

ARTICLE

<https://doi.org/10.1038/s41467-019-10873-y>

OPEN

# Upper tract urothelial carcinoma has a luminal-papillary T-cell depleted contexture and activated FGFR3 signaling

Brian D. Robinson<sup>1,2,3,15</sup>, Panagiotis J. Vlachostergios<sup>4,15</sup>, Bhavneet Bhinder<sup>5,6</sup>, Weisi Liu<sup>4</sup>, Kailyn Li<sup>4</sup>, Tyler J. Moss<sup>7</sup>, Rohan Bareja<sup>5,6</sup>, Kyung Park<sup>1</sup>, Peyman Tavassoli<sup>1</sup>, Joanna Cyrta<sup>1,3</sup>, Scott T. Tagawa<sup>2,3,4,8</sup>, David M. Nanus<sup>2,3,4,8</sup>, Himisha Beltran<sup>2,3,4,8</sup>, Ana M. Molina<sup>2,3,4,8</sup>, Francesca Khani<sup>1,2,3</sup>, Juan Miguel Mosquera<sup>1,2,3</sup>, Evangelos Xylinas<sup>2,9</sup>, Shahrokh F. Shariat<sup>2,10</sup>, Douglas S. Scherr<sup>2</sup>, Mark A. Rubin<sup>1,2,3,8,11</sup>, Seth P. Lerner<sup>12</sup>, Surena F. Matin<sup>13</sup>, Olivier Elemento<sup>3,5,6,8,16</sup> & Bishoy M. Faltas<sup>3,4,8,14,16</sup>

Upper tract urothelial carcinoma (UTUC) is characterized by a distinctly aggressive clinical phenotype. To define the biological features driving this phenotype, we performed an integrated analysis of whole-exome and RNA sequencing of UTUC. Here we report several key insights from our molecular dissection of this disease: 1) Most UTUCs are luminal-papillary; 2) UTUC has a T-cell depleted immune contexture; 3) High FGFR3 expression is enriched in UTUC and correlates with its T-cell depleted immune microenvironment; 4) Sporadic UTUC is characterized by a lower total mutational burden than urothelial carcinoma of the bladder. Our findings lay the foundation for a deeper understanding of UTUC biology and provide a rationale for the development of UTUC-specific treatment strategies.

<sup>1</sup>Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY 10065, USA. <sup>2</sup>Department of Urology, Weill Cornell Medicine, New York, NY 10065, USA. <sup>3</sup>Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY 10021, USA. <sup>4</sup>Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY 10065, USA. <sup>5</sup>Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10065, USA. <sup>6</sup>Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA. <sup>7</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. <sup>8</sup>Sandra and Edward Meyer Cancer Center at Weill Cornell Medicine, New York, NY 10065, USA. <sup>9</sup>Department of Urology, Cochin Hospital, APHP, Paris Descartes University, Paris 75014, France. <sup>10</sup>Department of Urology, Medical University of Vienna, Vienna 1190, Austria. <sup>11</sup>Department of Biomedical Research, University of Bern, Bern 3008, Switzerland. <sup>12</sup>Scott Department of Urology, Baylor College of Medicine, Houston, TX 77030, USA. <sup>13</sup>Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. <sup>14</sup>Department of Cell and Developmental Biology, Weill Cornell Medicine, New York, NY 10065, USA. <sup>15</sup>These authors contributed equally: Brian D. Robinson, Panagiotis J. Vlachostergios. <sup>16</sup>These authors jointly supervised this work: Olivier Elemento, Bishoy M. Faltas. Correspondence and requests for materials should be addressed to B.M.F. (email: [bmf9003@med.cornell.edu](mailto:bmf9003@med.cornell.edu))

Upper tract urothelial carcinoma (UTUC) accounts for 5–10% of all urothelial carcinomas (UCs)<sup>1</sup>. UTUC is a distinct clinical entity with an aggressive clinical behavior and a more advanced presentation compared to urothelial carcinoma of the bladder (UCB)<sup>1</sup>. Recently, the Cancer Genome Atlas (TCGA) study classified UCB into five molecular subtypes (luminal-papillary, luminal-infiltrated, luminal, basal/squamous, neuronal). Since TCGA did not include UTUC<sup>2</sup>, it is currently unknown whether UTUC recapitulates the same molecular subtypes<sup>2–6</sup>. Furthermore, our understanding of the immune milieu of UTUC is incomplete.

These knowledge gaps have hindered the development of effective UTUC-specific therapeutic strategies. To dissect the central biological features of UTUC's tumor and immune cell compartments, we analyzed whole-exome sequencing (WES) and RNA sequencing (RNAseq) data from high-grade UTUC tumors from patients at three different institutions [Weill Cornell Medicine (WCM), Baylor College of Medicine and MD Anderson Cancer Center (BCM-MDACC)]. We used whole-exome and RNAseq data from UCB tumors from the TCGA cohort as a comparison cohort<sup>5</sup>. This enabled us to define the biological differences between UC arising from the upper and lower urinary tracts and to gain insights into the unique mechanisms that drive UTUC biology. We show that UTUC is predominantly luminal-papillary and T-cell depleted. We identify *FGFR3* as a putative regulator of UTUC's immune contexture through attenuation of interferon gamma (IFNG) signaling. Finally, we report that the tumor mutational burden in sporadic UTUC is lower than UCB, despite reduced expression of DNA mismatch repair (MMR) transcripts and proteins.

## Results

**UTUC and UCB mutational profiles.** We performed whole exome sequencing (WES) of 37 UTUC primary tumor–normal pairs. We only included patients with high-grade UTUC tumors. Most patients were former or current smokers (64.8%) (Supplementary Table 1). We identified *FGFR3* mutations in 11/37 (29.7%) (Fig. 1a), a significantly higher frequency compared to the 17/124 (13.7%) mutations detected in TCGA UCB (Wilcoxon test  $P = 0.04$ ) (Fig. 1b). In contrast, we found no significant difference in the prevalence of mutations in chromatin modifying (*KMT2D*, *ARID1A*, *KDM6A*), receptor tyrosine kinase pathway (*PIK3CA*, *HRAS*), transcription factor (*RXRA*, *KLF5*, *ELF3*), and cell cycle regulation (*TP53*, *RB1*, *CDKN1A*, *CDKN2A*) genes between our UTUC and TCGA UCB cohorts (Fig. 1a, b).

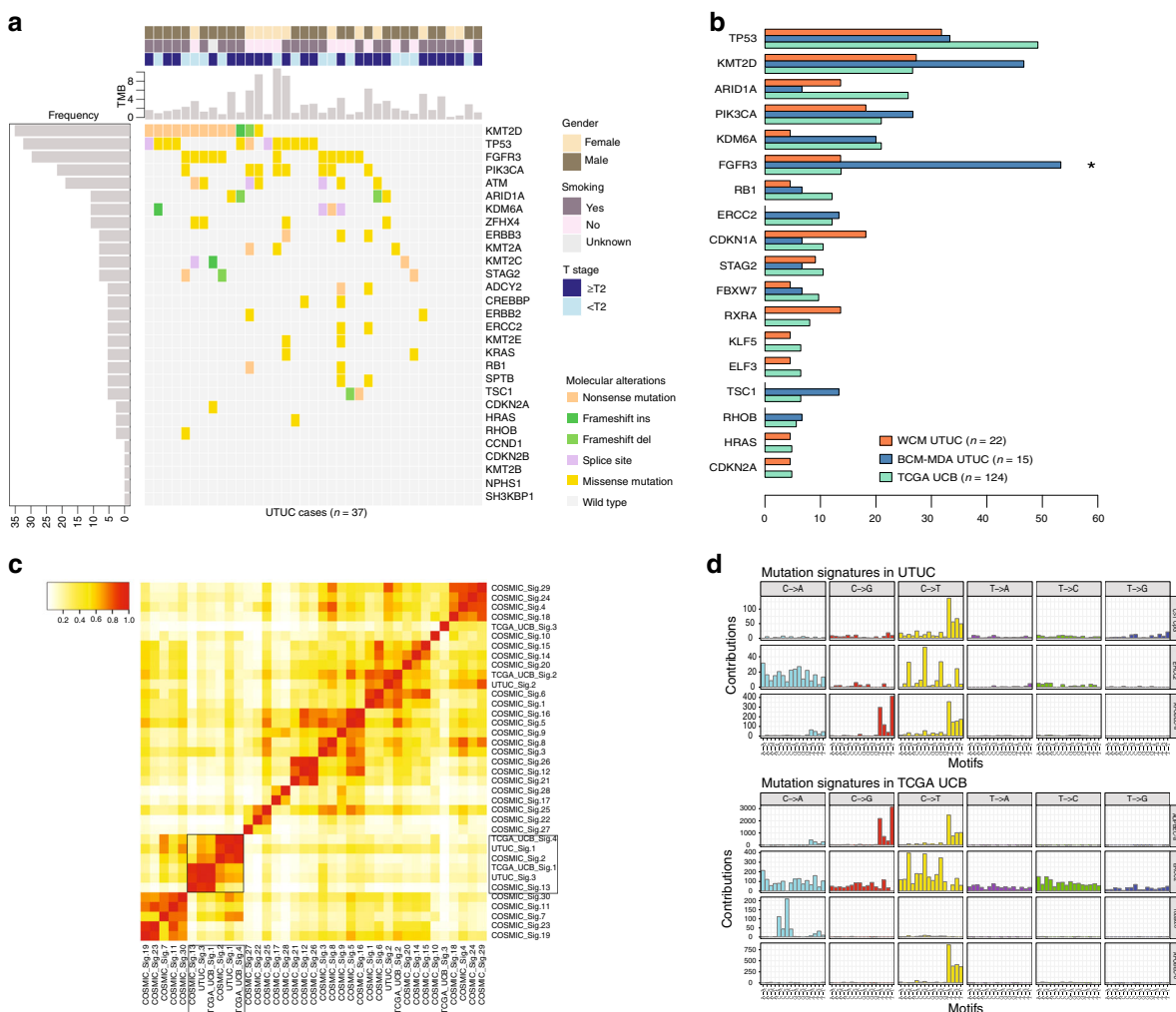
To define the mutagenic mechanisms that shape the genomic landscape of UTUC, we performed a mutational signature analysis to identify the prevalent Catalog of Somatic Mutations in Cancer (COSMIC) signatures in the WCM UTUC, BCM-MDA UTUC, and TCGA UCB cohorts<sup>7</sup>. We identified the COSMIC APOBEC-associated signatures 2 and 13 as the dominant mutational signatures in UTUC (Fig. 1c). We also identified C>T transitions at CpG dinucleotides. This signature is characterized by high numbers of small indels at mono/polynucleotide repeats and is associated with defective MMR (Fig. 1d). Another mutational signature in UTUC was found to be related to defective nucleotide excision repair (NER)<sup>8</sup> (Fig. 1d). Collectively, these data suggest that these three mutational processes (APOBEC, MMR, NER) (Fig. 1d) are responsible for the majority of mutations in UTUC.

**Somatic downregulation of DNA damage repair (DDR) genes in UTUC.** The association between germline mutations in MMR genes that cause microsatellite instability (MSI) and Lynch syndrome and increased susceptibility to the development of

UTUC is well established<sup>9–11</sup>. However, it is unclear whether non-Lynch (sporadic) UTUC patients have defective DDR and an increased mutational burden. To define whether somatic dysregulation of DDR genes could play a similar role in inducing a hypermutated phenotype in non-Lynch UTUC patients, we assessed the mRNA expression level of DDR pathway genes in UTUC (WCM, BCM-MDA) and UCB (TCGA). We focused our analysis on identifying differentially expressed genes in the canonical DDR pathways [MMR, base-excision repair (BER), NER, homologous recombination (HR), non-homologous end joining (NHEJ), Fanconi anemia (FA), translesion synthesis (TLS)] comparing UTUC with UCB (Fig. 2a). We identified a significant somatic dysregulation of 35 canonical DDR genes in UTUC (Fig. 2a). We analyzed germline WES data from patients in our UTUC cohorts (WCM, BCM-MDA) and identified no germline mutations in the canonical MMR genes (*MLH1*, *PMS2*, *MSH2*, *MSH6*). Interestingly, we observed significantly lower somatic mRNA expression of three canonical MMR genes: *MLH1*, *MSH2*, and *MSH6* in UTUC tumors compared to TCGA UCB tumors (Fig. 2a). To determine whether this decreased mRNA levels further translated into a low expression of MMR proteins, we used immunohistochemistry (IHC) to quantify their expression in WCM UTUC tumors ( $n = 16$ ) compared to stage-matched WCM UCB tumors ( $n = 14$ ). We found that the levels of *MLH1*, *PMS2*, *MSH2*, and *MSH6* proteins were significantly lower in UTUC tumors (Fig. 2b, c). This confirmed that lower expression of these proteins is a characteristic feature of UTUC even in the absence of germline or somatic mutations in the respective genes in the WCM UTUC cohort.

To determine whether downregulation of these proteins impaired DNA MMR, we examined WES data for short tandem repeats (MSI) using the MSI sensor program<sup>12</sup> that calculates the percentage of unstable microsatellites from UTUC and UCB tumor–normal paired WES data. Surprisingly, the median MSI sensor scores were similar in WCM UTUC and TCGA UCB. Both were below the previously defined threshold of 3.5, which was shown to accurately classify microsatellite unstable tumors<sup>12</sup> (Fig. 2d). Furthermore, WCM UTUC samples harbored a significantly lower mean total mutational burden (TMB) compared to TCGA UCB tumors (2.91 versus 5.46 mutations per MB) (Fig. 2e). This suggests that the somatic downregulation of MMR proteins is insufficient to produce MSI and is not a major driver of mutagenesis in non-Lynch UTUC. Collectively, these findings indicate that the decrease in the mRNA and protein levels of MMR genes in sporadic UTUC does not translate into MSI or a higher TMB.

**UTUCs are predominantly luminal-papillary.** To characterize the gene expression profiles which define UTUC, we generated an RNAseq meta-dataset from 32 UTUC tumors and the TCGA UCB cohort. To ensure homogeneity, we performed sample normalization of the data prior to standardization and fitted the quantiles of each sample's raw data to be similar. We also compared the z-scores of the mRNA expression of 40 housekeeping genes among tumor samples from WCM UTUC, BCM-MDA UTUC, and TCGA UCB and found no significant differences for the majority of these genes (Supplementary Fig. 1). Using the University of North Carolina (UNC) 47-gene signature (BASE47) classifier<sup>3</sup>, we found that 27/32 (84.3%) of the UTUC tumors clustered with the luminal subtype (Fig. 3a and Supplementary Table 2) as opposed to only 59/128 (46.1%) in UCB<sup>2</sup>. To ensure that this luminal expression pattern is a consistent biological property of UTUC, we confirmed this finding using several different methods. First, we interrogated the same meta-dataset using two additional validated classifiers of urothelial carcinoma

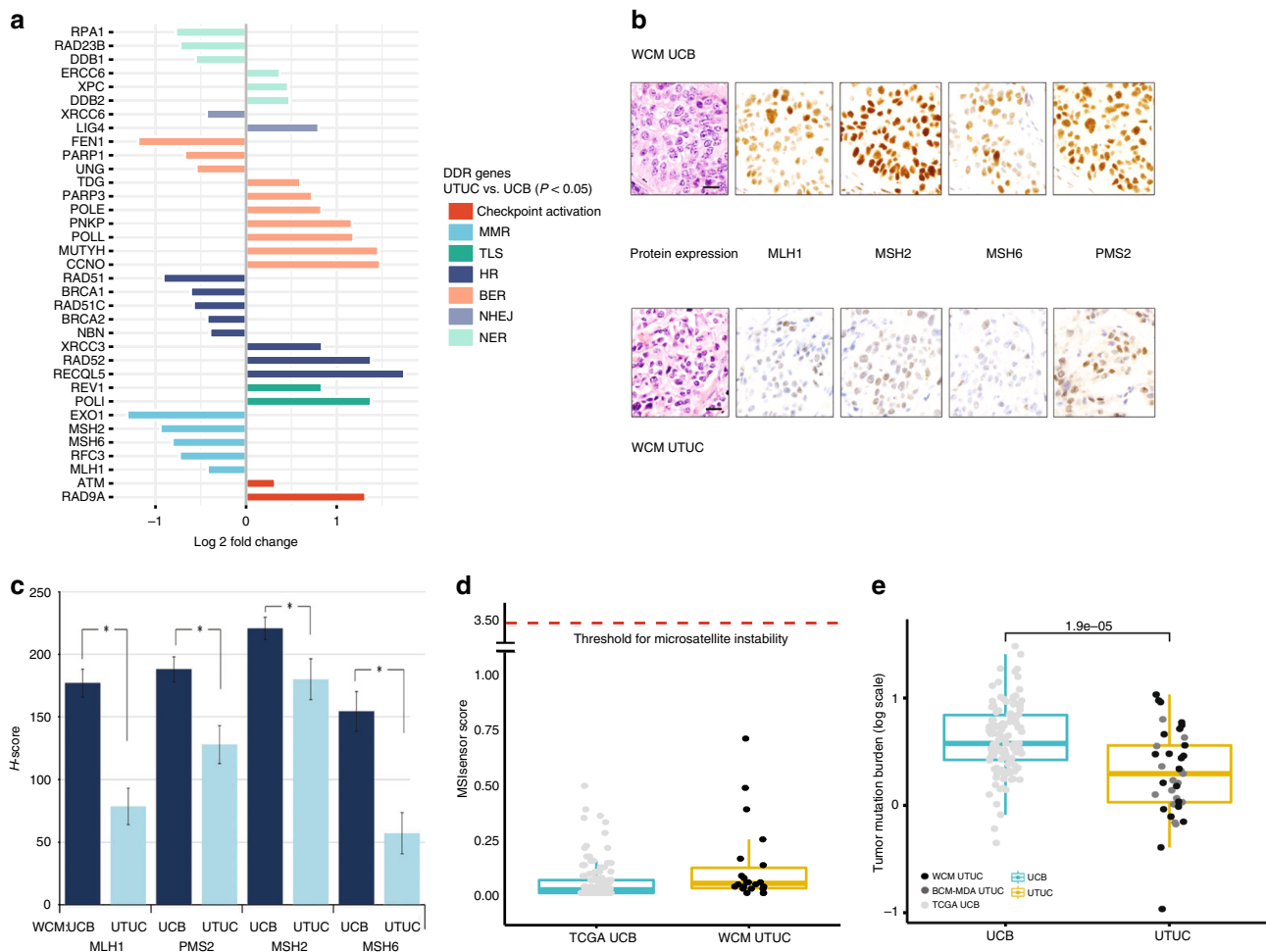


**Fig. 1** Genomic differences between upper tract urothelial carcinoma (UTUC) and urothelial bladder carcinoma (UCB). **a** Prevalence of frequent somatic genomic alterations in 37 patients with UTUC from the Weill Cornell Medicine (WCM UTUC) and Baylor College of Medicine–MD Anderson Cancer Center (BCM-MDACC UTUC) cohorts. The somatic mutational rate for each tumor is represented by vertical barplots (total number of non-silent mutations per megabase (MB)). Patient and tumor characteristics are represented on the top right. **b** Horizontal barplot showing differences in mutational frequencies between WCM UTUC (orange bars), BCM-MDA UTUC (blue bars), and TCGA UCB (green bars) cohorts for frequently mutated genes. Asterisk indicates statistically significant changes in paired comparison between BCM-MDA UTUC and TCGA UCB (\*Fisher's Exact test  $P = 0.001$ ). **c** Heatmap of a cosine similarity matrix of COSMIC mutational signatures and observed mutational signatures in UTUC (WCM, BCM-MDA) and TCGA UCB cohorts. UTUC signature clusters with COSMIC signatures 2, 13 (APOBEC-associated). **d** Dominant mutational signatures in UTUC (top panel) include the C>T CpG signature, attributable to mutagenesis via defective mismatch repair (MMR) (COSMIC signature 6), the ERCC2 signature, consistent with COSMIC signature 5, and the APOBEC signatures (COSMIC signatures 2 and 13) resulting from the endogenous cytidine deamination induced by the APOBEC3 enzyme family. TCGA UCB mutational signatures are shown for comparison (bottom panel)

subtypes. When we applied the MDACC classifier which divides UC into luminal, basal, and p53-like subtypes<sup>4</sup>, we found that 22/32 (68.7%) of UTUC tumors clustered with the luminal subtype (Supplementary Fig. 2 and Supplementary Table 2) versus only 47/128 (36.7%) of UCB<sup>2</sup>. Using the recent TCGA classifier which segregates UC into luminal-papillary, luminal-infiltrated, luminal, basal-squamous, and neuronal subtypes<sup>2</sup>, we confirmed that UTUC has a luminal-papillary phenotype (20/32, 62.5%), with the majority of remaining UTUC tumors also exhibiting a luminal expression profile (8/12, 67%) (Supplementary Fig. 3 and Supplementary Table 2). This is in contrast to 35/128 (27.3%) of luminal-papillary UCB tumors in the TCGA UCB cohort, with only a minority of the rest (35/93, 37.6%) also segregating with luminal or luminal-infiltrated subtypes<sup>2</sup>. To confirm these findings using a different approach, we used non-negative matrix factorization (NMF)<sup>13</sup>, a sensitive unsupervised statistical method to dissect and extract the key biological features of UTUC from

our high dimensional RNAseq dataset<sup>13</sup>. The NMF analysis revealed three principal components characterized by luminal/carcinoma in situ (CIS)-low, basal/squamous-like, and extracellular matrix (ECM)/epithelial-mesenchymal transition (EMT)-related gene sets<sup>2</sup>. Principal component analysis (PCA) was then performed on the coefficients obtained from NMF to visualize the separation of the three components (Fig. 3b). This demonstrated that the luminal-papillary component is a defining feature of UTUC. Our results suggest that the majority of UTUCs represent a distinct subset within the continuum of UC differentiation that shares similar characteristics with the luminal-papillary subtype of UCB.

**UTUC has a T-cell depleted immune contexture.** The immune contexture of tumors is an essential determinant of the host's anti-cancer immune response<sup>14</sup> and clinical outcomes. To dissect



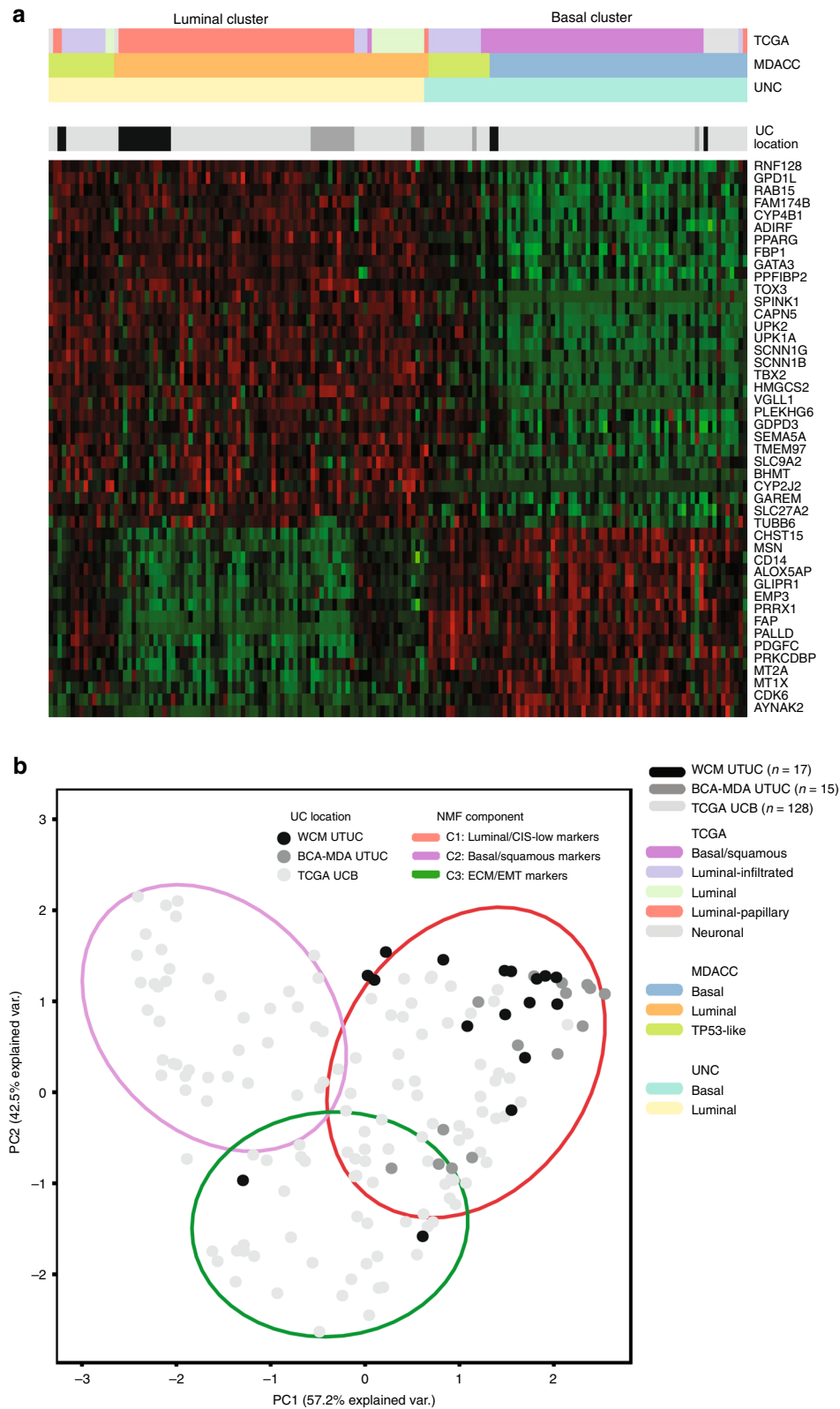
**Fig. 2** UTUC is characterized by decreased expression of canonical MMR proteins. **a** Comparative histogram of DNA damage response (DDR) genes with statistically significant differential expression ( $\log_2$  fold) highlighting significantly lower expression of mismatch repair (MMR) pathway genes (*MLH1*, *MSH2*, *MSH6*, *RFC3*, *EXO1*) in WCM UTUC compared to TCGA UCB (one-way ANOVA  $\log_2$ FC adjusted  $P \leq 0.05$ ). **b** Representative micrographs (magnification  $\times 400$ ) showing lower expression of MMR proteins in a UTUC tumor compared to a UCB tumor by immunohistochemistry. Scale bars represent  $25 \mu\text{m}$ . **c** Mean *H*-scores of MMR proteins are significantly lower in WCM UTUC tumors compared to WCM UCB tumors. Error bars show standard deviation (S.D.) for each MMR protein. Asterisks indicate statistically significant changes in paired comparison (*t*-test  $P < 0.05$ ). **d** Mean microsatellite instability (MSI) score is below the threshold of 3.5 in both WCM UTUC and TCGA UCB tumors. **e** Mean total mutational burden (TMB), expressed in log scale is lower in UTUC (WCM, BCM-MDA) compared to TCGA UCB tumors (Mann-Whitney test  $P = 1.9 \times 10^{-5}$ ). The horizontal lines within the boxes in the boxplots indicate the mean, boundaries of the boxes indicate the 25th-percentile and 75th-percentile, and the whiskers indicate the highest and lowest values of the results

the gene expression profile of UTUC that characterizes its immune contexture, we developed a 170-gene classifier comprising key immune genes (Supplementary Table 3). This classifier separated tumors independent of their anatomical origin into T-cell inflamed, and T-cell-depleted clusters. Interestingly, the majority of UTUC (WCM, BCM-MDA) tumors (28/32, 87.5%) were T-cell depleted (Fig. 4a) with consistent downregulation of T-cell related (*CD8A*, *CCL2*, *CCL3*, *CCL4*, *CXCL9*, *CXCL10*)<sup>15</sup> and *IFNG* signaling genes<sup>16</sup> (Fig. 4a). In contrast, TCGA UCB tumors were almost evenly distributed between T-cell inflamed (57/128, 44.5%) and T-cell depleted (71/128, 55.5%) immune subtypes (UTUC vs. UCB (Fisher's exact test  $P = 0.0009$ ) (Fig. 4a).

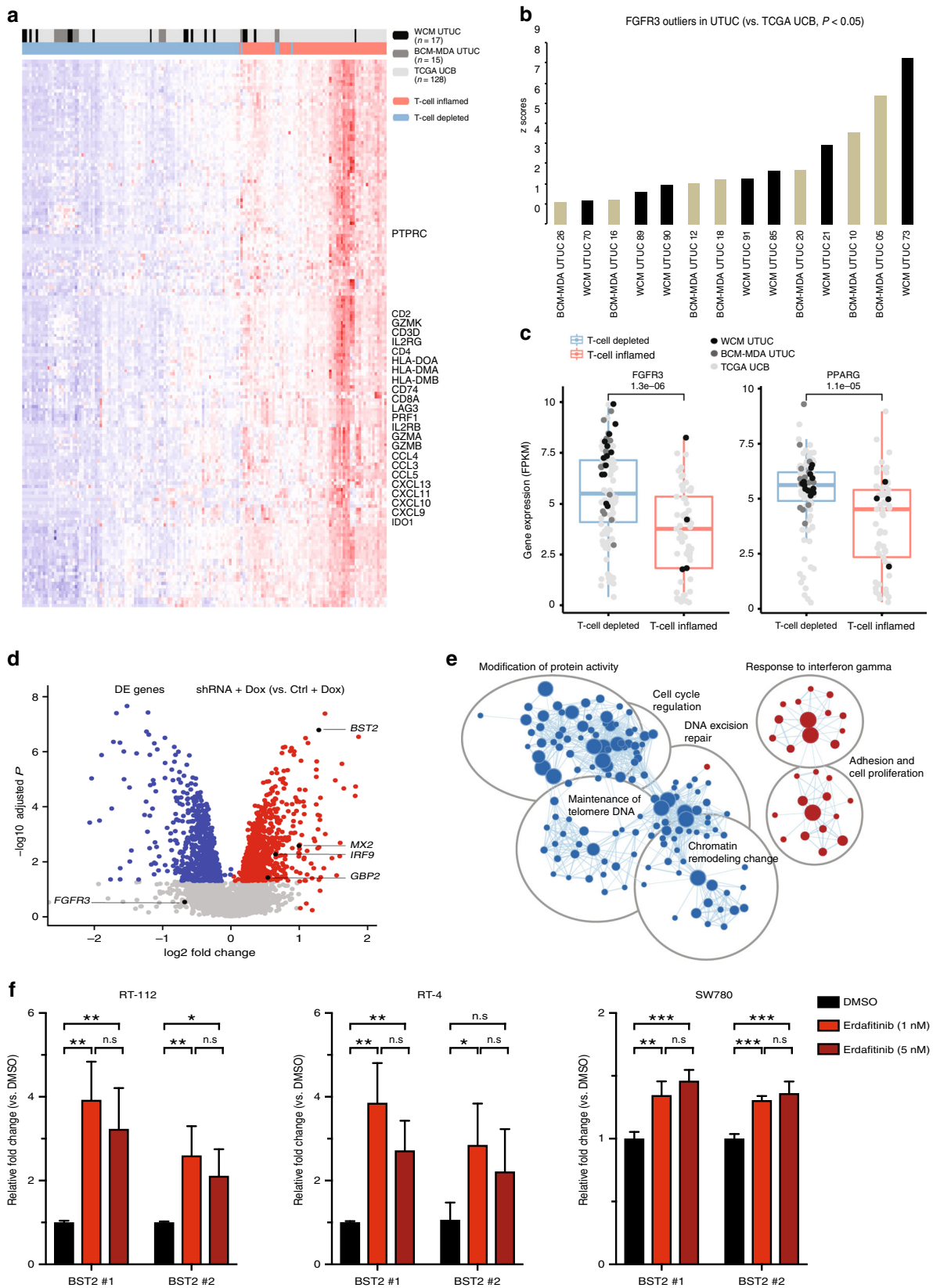
**FGFR3 is a putative driver of UTUC's immune-depleted contexture.** To identify the signaling pathways that characterize UTUC's immune contexture, we performed gene expression analysis to detect outliers. We detected outlier *FGFR3* mRNA

expression in 14/32 (43.7%) of the tumors in our UTUC (WCM, BCM-MDA) cohorts (Fig. 4b). We identified nine activating missense mutations in these tumors. These findings suggest that activated *FGFR3* signaling potentially plays a prominent role in the distinct gene expression profile of UTUC.

We then went on to assess whether *FGFR3* is differentially expressed between T-cell inflamed and T-cell-depleted tumors. Indeed, we identified significantly higher *FGFR3* expression in the T-cell-depleted cluster which harbored the majority of the UTUC tumors (Fig. 4c). We interrogated the functional link between *FGFR3* signaling and the T-cell-depleted phenotype observed in UTUC. We reanalyzed a previously published dataset of mRNA expression from the RT-112 UC cell line following doxycycline (dox)-inducible short hairpin RNA (shRNA) knockdown of *FGFR3*<sup>17,18</sup>. We found that several *IFNG* response genes including *BST2*, *MX2*, *IRF9*, *GBP2* were upregulated after *FGFR3* knockdown (Fig. 4d). Confirming these results, we found that *BST2* and *IRF9* were significantly downregulated in



**Fig. 3** UTUC is predominantly luminal. **a** Supervised consensus clustering and heatmap of mRNA expression data from WCM UTUC, BCM-MDA UTUC, and TCGA UCB metadataset. BASE47 classifier (UNC) genes are listed on the right. Assigned TCGA, MDACC, and UNC clusters are represented on top (color key, bottom right). WCM UTUC and BCM-MDA UTUC cluster with the luminal subtype (yellow vertical bars) by UNC criteria, luminal subtype by MDACC criteria (orange horizontal bars) and luminal-papillary subtype by TCGA classification (red horizontal bars). **b** Non-negative matrix factorization (NMF) of WCM UTUC, BCM-MDA UTUC, and TCGA UCB tumors segregates gene expression along three principal components: basal/squamous (purple), luminal/CIS-low (pink), and extracellular matrix/epithelial-mesenchymal transition (ECM/EMT) (green). WCM UTUC tumors represented as black dots and BCM-MDA tumors represented as dark gray dots cluster with the luminal/CIS-low component



the T-cell-depleted cluster which harbored the majority of UTUCs in our patient cohort (Supplementary Fig. 4). We also found a statistically significant upregulation of other IFNG response genes in the *FGFR3* shRNA + dox dataset when

compared to the control (ctl) + dox dataset (Fig. 4e). To confirm this observation and define whether pharmacologic inhibition of *FGFR3* will have a similar effect on IFNG response genes (Supplementary Table 4), we tested the effects of erdafitinib, a

**Fig. 4** FGFR3 plays an important role in the T-cell-depleted immune contexture of UTUC. **a** UTUC is T-cell depleted. Supervised consensus clustering of WCM UTUC, BCM-MDA UTUC, and TCGA UCB tumors according to a 170-immune gene signature classifies tumors into T-cell depleted (with lower expression of classifier genes), and T-cell inflamed (with higher expression of classifier genes) clusters (Fisher's exact test  $P = 9 \times 10^{-5}$ ). **b** FGFR3 is an expression outlier in UTUC tumors. WCM UTUC and BCM-MDA UTUC tumors are represented on the x-axis, and normalized z-scores of gene's expression represented on the y-axis. **c** Boxplots of mean expression of *FGFR3* and *PPARG* genes [in Fragments Per Kilobase of transcript per Million mapped reads (FPKM)] within the T-cell-depleted versus T-cell inflamed clusters (*FGFR3*: Wilcoxon test  $P = 1.3 \times 10^{-6}$ ; *PPARG*: Wilcoxon test  $P = 1.1 \times 10^{-5}$ ). The horizontal lines within the boxplots indicate the mean, boundaries of the boxes indicate the 25th-percentile and 75th-percentile, and the whiskers indicate the highest and lowest values of the results. **d** Interferon gamma (IFNG)-response genes are upregulated in response to FGFR3 knockdown. Volcano plot of differential fold expression of genes ( $\log_{FC} < 0$  vs.  $\log_{FC} > 0$ ;  $t$ -test adjusted  $P < 0.05$ ) in *FGFR3* shRNA + Doxycycline compared to control + Doxycycline UCB RT-112 cells (*BST2*:  $P < 0.001$ ; *GBP2*:  $P = 0.038$ ; *IRF9*:  $P = 0.005$ ; *MX2*:  $P = 0.003$ ). **e** Enrichment map of cancer-related pathways with significant positive and negative enrichment in *FGFR3* shRNA UCB RT-112 cells. Node size corresponds to the number of genes within each gene set. The up-regulated nodes were represented in red while the down-regulated clusters were represented in blue. Overall, 476/3534 gene sets were upregulated, and 651/3534 gene sets were downregulated ( $t$ -test  $P < 0.05$ ; false discovery rate (FDR)  $< 0.25$ ). Several IFNG response gene sets were enriched after *FGFR3* blockade ( $t$ -test  $P < 0.001$ ; FDR = 0.062). **f** Pharmacologic inhibition of FGFR3 upregulates IFNG-response gene *BST2* using two different primer pairs (*BST2* #1, *BST2* #2). Barplots showing relative fold increase (mean  $\pm$  SD) of mRNA levels of *BST2* (*BST2* #1 and *BST2* #2) after treatment with erdafitinib at 1 and 5 nM compared to DMSO vehicle in RT-112, RT-4, and SW780 cells. No statistically significant differences were observed in the expression of *BST2* between the 1 and 5 nM erdafitinib conditions in any of the tested cell lines (statistical significance level is denoted by asterisks \* $t$ -test  $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , n.s.: non-significant). Error bars show standard deviation (S.D.) for each condition

small molecule FGFR3 inhibitor, in three different UC cell lines (RT-112, RT-4, and SW780). These cell lines harbor FGFR3 fusions (RT-112, RT-4: FGFR3-TACC3; SW780: FGFR3-BAIAP2L1) resulting in constitutively activated FGFR3 signaling<sup>19</sup>. We found that treatment with erdafitinib led to a significant upregulation of *BST2*, a hallmark of activated interferon signaling (Fig. 4f). Collectively, these findings show that FGFR3 plays an important role in shaping the T-cell-depleted phenotype in UTUC in a cancer-cell autonomous manner.

## Discussion

We performed a comprehensive genomic and transcriptomic analysis of UTUC to identify the key biological features that differentiate UTUC from UCB. We found that the majority of UTUC tumors are luminal-papillary and T-cell depleted.

Previous studies identified a link between Lynch syndrome caused by germline mutations in MMR genes and UTUC<sup>9–11</sup>. However, the vast majority of UTUCs arise sporadically (in non-Lynch syndrome patients). Our analysis aimed to define whether sporadic UTUC patients had MSI or a higher TMB, which are independent predictors of response to immune checkpoint inhibition<sup>20,21</sup>. Our study revealed a lower TMB in non-Lynch syndrome UTUC compared to UCB. We found that the lower expression of canonical MMR mRNA and proteins in non-Lynch UTUC was insufficient to produce significant MSI or a higher TMB compared to UCB. Unlike the complete absence of MSH2, MLH1, MSH6, or PMS2 or MLH1 protein expression observed in Lynch syndrome patients caused by loss-of-function germline mutations, the incomplete loss of these proteins observed in sporadic UTUC is not sufficient to cause MSI<sup>22</sup>. Even low MMR protein expression is adequate for maintaining functional MMR and preserving microsatellite stability<sup>23</sup>. These results are also consistent with a recent study showing MSI sensor scores  $< 3.5$  in the majority of non-Lynch syndrome UTUC patients<sup>24</sup>. Taken together, our findings suggest that, contrary to the prevalent notion, sporadic UTUC is not hypermutated. This is especially important when considered in conjunction with our data showing that the majority of sporadic UTUCs are also consistently luminal-papillary and T-cell depleted. This new understanding of the mutational landscape and immune contexture in non-Lynch syndrome UTUCs (which constitute the majority of cases) further explains the lack of higher response rates of UTUC compared to UCB in clinical trials of immune checkpoint inhibitors (ICIs)<sup>20,21</sup>. In a prospective phase 2 study

of advanced, post-platinum UC patients (including UTUC), atezolizumab demonstrated the lowest response rates in the luminal-papillary (cluster I) subtype compared to other subtypes<sup>20</sup>. In a different trial, response to nivolumab was lower in the luminal 1 (cluster I) UC tumors with low expression of IFNG signature genes<sup>15</sup>.

We identified a putative role for upregulated FGFR3 in UTUC in shaping the immune contexture of T-cell-depleted UTUC tumors. This observation is consistent with the previously described enrichment of the *FGFR3* gene signature in luminal UCB tumors<sup>2–6,25</sup> and the association of FGFR upregulation with T-cell depletion in tumors of pancreatic and breast origin<sup>14</sup>. We observed a consistent increase in *BST2* following pharmacologic FGFR3 inhibition in three different UC cell lines that harbor activating FGFR3 fusions. *BST2* is a viral restriction factor which is canonically induced by interferon<sup>26</sup>. This is also consistent with the role of FGFR3 in blocking the Y701 tyrosine phosphorylation required for STAT1 activation<sup>27</sup>. Taken together, these findings provide putative mechanistic links between FGFR3 and IFNG signaling and suggest that FGFR3 inhibition potentially remodels the immune contexture of UTUC by upregulating interferon response genes to reverse its T-cell-depleted phenotype.

Our observations also provide a rationale for combining FGFR3 inhibitors with PD-1/PD-L1 inhibitors as a targeted therapeutic strategy to modulate the T-cell-depleted phenotype of UTUC. Preliminary clinical trial results using two pan-FGFR inhibitors, erdafitinib, and BGJ398 in several cancers enriched for *FGFR* genomic alterations including urothelial carcinoma, are encouraging<sup>28,29</sup>. Erdafitinib was granted accelerated approval by the FDA in relapsed/refractory metastatic bladder cancer on the basis of phase 2 trial results showing a response rate of 32.2% in 87 patients with tumors that harbored actionable *FGFR* alterations<sup>30</sup>. Our findings suggest that clinical trials of FGFR3 inhibitors as single agents or in combination with immune checkpoint blockade as a UTUC-targeted therapeutic strategy is warranted. This strategy is also potentially applicable to other tumor types harboring *FGFR3*-activating molecular alterations.

Our study has multiple strengths. We used different approaches to confirm that the predominantly luminal-papillary phenotype of UTUC is a consistent biological feature. In a previous study, unsupervised clustering of RNAseq data from both high-grade and low-grade UTUC was used to divide UTUC into four molecular subtypes<sup>31</sup>. Here, we used an alternative approach to position high-grade UTUC within the continuum of UC biology.

Our study also has several limitations. Even though we included patients from three major academic institutions, our cohort is still limited by sample size due to the relative rarity of this tumor type. Further confirmation of our findings in larger UTUC cohorts is warranted. Future studies also need to examine the stability of the molecular profiles we identified across matched primary and metastatic UTUC tumors.

In summary, our findings lay the foundation for a deeper understanding of the key features of the biology of UTUC. Based on this knowledge, we provide a roadmap for the rational clinical development of targeted and immunotherapeutic strategies that are specific to UTUC but also potentially applicable to other tumor types harboring *FGFR3*-activating molecular alterations.

## Methods

**Patient enrollment and tissue acquisition.** The study was approved by our Institutional Review Boards (Weill Cornell Medicine (WCM)/New York-Presbyterian (NYP) IRB protocols for Tumor Biobanking—0201005295, GU tumor Biobanking—1008011210, Urothelial Cancer Sequencing—1011011386, Comprehensive Cancer Characterization by Genomic and Transcriptomic Profiling—1007011157 and Precision Medicine—1305013903). Banked excess tissue was collected from nephroureterectomy specimens of patients with a diagnosis of high-grade UTUC. UTUC high-grade samples were obtained from patients under protocols approved by institutional review boards using endoscopic biopsy or surgical resection at BCM and MDACC<sup>31</sup>. All tumor samples consisted of conventional UC. Samples were selected based on pathologic diagnosis according to standard guidelines for UTUC<sup>1,32</sup>. All pathology specimens were reviewed and reported by board-certified genitourinary pathologists in the Department of Pathology at WCM/NYP, BCM and MDA. Clinical charts were reviewed by the authors (P.J.V., T.J.M., S.F.M., S.P.L., B.M.F.) to record patient demographics, tobacco use, treatment history, anatomic site, the presence of concurrent bladder cancer, pathologic grade and stage using tumor, node, metastasis (TNM) system. DNA for WES was extracted from tumors and matched normal tissues and RNA was purified from tumors for RNAseq.

**DNA extraction and WES.** For WCM UTUC samples, we used our established WES protocol<sup>33,34</sup>. After macrodissection of target lesions, tumor DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) or cored OCT-cryopreserved tumors using the Promega Maxwell 16 MDx (Promega, Madison, WI, USA). Germline DNA was extracted from normal kidney tissue adjacent to the tumor, using the same method. Pathological review by one of the study pathologists (B.D.R., J.M.M., M.A.R.) confirmed the diagnosis and determined tumor content. A minimum of 200 ng of DNA was used for WES. DNA quality was determined by TapeStation Instrument (Agilent Technologies, Santa Clara, CA) and was confirmed by real-time PCR before sequencing. Sequencing was performed using Illumina HiSeq 2500 (2 × 100 bp). A total of 21,522 genes were analyzed with an average coverage of 85× using Agilent HaloPlex Exome (Agilent Technologies, Santa Clara, CA). For BCM-MDA samples, DNA was purified from tumor and matched normal tissues and used for WES<sup>31</sup>.

**WES data processing pipeline.** All the WCM samples data were processed through the computational analysis pipeline of the Institute for Precision Medicine at Weill Cornell, New York Presbyterian Hospital (IPM-Exome-pipeline)<sup>32</sup>. Raw reads quality was assessed with FASTQC<sup>33</sup>. Pipeline output includes segment DNA copy number data, somatic copy-number aberrations (CNAs) and putative somatic single-nucleotide variants (SNVs). Bioinformatic analysis of BCM-MDA samples data was performed<sup>31</sup>.

**Single nucleotide variants.** We developed a consensus somatic SNVs calling pipeline to enhance the accuracy of these calls for WCM samples. SNVs were identified in the paired tumor–normal samples using MuTect2, Strelka, VarScan, and SomaticSniper, and only the SNVs identified by at least two mutation callers were retained. Indels (insertions or deletions) were identified using Strelka and VarScan and only those identified by both tools were retained. The identified somatic alterations were further filtered using the following criteria: (a) read depth for both tumor and matched normal samples was ≥ 10 reads, (b) the variant allele frequency (VAF) in tumor samples was ≥ 5% and >3 reads harboring the mutated allele, (c) the VAF of matched normal was ≤ 1% or there was just one read with mutated allele. The variants were annotated using Oncotator (version 1.9); the dbSNPs amongst the mutation calls, unless also found in the COSMIC database, were filtered out. For the IPMs samples, the promiscuous mutation calls, previously identified internally as artifacts for HaloPlex were also excluded from the final list of mutations. Tumor mutation burden (TMB) was calculated for each sample as the number of mutations divided by the number of bases in the coverage space per million. Somatic mutations were called via a standard cancer analysis pipeline at

the BCM Human Genome Sequencing Center and by using VARSCAN2 for BCM-MDA samples<sup>31</sup>.

**Mutational signature analysis.** Somatic alterations identified from the WES analysis pipeline were used to identify underlying patterns of mutational signatures. The nonsynonymous SNVs were classified into the six base substitution classes and the bases immediately to the 5' and 3' of the mutated base gave 16 different mutational contexts; hence there were 96 different base substitution patterns in total. All base substitutions were reported in context of pyrimidines and in 5' to 3' direction. The signatures were identified using the counts of 96 base substitutions for each sample, based on the Bayesian NMF<sup>8</sup>. The signatures discovered in the WCM UTUC, BCM-MDA UTUC and TCGA UCB cohorts were compared to the 30 COSMIC signatures using hierarchical clustering of cosine similarity amongst these signatures with 'ward.D2' linkage.

**MMR histochemical expression and H-score calculation.** Expression of MMR proteins was assessed in 16 WCM UTUC tumors and 14 matched archival WCM UCB. IHC was performed on 4-μm-thick formalin-fixed paraffin-embedded tissue sections using a Leica Bond III automated stainer. Mouse antibodies against MLH1 (G168-728, 1:25 dilution, BD Biosciences), PMS2 (A16-4, 1:100 dilution, BD Biosciences), MSH2 (FE11, 1:200, EMD Millipore), and MSH6 (44/MSH6, 1:200, BD Biosciences) were used. IHC slides were scanned at ×200 total magnification using a single z-plane via an Aperio AT2 whole slide scanner (Leica Biosystems, San Diego, CA, USA). The scanned images were loaded onto the HALO<sup>TM</sup> imaging analysis platform (Indica Labs, Corrales, New Mexico, USA). Study pathologists manually selected tumor areas for automated image scoring, and the HALO<sup>TM</sup> analysis software determined the staining intensity of each tumor cell (0, 1+, 2+, 3+) and percentage of tumor cells for each intensity level. H-scores were then calculated using the formula [1×(% of cells with intensity of 1+)+2×(% cells 2+)+3×(% cells 3+)] with possible scores thus ranging from 0 to 300.

**Computational detection of MSI.** MSI in WCM UTUC and TCGA UCB samples was detected by MSI sensor. MSI sensor is a software tool that quantifies MSI in paired tumor–normal genome sequencing data and reports the somatic status of corresponding microsatellite sites in the human genome<sup>12</sup>. MSIsensor score was calculated by dividing the number of microsatellite unstable by the total number of microsatellite stable (MS) sites detected. The cut-off for defining MSI-high (MSI-H) versus MS stable (MSS) samples was 3.5 (MSI-H > 3.5, MSS < 3.5)<sup>12</sup>.

**Germline variant calling pipeline.** Germline samples used in this study were normal tissue from fresh frozen or formalin-fixed paraffin-embedded tissue from nephroureterectomy archival specimens of patients with a diagnosis of UTUC at WCM<sup>35</sup>. We applied a germline variant calling pipeline based on the Burrows–Wheeler Aligner (BWA), and the Genome Analysis Toolkit (GATK), for base recalibration, realignment around indels and variant calling. We devised a variant filtering strategy to narrow down the most important and likely clinically relevant variants. For each variant, we collected annotations from databases including ClinVar and Exome Aggregation Consortium (ExAC, <http://exac.broadinstitute.org>)<sup>36</sup>. Following the ACMG Standards and Guidelines regarding classification and interpretation of sequence variants<sup>37</sup>, the variants were classified into five categories: Pathogenic Likely Pathogenic, Likely Benign, Benign and variants of unknown significance. Germline samples from the BCM-MDA cohort were previously analyzed<sup>31</sup>. None of the included BCM-MDA UTUC patients had a diagnosis of Lynch syndrome<sup>31</sup>.

**RNA extraction, RNAseq, and data analysis.** RNA was extracted from frozen material for RNA-sequencing (RNA-seq) using Promega Maxwell 16 MDx instrument, (Maxwell 16 LEV simplyRNA Tissue Kit (cat. # AS1280)) from WCM UTUC tumors. Specimens were prepared for RNAseq using TruSeq RNA Library Preparation Kit v2 or riboZero. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Sequencing was then performed on GAI, HiSeq 2000, or HiSeq 2500. All reads were independently aligned with STAR\_2.4.0f1<sup>38</sup> for sequence alignment against the human genome sequence build hg19, downloaded via the UCSC genome browser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/>), and SAMTOOLS v0.1.19<sup>39</sup> for sorting and indexing reads. Cufflinks (2.0.2)<sup>40</sup> was used to estimate the expression values (FPKM), and GENCODE v23<sup>41</sup> GTF file for annotation. Rstudio (1.0.136) with R (v3.3.2) and ggplot2 (2.2.1) were used for the statistical analysis and the generation of figures. For fusion analysis, we used STAR-fusion (STAR-Fusion\_v0.5.1)<sup>42,43</sup>. Fusions with significant support of junction reads and spanning pairs are then selected and manually reviewed. RNA was purified from BCM-MDA UTUC tumors and mRNA expression was computed for all genes from RNAseq data<sup>31</sup>. Gene fusions were detected in the RNAseq data using deFuse and SOAPfuse<sup>31</sup>.

**RNAseq data quantification, integration, and expression analysis.** The mRNA gene expression for high-grade WCM UTUC and TCGA UCB tumors was quantified as Reads Per Kilobase of transcript per Million mapped reads (RPKM).



Similarly, RPKMs for high-grade tumors from a previously published UTUC cohort at BCM and MDACC were calculated<sup>31</sup>. The RPKMs from these three institutions were combined and quantile normalized to reduce any batch effects among the samples while maintaining their individual biological variability. The quantile normalized data were log transformed for further analyses. To rule out batch effects, we examined the normalized expression values for a set of 40 housekeeping genes expected to have comparable expression value distributions among the three datasets (WCM UTUC, BCM-MDA UTUC, TCGA UCB) (Supplementary Fig. 1). Differential gene expression (DGE) between UTUC and UCB cohorts was performed on the counts data using the Bioconductor package DESeq2. The threshold to select differentially regulated genes was determined at a fold change of >1 for upregulated and <-1 for downregulated genes and results were deemed significant at an adjusted *p*-value of 0.05 (Benjamini-Hochberg correction).

**Hierarchical clustering to infer UTUC subtypes.** The subtypes in the UTUC cohorts, namely WCM and BCM-MDA, were inferred using previously published subtype-specific gene signatures together with previously reported<sup>2</sup> subtype classifications for TCGA UCB samples. To this end, first, the normalized expression values corresponding to the BASE47-signature, MDACC-signature, and TCGA-signature genes were extracted from the three integrated data sets (namely WCM UTUC, BCM-MDA UTUC, and TCGA UCB). For each signature gene set, the subtype for each of the WCM UTUC and BCM-MDA UTUC samples was inferred as follows: (1) The samples were clustered based on Pearson correlation and average linkage, and the cophenetic distances among them were calculated, (2) TCGA samples with minimum cophenetic distances to each UTUC sample were identified, and corresponding TCGA subtype labels were assigned to the specific UTUC samples. Subsequently, the results corresponding to each signature gene set were visualized in heatmaps, scaled across signature genes, and grouped based on inferred subtypes.

**Unsupervised NMF.** Unsupervised NMF was applied to the FPKM expression matrix for TCGA and UTUC samples. The input expression matrix was first filtered to retain only the top 25% of the genes with the highest expression variability. The optimal number of ranks was estimated to three based on 30 randomly initialized instances using the NMF R package. The NMF was then run with a rank  $k = 3$  over 100 iterations, to obtain the final deconvolution into three resulting components. Amongst other genes, component 1 contained ECM/EMT markers (C7, COMP, DES, PGM5, SFRP4, CLDN3, TWIST1), component 2 contained luminal/papillary markers (CIS.Down (CRTAC1, CTSE), luminal (FGFR3, KRT20, SNX31, UPK1A, UPK2), and Sonic Hedgehog (SHH)), and component 3 contained basal/squamous markers (basal (KRT5, KRT14, KRT6A), immune (CXCL11, SAA1), and squamous (DSC3, GSDMC, PI3, TGM1). PCA was then performed on the coefficients obtained from NMF to visualize the separation of the three components.

**Outlier analysis.** Genes expression outliers *Z*-scores were calculated for a list 74 cancer-related genes generated from the intersection of Sanger database and Drugbank (<https://www.drugbank.ca/> <https://www.sanger.ac.uk/science/tools/gdsc-genomics-drug-sensitivity-cancer>). *Z*-scores were calculated across the WCM UTUC and BCM-MDA UTUC cohorts for these 74 cancer-related genes. For each sample, the quantiles were calculated and then used to compute the lower and upper bound to define an outlier. A cut-off of *Z*-score > 1 and FPKMS > 50 was reported for the UTUC outlier genes after comparison with TCGA UCB samples.

**Identification of T-cell-inflamed and T-cell-depleted subtypes.** The top 5000 genes with the most variable normalized expression levels across TCGA UCB and UTUC samples (WCM and BCM-MDA) were selected based on their median absolute deviations. These genes data were then median centered and used as an input for hierarchical clustering and Euclidean distance (linkage = ward.D2,  $k = 20$ ). A 170-gene cluster containing CD8A and other key immune genes was identified. A *k*-means consensus clustering of these 170 genes across both UTUC and UCB cohorts revealed the presence of two prominent subclusters that we labeled as T-cell inflamed (with higher expression of cluster genes) and T-cell depleted (with lower expression of cluster genes).

**Differential expression analysis of FGFR3 shRNA dataset.** To study the role of FGFR3 in up-regulation of the interferon response, we obtained the publicly available Affymetrix microarray dataset from the RT-112 bladder cancer cell line, with or without shRNA-mediated knockdown of FGFR3<sup>17,18</sup>. The dataset comprised of 12 samples transduced with doxycycline-inducible shRNAs, which were either a shRNA-targeting EGFP (control) or one of three distinct shRNAs-targeting FGFR3 (FGFR3-shRNA); each condition had three biological replicates. This data was downloaded as raw signal intensity values for 54,675 probesets (Affymetrix Human Genome U133 Plus 2.0 Array). We used robust multiarray average (RMA) for background correction, normalization, and probe level intensity calculation from Affy Bioconductor Package (Version 1.52), in R statistical environment<sup>44</sup>. The normalized expression profiles were then used to

identify differentially expressed probes between FGFR3-shRNA versus control samples using the limma package (version 3.30.13)<sup>45</sup>. Probes were collapsed to gene level after taking the median fold change of the probes, utilizing hgu133-plus2 annotation data<sup>46</sup>. Genes that were differentially expressed after doxycycline induction in all three FGFR3-depleted cell lines but not in the control cell line were considered putative FGFR3-regulated genes. We identified 58 up-regulated genes (log-fold change > 1 and adjusted *P*-value < 0.05) and 45 downregulated genes. The log-fold change values of >1 or <-1 were used as thresholds to select for up-regulated or down-regulated genes, respectively (adjusted *P*-value < 0.05).

**Gene set enrichment analysis.** A pre-ranked gene set enrichment analysis (GSEA) was applied to the differentially expressed genes, ordered based on their log-fold change values, to identify the cellular pathways significantly altered after shRNA-mediated knockdown of FGFR3<sup>47</sup>. Gene sets available through the Gene Ontology Biological Pathways collection in the Molecular Signatures Database<sup>48</sup> were used for the GSEA analysis.

**Network analysis.** Gene sets found to have a statistically significant enrichment using GSEA were visualized using network-based enrichment maps in Cytoscape<sup>49</sup>, where each node in the network was representative of an individual gene set<sup>50</sup>. In addition, the enrichment map also grouped redundant gene sets into distinct clusters enabling the identification of broader functional categories. The clusters from the enrichment maps were further refined and labeled using AutoAnnotate<sup>51</sup>. We only focused on cancer-related and immune-related pathways in the network.

**Real-time PCR.** RT-4 and SW780 were purchased from ATCC (HTB-2 and CRL-2169) and RT-112 was purchased from Sigma (85061106). RT-4, SW780 and RT-112 were cultured in McCoy's 5A (modified) medium (Thermo Fisher Scientific, 16600108), DMEM (Thermo Fisher Scientific, 11965118), and EMEM (ATCC, 302003) with 10% FBS, respectively. All the cell lines were mycoplasma negative and validated by STR testing. RT-112, RT4, and SW780 cells were treated with DMSO, Erdafitinib 1  $\mu$ M, and Erdafitinib 5  $\mu$ M for 48 hours, respectively. Total RNA was isolated with the RNeasy Plus Mini Kit (Qiagen #74134) according to the manufacturer's protocol. Total RNA concentration was measured by NanoDrop (Thermo Fisher Scientific). cDNA was synthesized using SuperScript<sup>TM</sup> III First-Strand Synthesis system (Invitrogen #18080051). Real-time PCR was performed by LightCycler 480 (Roche) using Power SYBR Green Master Mix (Applied Biosystems #4367659). All calculations were collected and analyzed with LightCycler 480 software (Roche) using the delta-delta Ct method. The reaction components, conditions, and primers used are listed in Supplementary Table 4. All data were normalized to the expression of the housekeeping gene beta-actin ( $\beta$ -actin) and then compared to the expression in the DMSO-treated group. Two independent experiments were performed, each with two technical replicates. The data were presented by mean  $\pm$  SD. *P*-values were calculated using the *t*-test and corrected for multiple comparisons using the Holm-Sidak method. All analyses were performed using GraphPad Prism statistical software.

**Statistical analyses.** For statistical tests, two-sided Mann-Whitney-Wilcoxon test was used to check for significant differences between two distributions. The two-sided Fisher's exact test was applied to determine whether the deviations between the observed and the expected counts were significant. When appropriate, *P*-values were adjusted for multiple hypotheses testing with Benjamini-Hochberg procedure. Boxplot statistics were computed with the function "boxplot" of R programming language.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The genomic data that support the findings of this study are available in the database of Genotypes and Phenotypes (dbGaP) and on cBioPortal for Cancer Genomics with the identifier [https://www.cbioportal.org/study?id=utuc\\_cornell\\_baylor\\_mdacc\\_2019](https://www.cbioportal.org/study?id=utuc_cornell_baylor_mdacc_2019). The source data underlying Figs. 1a-d, 2a-e, 3a, b, and 4a-f and Supplementary Figs. 1, 2, 3 and 4 are provided as a Source Data file.

Received: 18 April 2018 Accepted: 4 June 2019  
Published online: 05 July 2019

## References

1. Rouprêt, M. et al. European Association of Urology guidelines on upper urinary tract urothelial carcinoma: 2017 update. *Eur. Urol.* **73**, 111-122 (2018).

2. Robertson, A.G. et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017).
3. Damrauer, J. S. et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl Acad. Sci. USA* **111**, 3110–3115 (2014).
4. Choi, W. et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* **25**, 152–165 (2014).
5. Cancer Genome Research Atlas Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
6. Sjö Dahl, G. et al. A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res.* **18**, 3377–3386 (2012).
7. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
8. Kim, J. et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet.* **48**, 600–606 (2016).
9. Harper, H. L. et al. Upper tract urothelial carcinomas: frequency of association with mismatch repair protein loss and lynch syndrome. *Mod. Pathol.* **30**, 146–156 (2017).
10. Iyer, G. et al. Mismatch repair (MMR) detection in urothelial carcinoma (UC) and correlation with immune checkpoint blockade (ICB) response. *J. Clin. Oncol.* **35**, 4511–4511 (2017).
11. Donahue, T.F. et al. Genomic characterization of upper-tract urothelial carcinoma in patients with Lynch syndrome. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00143> (2018).
12. Niu, B. et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–6 (2014).
13. Jia, Z. et al. Gene ranking of RNA-Seq data via discriminant non-negative matrix factorization. *PLoS One* **10**, e0137782 (2015).
14. Wellenstein, M. D. & de Visser, K. E. Cancer-cell-intrinsic mechanisms shaping the tumor immune landscape. *Immunity* **48**, 399–416 (2018).
15. Sweis, R. F. et al. Molecular drivers of the non-T-cell-Inflamed tumor microenvironment in urothelial bladder cancer. *Cancer Immunol. Res.* **4**, 563–568 (2016).
16. Sharma, P. et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* **18**, 312–322 (2017).
17. Cao, W., Ma, E., Zhou, L., Yuan, T. & Zhang, C. Exploring the FGFR3-related oncogenic mechanism in bladder cancer using bioinformatics strategy. *World J. Surg. Oncol.* **15**, 66 (2017).
18. Du, X. et al. FGFR3 stimulates stearyl CoA desaturase 1 activity to promote bladder tumor growth. *Cancer Res.* **72**, 5843–5855 (2012).
19. Williams, S. V. et al. Oncogenic FGFR3 gene fusions in bladder cancer. *Hum. Mol. Genet.* **22**, 795–803 (2013).
20. Rosenberg, J. E. et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909–1920 (2016).
21. Balar, A. V. et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* **389**, 67–76 (2017).
22. Catto, J. W., Xinarianos, G., Burton, J. L., Meuth, M. & Hamdy, F. C. Differential expression of hMLH1 and hMSH2 is related to bladder cancer grade, stage and prognosis but not microsatellite instability. *Int. J. Cancer* **105**, 484–490 (2003).
23. Shin, K. H. & Park, J. G. Microsatellite instability is associated with genetic alteration but not with low levels of expression of the human mismatch repair proteins hMSH2 and hMLH1. *Eur. J. Cancer* **36**, 925–931 (2000).
24. Audenet, F. et al. Clonal relatedness and mutational differences between upper tract and bladder urothelial carcinoma. *Clin. Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-18-2039> (2018).
25. Marzouka, N. A. et al. A validation and extended description of the Lund taxonomy for urothelial carcinoma using the TCGA cohort. *Sci. Rep.* **8**, 3737 (2018).
26. Tokarev, A. et al. Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS Res. Hum. Retrovir.* **25**, 1197–210 (2009).
27. Krejci, P. et al. Fibroblast growth factor inhibits interferon gamma-STAT1 and interleukin 6-STAT3 signaling in chondrocytes. *Cell. Signal.* **21**, 151–160 (2009).
28. Karkera, J. D. et al. Oncogenic characterization and pharmacologic sensitivity of activating fibroblast growth factor receptor (FGFR) genetic alterations to the selective FGFR inhibitor erdafitinib. *Mol. Cancer Ther.* **16**, 1717–1726 (2017).
29. Nogova, L. et al. Evaluation of BGJ398, a fibroblast growth factor receptor 1–3 kinase inhibitor, in patients with advanced solid tumors harboring genetic alterations in fibroblast growth factor receptors: results of a global phase I, dose-escalation and dose-expansion study. *J. Clin. Oncol.* **35**, 157–165 (2017).
30. FDA Drugs. *FDA Grants Accelerated Approval To Erdafitinib for Metastatic Urothelial Carcinoma.* <https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm635910.htm> (2019).
31. Moss, T. J. et al. Comprehensive genomic characterization of upper tract urothelial carcinoma. *Eur. Urol.* **72**, 641–649 (2017).
32. Humphrey, P. A., Moch, H., Cubilla, A. L., Ulbright, T. M. & Reuter, V. E. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: prostate and bladder tumours. *Eur. Urol.* **70**, 106–119 (2016).
33. Rennert, H. et al. Development and validation of a whole-exome sequencing test for simultaneous detection of point mutations, indels and copy-number alterations for precision cancer care. *NPJ Genom. Med.* **1**, 16019–16030 (2016).
34. Beltran, H. et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* **1**, 466–474 (2015).
35. Zhang, T. et al. Discovery and reporting of clinically-relevant germline variants in advanced cancer patients assessed using whole-exome sequencing. bioRxiv 112672, <https://doi.org/10.1101/112672> (2017).
36. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
37. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
38. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
39. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
41. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
42. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).
43. STAR-Fusion/STAR-Fusion, <https://github.com/STAR-Fusion/STAR-Fusion> (2016).
44. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
45. Diboun, I., Wernisch, L., Orengo, C. A. & Koltzenburg, M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genom.* **7**, 252 (2006).
46. Carlson, M. *hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2).* (R package version 3.2.3., 2016).
47. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
48. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
49. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
50. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
51. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: a Cytoscape app for summarizing networks with semantic annotations. *PLoS ONE* **5**, 1717 (2016).

## Acknowledgements

The work conducted at WCM was supported by the Conquer Cancer Foundation and the John and Elizabeth Leonard Family Foundation Young Investigator Award (B.M.F.). This work was also supported by a Charles, Lilian and Betty Newirth Foundation Fellowship in Oncology Award (P.J.V.), by the Translational Research Program at WCM Department of Pathology and Laboratory Medicine (B.D.R., J.M.M.) and by the Englander Institute for Precision Medicine at WCM (O.E., M.A.R.). The part of this work that was conducted at BCM-MDA was funded in part by the Robert J. Kleberg Jr. and Helen C. Kleberg Foundation, the Khalifa Bin Zayed Al Nahyan Foundation, and the Eleanor and Scott Petty Fund for UTUC Research, University of Texas MD Anderson Cancer Center. This work was also funded in part by the Partnership for Bladder Cancer Research, Scott Department of Urology, Dan L. Duncan Cancer Center Baylor College of Medicine.

## Author contributions

Initiation and design of the study: B.D.R., P.J.V., O.E. and B.M.F. Subject enrollment, sample, and clinical data collection: P.J.V., B.D.R., K.P., P.T., J.C., S.T.T., D.M.N., H.B., A.M.M., F.K., J.M.M., E.X., S.F.S., D.S.S., M.A.R., S.P.L., S.F.M. and B.M.F. Lab data

collection and analysis: W.L., K.L. and B.M.F. Statistical and bioinformatic analyses: P.J.V., B.B., R.B., T.J.M. and O.E. Supervision of research: O.E. and B.M.F. Writing of the first draft of the manuscript: B.D.R., P.J.V., O.E. and B.M.F. All authors contributed to the writing and editing of the revised manuscript and approved the manuscript.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-10873-y>.

**Competing interests:** B.D.R. has received consulting fees from BMS, S.T.T. received honoraria from Janssen. D.M.N. served on the data safety monitoring board for Genentech and Roche. H.B. has received research funding from Janssen, Abbvie/Stemcentryx, Astellas, Eli Lilly, Millennium and has served as advisor/consultant for Janssen, Astellas, Amgen, Astra Zeneca, Sanofi Genzyme. S.T.T., D.M.N., A.M.M., and B.M.F. has received research funding from Janssen for Weill Cornell Medicine for clinical trials. E.X. is a consultant for Janssen. S.F.S. has received honoraria, consulted or served on advisory boards for Astellas, Astra Zeneca, Bayer, BMS, Cepheid, Ferring, Ipsen, Janssen, Lilly, MSD, Olympus, Pfizer, Pierre Fabre, Richard Wolf Roche, Sanochemia, Sanofi, Takeda, Urogen. S.P.L. research funding for clinical trials from Endo, FKD, JBL (SWOG), Roche/Genentech (SWOG), UroGen, Viventia and is consultant for Archiano Therapeutics, UroGen, Vaxion. S.P.L. is on the advisory board of Archiano Therapeutics, Ferring, miR Scientific, QED Therapeutics, UroGen. S.F.M. is a consultant for QED Therapeutics. B.M.F. has received research support for Weill Cornell from Eli Lilly. The remaining authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Peer review information:** *Nature Communications* thanks John Sfakianos and other anonymous reviewer(s) for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019