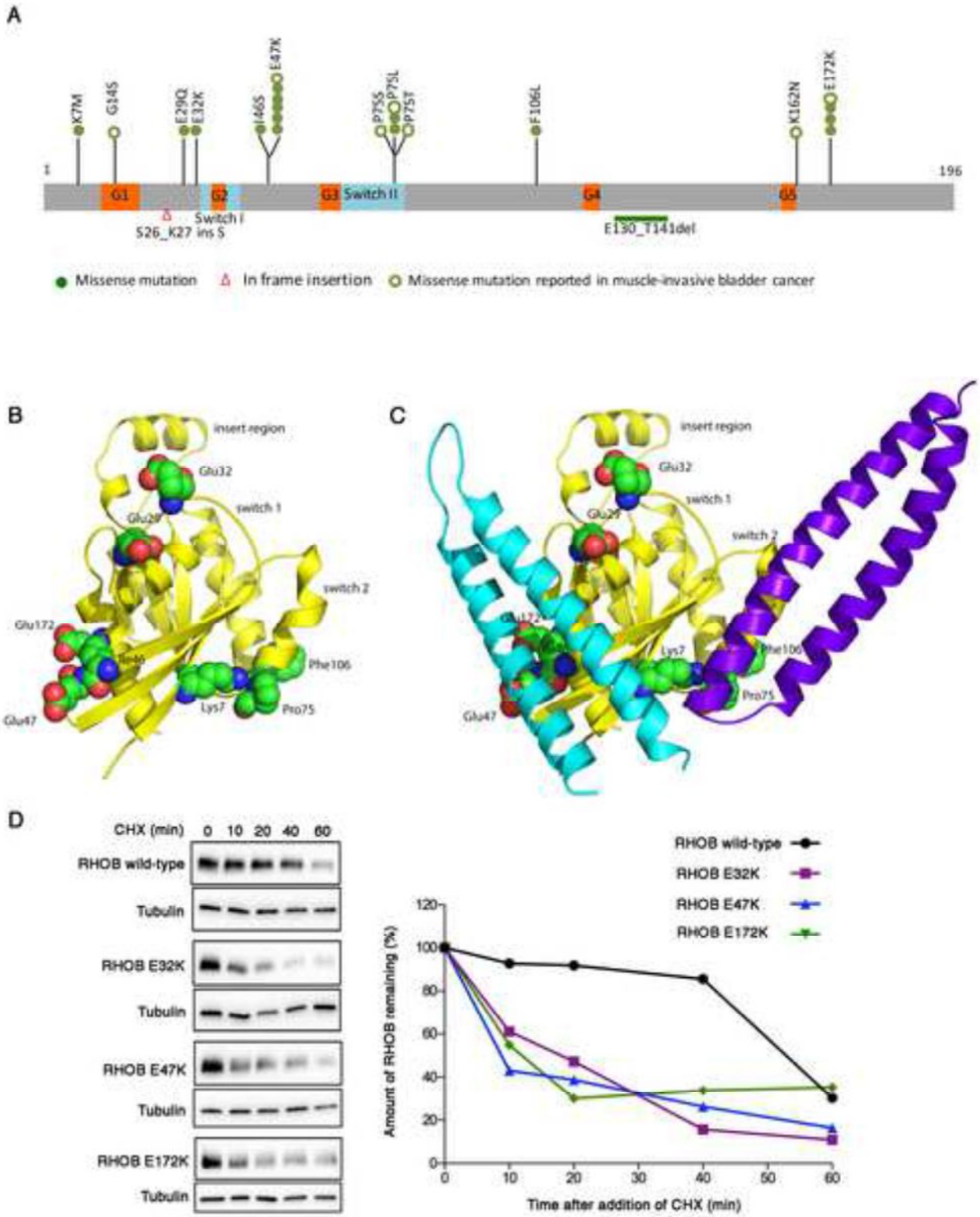
Figure

Click here to download Figure Figure 4 revised25_7_17.tif



Figure

Click here to download Figure Figure 5 revised 25_7_17.tif



Figure

Click here to download Figure Figure 6 revised 25_7_17.tif

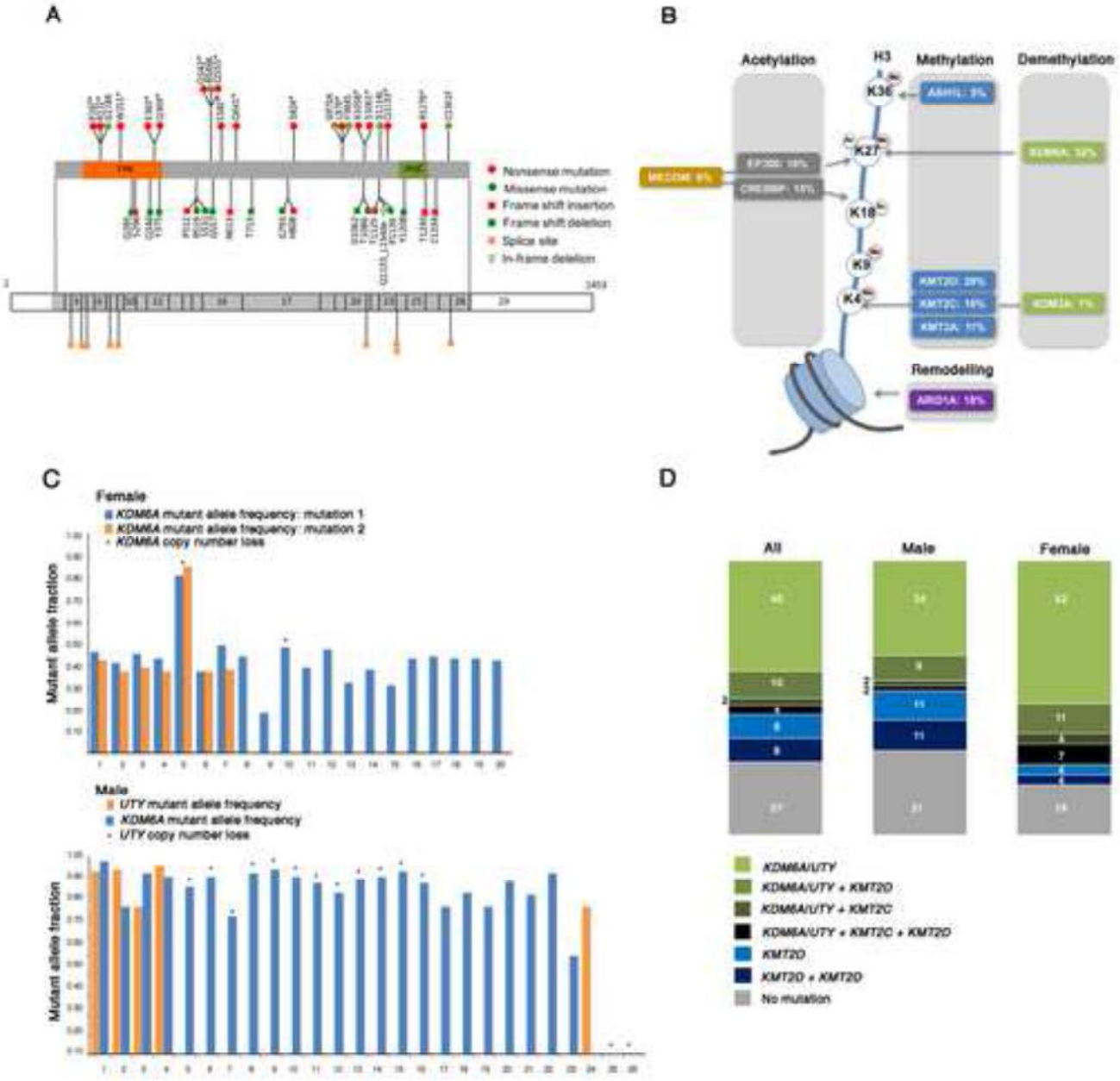


Table S1 related to Figure 1. Details of patients, tumor samples and analyses carried out.
Provided as an Excel File.

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
2004	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
2088	2	INFIL/MS2b1	2
1932	2	GU/MS2a2	2
2082	2	GU/MS2a1	2

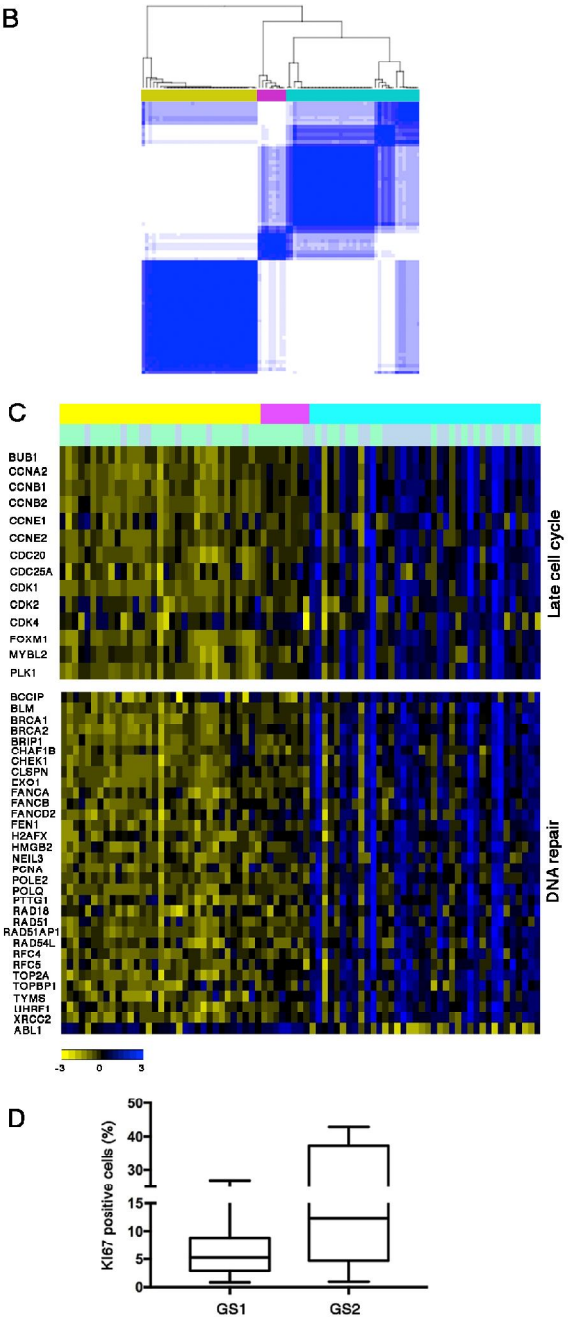


Figure S1 related to Figure 2. Alignment of GS1 and GS2 expression profiles with previously identified NMIBC expression subtypes.

(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.

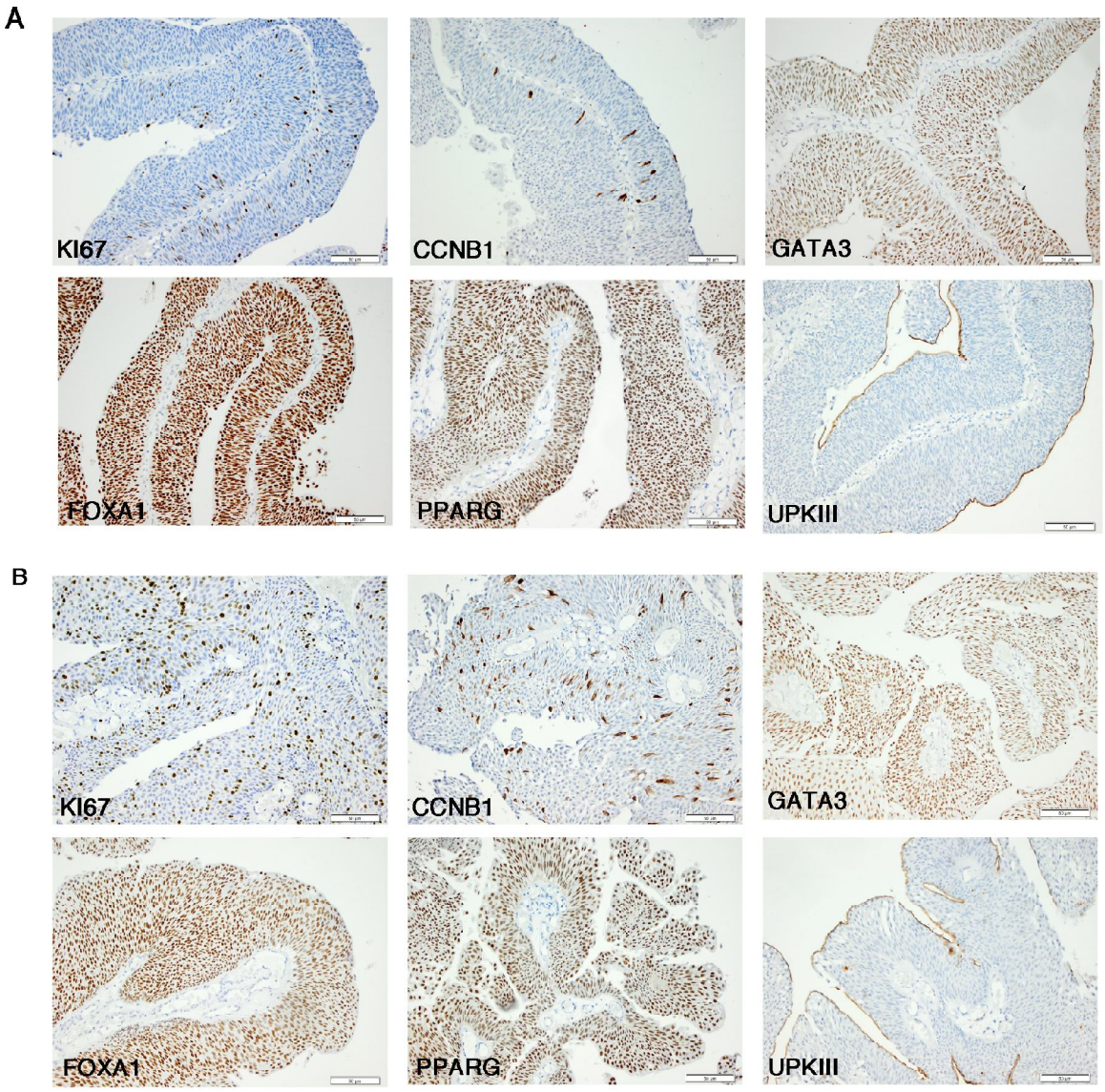


Figure S2, related to Figure 2. Immunohistochemistry for selected proteins in tumors from GS1 and GS2 subtypes.

(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.

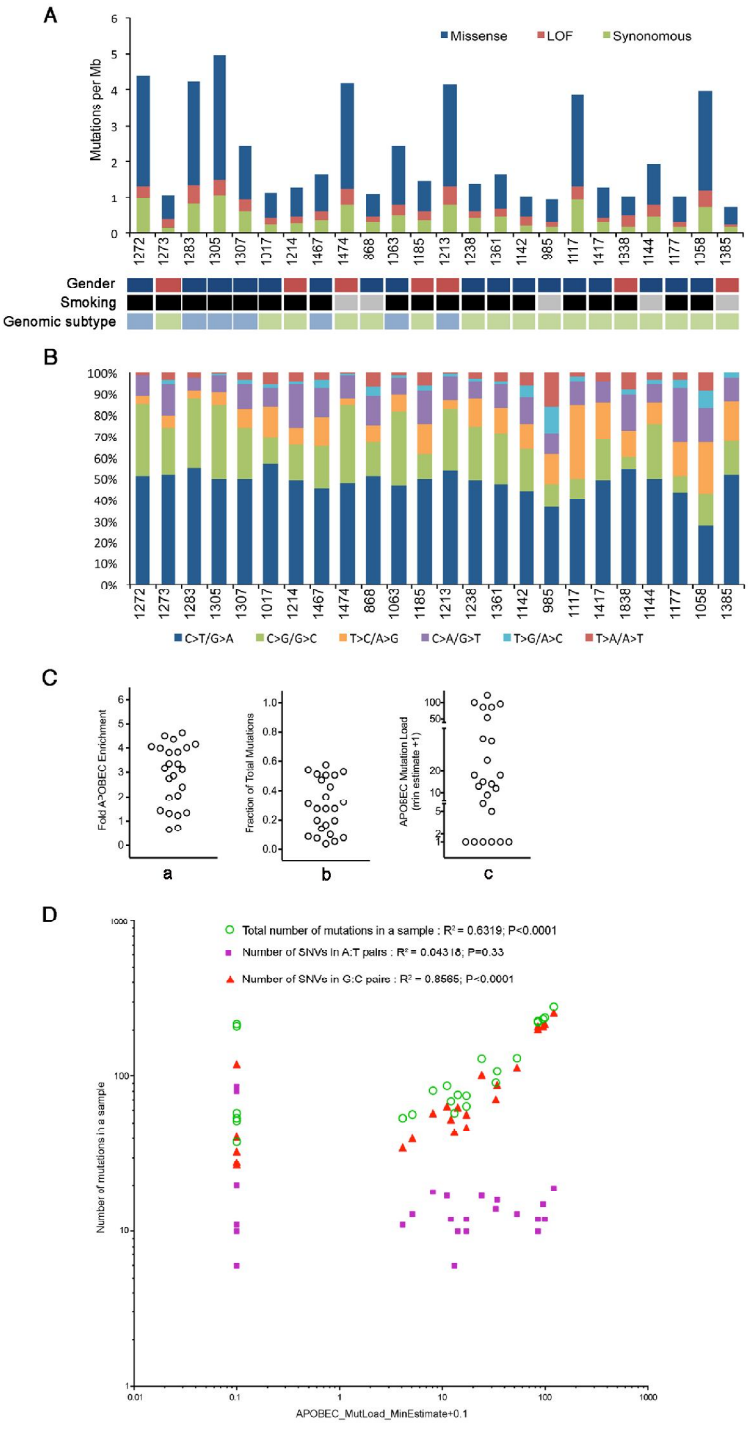


Figure S3, related to Figure 3. Distribution of mutation type and nucleotide substitutions in stage Ta bladder cancer exome sequences.

(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/

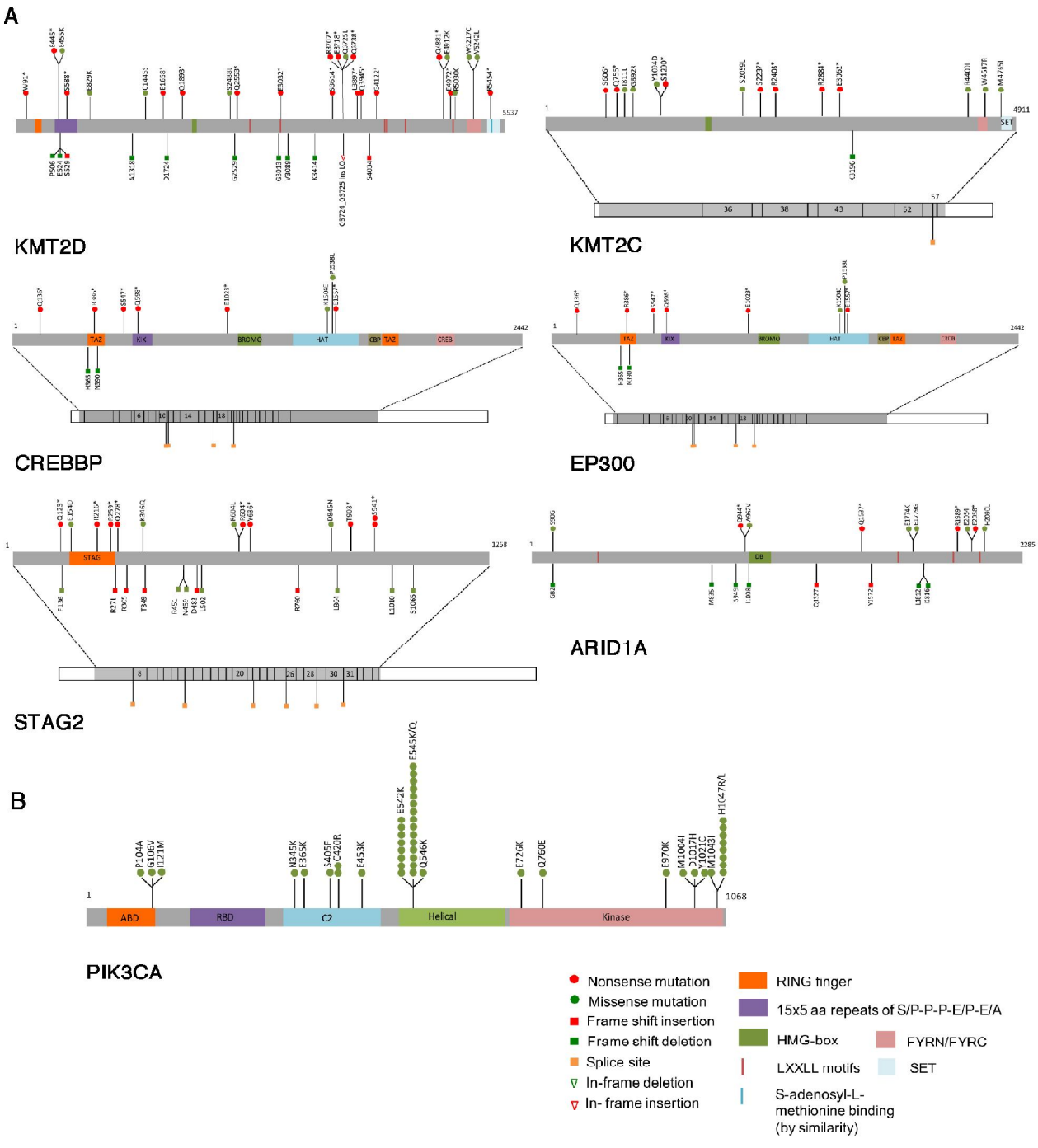
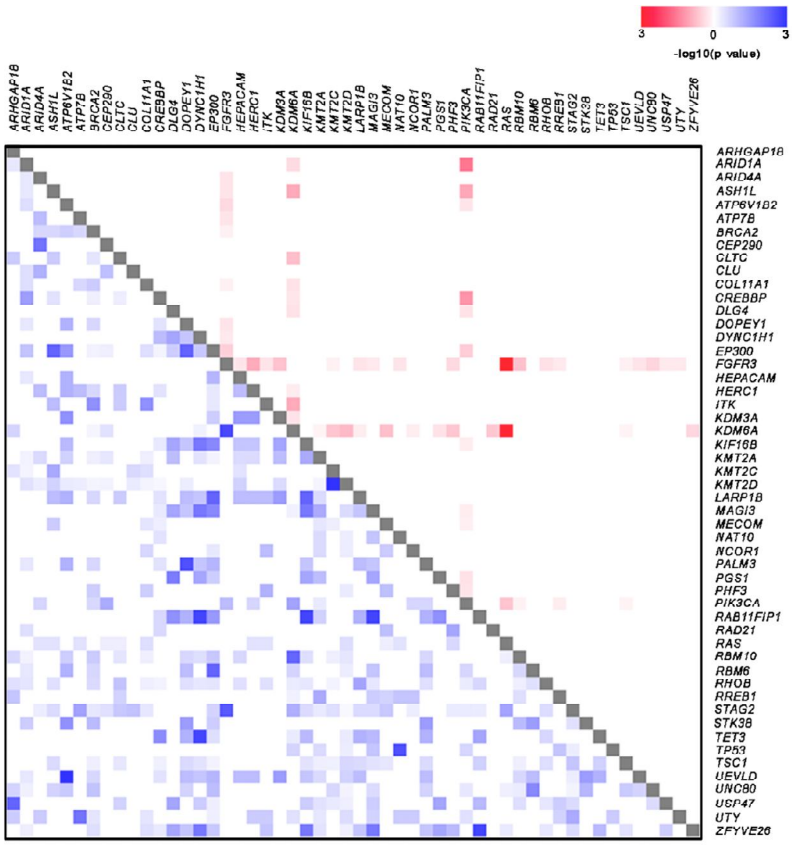


Figure S4, related to Figure 3. Schematics showing the distribution of mutations identified by exome sequencing and targeted re-sequencing in 82 stage Ta bladder tumors.

(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	0.03830349	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

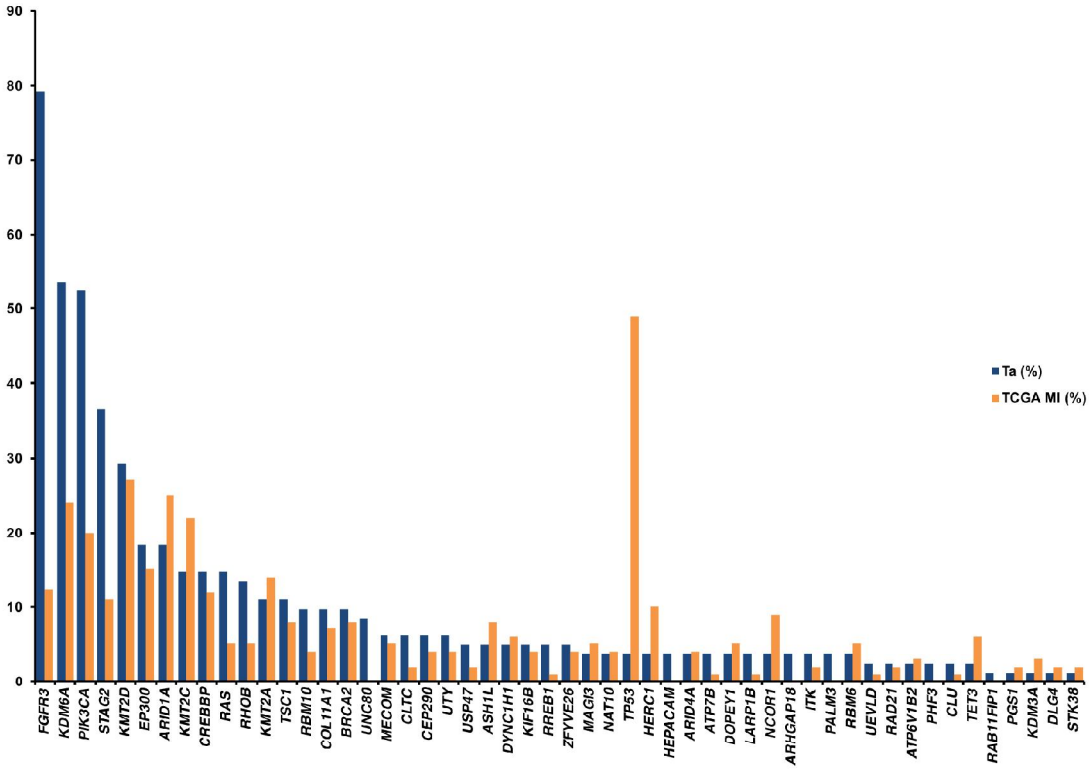


Figure S6, related to Figure 3. Comparison of mutations identified in NIBC and MIBC.

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/

Click here to access/download

Supplemental Movies and Spreadsheets

Final Table S1 copy.xlsx

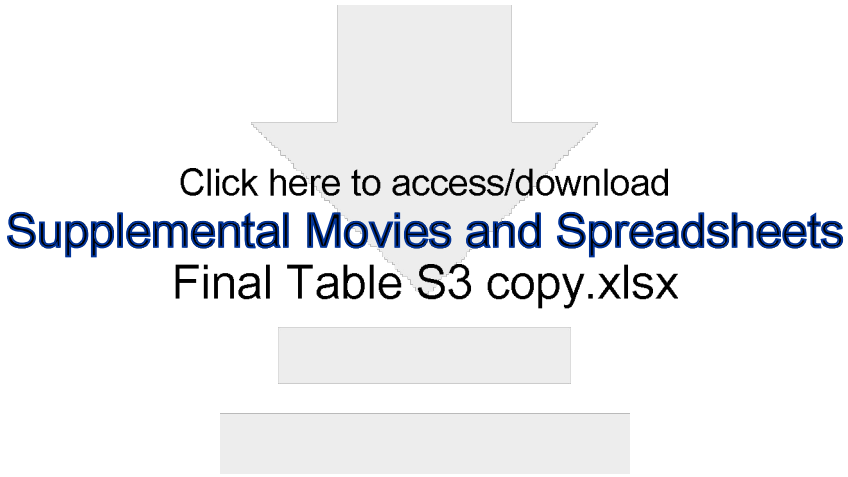
Supplemental Movies and Spreadsheets

Click here to access/download

Supplemental Movies and Spreadsheets

Final Table S2 copy revised 7_8_17.xlsx

Supplemental Movies and Spreadsheets



Click here to access/download

Supplemental Movies and Spreadsheets

Final Table S3 copy.xlsx