iScience, Volume *23*

Supplemental Information

Urothelial-to-Neural Plasticity Drives Progression to Small Cell Bladder Cancer

Guoliang Yang, Jolanta Bondaruk, David Cogdell, Ziqiao Wang, Sangkyou Lee, June Goo Lee, Shizhen Zhang, Woonyoung Choi, Yan Wang, Yu Liang, Gang Wang, Ying Wang, Hui Yao, Vipulkumar Dadhania, Jianjun Gao, Christopher Logothetis, Arlene Siefker-Radtke, Ashish Kamat, Colin Dinney, Dan Theodorescu, Marek Kimmel, Peng Wei, Charles C. Guo, John N. Weinstein, David J. McConkey, and Bogdan Czerniak

SUPPLEMENTAL FIGURES AND LEGENDS

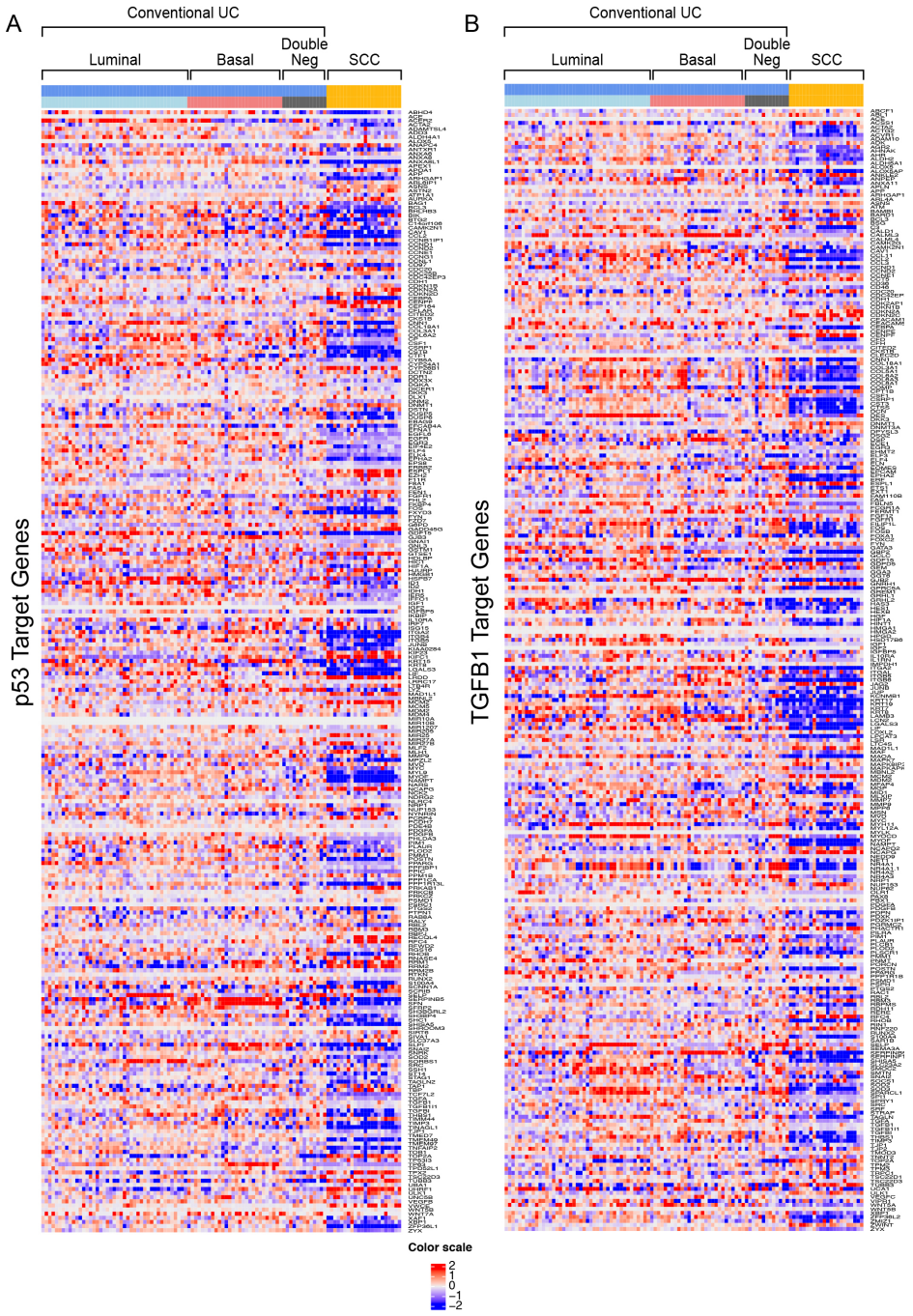

Figure S1: Expression patterns of TP53 and TGFB1 target genes in SCC and conventional UC, Related to Figure 2.

(A) Expression patterns of TP53 target genes in SCC and molecular subtypes of conventional UC. (B) Expression patterns of TGFB1 target genes in SCC and molecular subtypes of conventional UC.

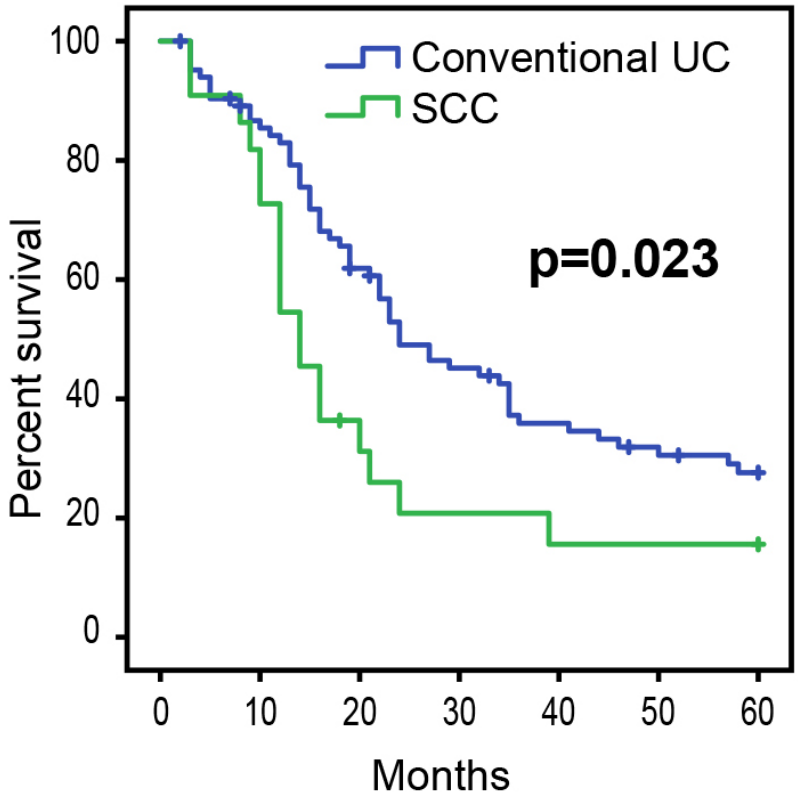

Figure S2: Survival analysis in SCC and conventional UC, Related to Figure 2.
Kaplan Meier analysis of survival in SCC and conventional UC and log-rank testing. A p-value <0.05 was considered statistically significant.

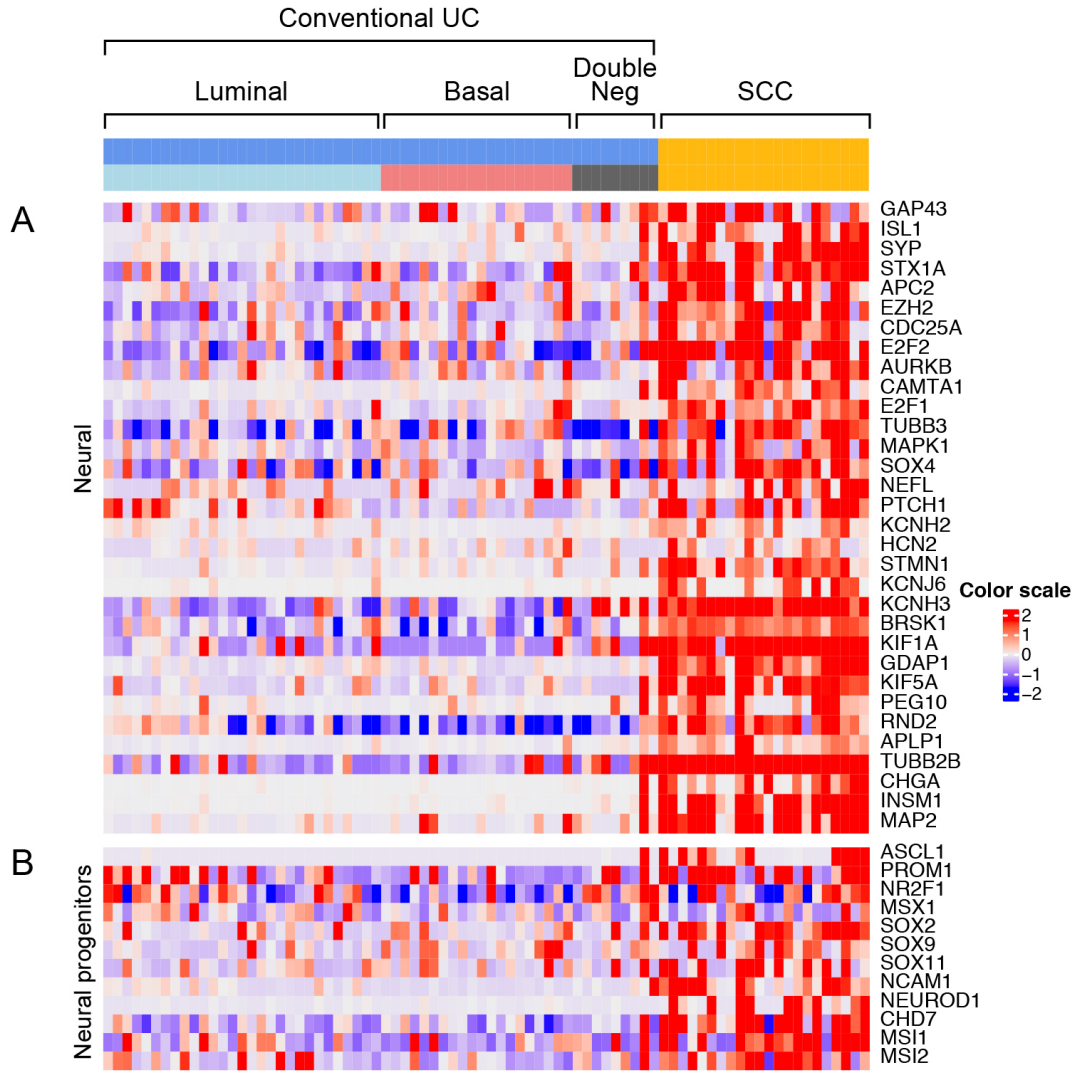

Figure S3: Expression patterns of genes involved in the regulation of neural phenotype, Related to Figure 4.

(A) Expression patterns of neural phenotype-related genes in SCC and molecular subtypes of conventional UC. (B) Expression patterns of transcription factors involved in neural progenitors development in SCC and molecular subtypes of conventional UC.

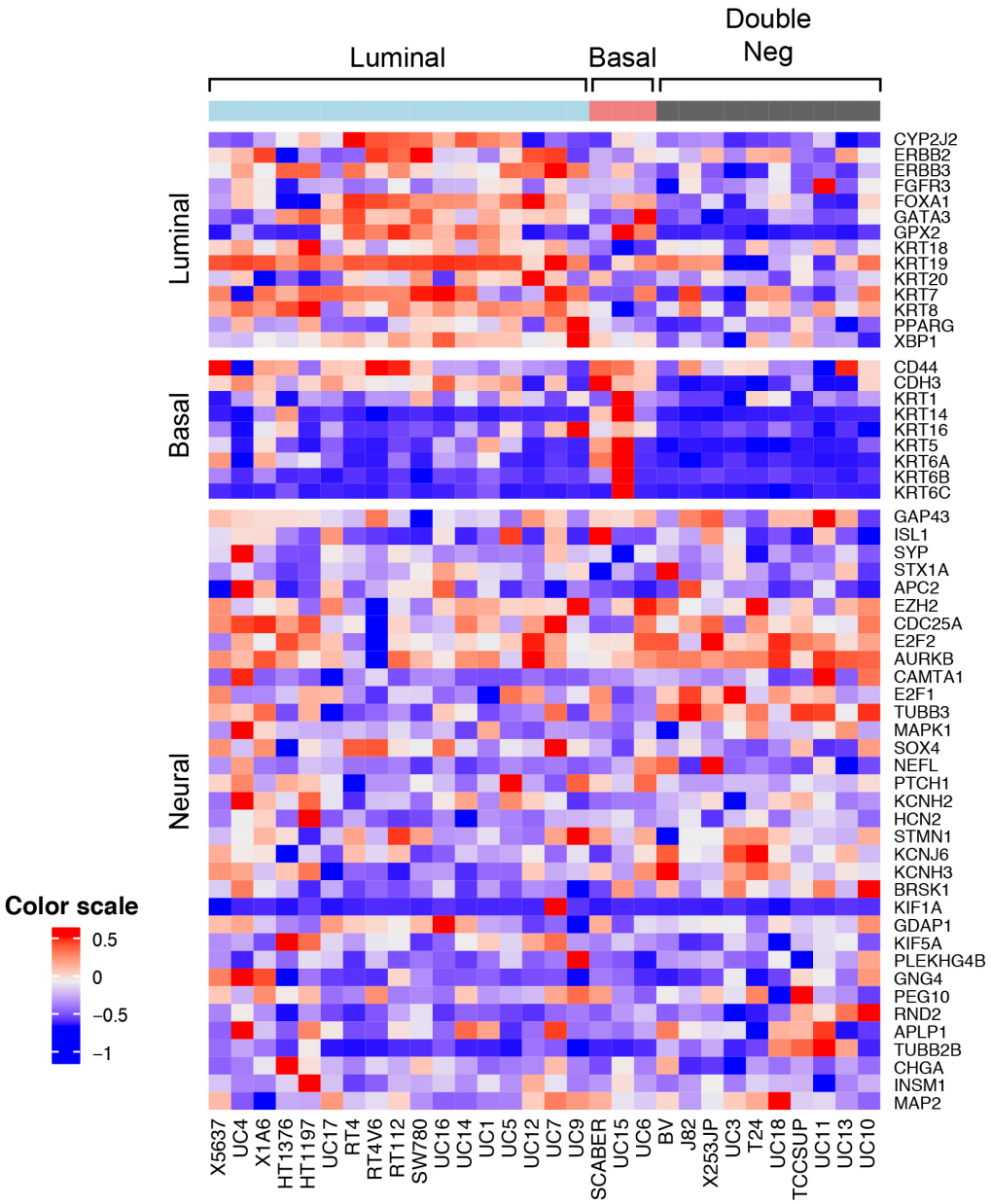

Figure S4. Molecular subtypes in 30 bladder cancer cell lines, Related to Figure 8.
Expression pattern of luminal, basal and neural markers in 30 bladder cancer cell lines revealed by RNASeq analysis.

Table S1: Summary of Clinical and Pathological Data for MD Anderson and TCGA Cohorts

MDACC Small Cell Carcinoma FFPE Cohort, Related to Figure 1.

Stage	Subtype	Gender F/M	Number of samples	Age, yr, mean ± SD
Invasive				
T1	Double negative	1/0	1	N/A
(T2 and higher)	Double negative	2/31	33	67.1 ± 10.8
Total		3/31	34	67.4 ± 10.8

CONVENTIONAL UROTHELIAL CARCINOMA COHORTS

MDACC FFPE Cohort

Invasive	Luminal	7/36	43	70.2 ± 11.7
(T2 and higher)	Basal	10/18	28	68.7 ± 10.9
	Double Negative	2/11	13	67.9 ± 8.8
Total		19/65	84	69.3 ± 11.0

MDACC Tissue Microarray FFPE Cohort

Invasive	Luminal	7/33	40	69.1 ± 11.5
(T2 and higher)	Basal	10/17	27	70 ± 10.3
	Double Negative	1/8	9	63.7 ± 6.1
Total		18/58	76	68.8± 10.6

TCGA Cohort

Invasive	Luminal	48/164	212	68.3 ± 11.0
(T2 and higher)	Basal	56/123	179	68.0 ± 10.1
	Double Negative	3/14	17	65.6 ± 10.2
Total		**107/301**	**408**	**68.1 ± 10.6**

F, female; M, male; yr, year; SD, standard deviation

Table S2: List of Mutations Identified in SCC (n=13) and Paired Samples of Conventional UC, Related to Figure 1.

	Paired Conventional UC/SCC				SCC										
	Conventional UC		SCC												
	1	2	1	2	3	4	5	6	7	8	9	10	11	12	13
	UC1	UC2	SCC1	SCC2	SCC3	SCC4	SCC5	SCC6	SCC7	SCC8	SCC9	SCC10	SCC11	SCC12	SCC13
TP53	p.H179Y	p.C176F	p.H179Y	p.C176F	p.E271K	p.E285K	p.N247I;p.A159fs	p.E286K	p.N210fs	p.E285K	p.R337C	p.R213*		p.P85fs	p.R248W
RB1	p.R798fs		p.R798fs				p.I369fs	p.Y325N; p.R579*;p.M708K			p.E398fs			p.R787*	p.S567L
FSCN3						p.E77Q			p.L259P						
PRAMEF1	p.P456L†											p.S462*			
BRD4											p.Q30E				Splice Site
ISLR2											p.D148N			p.225_226LP>LKRWP	
MAG								p.E178K			p.E572Q				
MAMDC2								p.FDSVLA35fs				p.D131Y			
TAF1D						p.E256*	p.E139del								
CIAPIN1											p.I6N			p.P164Q;p.K163K	
ARID1A					p.YP809fs						p.E1297*				
PKNOX1						p.G367A									
C16orf61															p.V28G
CDC23				p.S311fs											
PCDH18								p.Q744*							
NEUROG1						p.S121*									
SLC9A3R2							p.Q296*								
AR						p.Q488*									
TPX2											Splice Site				
CCDC158						p.S3*									
ADCY8						p.K703K									
LYZL4						p.Q132*									
MYST2						p.S216*									
TEF							p.D13E								
SERPINB12								p.DAPFCLN228fs							
RABEP1						Splice Site									
HOXB5					p.Y19*										
COG1						p.Q139*									
CRTAM						p.S326*									
PIGN								p.Y227*							
DLX2												p.Q195P			
RNF212							p.P147S								

*** Stop codon; fs, Frame shift; †, mutations present in the Conventional UC sample but not in the SCC paired sample**

TRANSPARENT METHODS

Clinical information and tissue samples

The use of human tissue samples for this study was approved by the Institutional Review Board of the UT MD Anderson Cancer Center. Thirty-four archival paraffin-embedded SCC and 84 invasive conventional bladder UC samples from the MD Anderson Cancer Center cohort were analyzed (Table S1). UCs were classified according to the histologic tumor grading system of the World Health Organization. Levels of invasion were defined according to the TNM staging system. All conventional UCs were invasive T2 and above high-grade tumors. The SCC and UC cohorts had similar age distributions with a male predominance. The mean age of the SCC cohort (31 men and 3 women) was 67 years (range, 34–90 years). The mean age of the conventional UC cohort (65 men and 19 women) was 69 years (range, 33–91 years). The median follow-up durations for the SCC and UC cohorts were 14 and 23 months, respectively. Sufficient high-quality DNA was available for 13 SCC cases and 2 paired SCC coexisting with conventional UC cases for whole-exome sequencing. Gene expression profiling was performed on 22 of the cases using Illumina's DASL platform and the data were merged with those obtained from a cohort of 84 conventional UCs. Quantitative RT-PCR was used to analyze the miRNA expression levels in 22 SCC samples and 80 conventional UC samples. To perform immunohistochemical validation studies of selected markers, we created a tissue microarray from the genomic profiling cohort comprising 14 of the cases of SCCs and 76 of the cases of conventional UC. Genomic, clinical, and pathological data from The Cancer Genome Atlas (TCGA) Bladder Cancer (BLCA) cohort of 408 muscle-invasive bladder cancers were used for reference.

DNA and RNA extraction

Genomic DNA and total RNA were extracted from FFPE tissue samples for DNA sequencing and microarray experiments using the MasterPure Complete DNA and RNA Purification Kit (Epicenter Biotechnologies, Madison, WI) according to the manufacturer's instructions as previously described(Guo et al., 2019). In brief, FFPE tissue cylinders were minced, deparaffinized, and digested with 300 μl Proteinase K digestion buffer with 10 μl Proteinase K (50ug/μl, Roche Diagnostics, Mannheim, Germany) at 55 °C overnight. DNA and RNA concentrations and quality were determined using the ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE) and the Quant-iT PicoGreen Kit (Life Technologies, Carlsbad, CA). Sufficient amounts of total RNA for gene expression analysis were extracted from all 47 SCC and 84 conventional UC samples. In addition, sufficient amount of genomic DNA were extracted from 13 cases of SCC, including 2 cases which also contained coexistent conventional precursor UC in the same specimens. DNA extracted from the peripheral blood lymphocytes or normal tissue of the resection specimen from the same patient was used as a control.

Whole-exome sequencing and processing pipeline

Genomic DNA from 13 cases of SCC and two cases of paired conventional UC were used for whole exome sequencing, which was performed on the HiSeq 2000 platform (Illumina, San Diego, CA, USA) at MD Anderson Cancer Center's Genomic Core. The TCGA data on 408 muscle-invasive conventional UC of the bladder were used as a reference set for mutational analyses. BWA-MEM (version 0.7.12) was used to align reads to the hg19 reference genome. Samtools (version 1.4) and Picard (version 2.5.0) were used to sort and convert between formats and remove duplicate reads(Etherington et al., 2015; Li et al., 2009). The Genome Analysis Toolkit (version3.4-46) was used to generate realigned and recalibrated BAM files(McKenna et al., 2010;

Van der Auwera et al., 2013). Somatic variants relative to the normal reference sample were detected by MuTect2(Cai et al., 2016; Callari et al., 2017). Oncotator (version 1.8.0.0) was used to produce gene-based and function-based annotations of the single nucleotide variants (SNVs) and insertions/deletions(Ramos et al., 2015). Similar analyses were performed for the genome-wide expression data from the TCGA cohort (n=408), and tumors were assigned to specific molecular subtypes by applying the sets of luminal and basal markers as described previously(Choi et al., 2014). Mutational data were downloaded from the TCGA portal (https://tcga-data.nci.nih.gov/tcga/). MutSigCV (version1.4; https://www.broadinstitute.org/cancer/cga/mutsig) and used to identify genes that were mutated more often than expected by chance given the background mutation processes(Lawrence et al., 2013). The significant gene list was obtained using a false discovery rate (FDR) cutoff 0.05. The statistical significance of associations between the mutations and the molecular subtypes was assessed by the Fisher's exact test.

Mutagenesis signatures

We used 432 SNVs identified in at least one sample and segregated them into six types of mutations corresponding to the following base pair substitutions: C>A, C>G, C>T, T>A, T>C, T>G. The Fischer's exact test was used to determine the distribution of these mutations in the three groups of samples corresponding to conventional UC in the TCGA cohort, paired UC and SCC cohort. The genomic context of SNVs, referred to as fingerprints which included the two flanking bases on 5' and 3' sides to each position for a total of 96 possible mutational fingerprints, was assembled. Wilcoxon Rank Sum tests were used to test the hypothesis of no difference in the frequency of any fingerprint between any two groups of mucosal samples. The Benjamini and Hochberg method was applied to control the FDR. For each sample, we used its mutational fingerprints (V) and the quadratic programming method to compute a weight score (H) for each of

30 canonical mutational signatures (*W*) available from the Sanger Institute (http://cancer.sanger.ac.uk/cosmic/signatures). We applied the 96-by-30 matrix of canonical signatures (W) and given the 96-by-1 mutational profile of a sample (V), we computed the 30-by-1 vector (H) for each of the canonical signatures' relative contributions to the sample profile by solving the following optimization formula:

$$\min_H (WH - V)^T(WH - V) \text{ such that } h_i \geq 0 \text{ and } \Sigma_i h_i = 1.$$

The optimization problems were solved using "quadprog" (version 1.5-5). Kruskal-Wallis test was used to test against the null hypothesis of no difference in weight scores among three groups of samples: conventional UC, paired UC and SCC.

mRNA expressions and data processing

RNAs from SCCs (n=22), and conventional UCs (n=84) were assessed using Illumina HumanHT-12 DASL Expression BeadChips as per the manufacturer's instructions, and Illumina BeadStudio v3.1.3 (Gene Expression Module V3.3.8) was used for transformation and normalization of the data. Comparisons were carried out using Welch's t-tests and Benjamini-Hochberg -controlled FDR-adjusted p values (<0.05) and fold changes. Unsupervised hierarchical clustering of log ratios was performed in R (Version 3.5.2), and the results were visualized with R package ComplexHeatmap. Pearson's correlation, mean centering, and average linkage were applied in all clustering applications. Genes within 0.5 standard deviations of the log-transformed ratios were discarded. To select specific and robust gene sets associated with SCC, we used the combination analysis with Welch's t-test and fold-change; genes having FDR-adjusted p-values <0.05 and showing fold-change >2.0 were selected. IPA software (Ingenuity Systems, Redwood City, CA) was used to determine dysregulated canonical pathways and predicted upstream

regulators by calculating z-scores and $-\log_{10}$ p values(Jimenez-Marin et al., 2009; Kramer et al., 2014). GSEA was used to evaluate the enrichment probability of the top canonical pathways and upstream regulators identified by IPA(Subramanian et al., 2005). Both SCC and UC samples were classified into luminal, and basal intrinsic molecular subtypes using an algorithm described previously.(Ochoa et al., 2016)

Immune gene expression signatures for SCC and conventional UC were established using unsupervised hierarchical clustering (Pan et al., 2019). Gene dendrogram nodes corresponding to genes characteristically expressed in specific immune cell types were identified and validated through DAVID functional annotation clustering and Ingenuity Systems Analysis (www.ingenuity.com). Immune gene signatures were used as previously reported (De Simone et al., 2016; Iglesia et al., 2014; Sherman et al., 2007; Torri et al., 2010). The immune expression signature was quantitatively assessed by calculating the immune scores for the expression profile of 128 genes shown in Figure 4. Specifically, the immune score for the *i*th sample was defined as m_i-(1/n) $\sum_{j=1}^{n} m_i$, where m_i is the median expression level across the *i*th sample's immune expression profile and (1/n) $\sum_{j=1}^{n} m_i$ is the grand mean of medians across all n samples. Additional analysis of immune infiltrate was performed by the CIBERSORT algorithm (http://cibersort.standford.edu/runcibersort.php). The expression profile of 547 genes using normalized mRNA levels with absolute mode and default parameters was used to assess the presence of 22 immune cell types in the conventional UC and SCCs.(Chen et al., 2018; Gentles et al., 2015) An empirical p-value was calculated using 500 permutations to test against the null hypothesis that no cell type is enriched in each sample. Then a Fisher's exact test was used to test against the null hypothesis of no association between sample types and their statistical significance.

To quantitatively assess the level of EMT (Garg and Singh, 2019; Mak et al., 2016), we calculated the EMT score based on a 76-gene expression signature reported in Byers et al. For each tumor sample, the score was calculated a weighted sum of 76 gene expression levels: $\sum_{i=1}^{76} w_i \, G_{ij}$, where w_i is the correlation coefficient between the *i*th gene expression in the signature and that of E-cadherin and G_{ij} is the *i*th gene's normalized expression in the *j*th tumor sample. We centered the scores by subtracting the mean across all tumor samples so that the grand mean of the score was zero.

miRNA analysis

miRNA analysis was performed on 22 SCC samples and 80 conventional UC samples (Dong et al., 2017). For miRNA cDNA synthesis, 400 ng of total RNA was reverse-transcribed using a miRNA reverse transcription kit (Applied Biosystems; catalogue No. 4366596) in combination with the stem-loop Megaplex primer pool (Applied Biosystems). For each cDNA sample, 381 small RNAs were profiled using TaqMan Human MicroRNA A Cards (Applied Biosystems; catalogue No. 4398965). Fold-change for each microRNA was determined using the ΔCt method and examined using Welch's t-test. An adjusted p value with FDR <0.05 was considered significant.

Regulon analysis

We performed regulon analysis to infer the relative activity of two sets of candidate transcription factors (TFs)(Castro et al., 2016). For a given TF, the set of its putative target genes is defined as a regulon. The first set of TFs were previously reported to be associated with bladder cancer and analyzed for the gene regulatory network: FOXA1, RXRA, FGFR3, RXRB, ERBB3, AR, GATA3, ESR2, ERBB2,PPARG, RARA, FGFR1, PGR, RARB, TP63, ESR1, GATA6,

STAT3, FOXM1, KLF4, EGFR, and HIF1A(Robertson et al., 2017). The second set of TFs were those reported to be related to neuronal phenotyping and manually curated based on the literature, including ASCL1, PROM1, NR2F1, MSX1, SOX2, SOX9, SOX11, NCAM1, NEUROD1, CHD7, MSI1, MSI2, E2F1, E2F2, and SOX4.

To construct the regulons for the TFs of interest, we employed the method developed and implemented in R package RTN1(Castro et al., 2016; Fletcher et al., 2013). In addition, we utilized the IPA to supplement the list of target genes of a TF if the number of target genes identified by the RTN method was too small to allow reliable downstream analyses.

We analyzed each regulon using the two-tailed GSEA(Castro et al., 2016; Subramanian et al., 2005). This method first divided the set of target genes for each TF into positive and negative targets associated with the phenotype of interest (for example, conventional UC vs SCC) using Spearman's correlation coefficient. The distribution of the positive and negative targets was tested respectively, producing enrichment scores (ES) for each sample. The difference between positive ES and negative ES resulted in the differential enrichment score (dES). For further analysis of the difference between the conventional UC and SCC, we extracted the dES for each sample. We assessed how the regulons were associated with bladder cancer subtypes by analyzing the dES with heatmaps and two-tailed GSEA.

Validation studies

Tissue microarrays, immunohistochemistry, and western blotting. The expression levels of selected genes were validated on parallel tissue microarrays (TMAs) comprising FFPE samples from 76 UCs and 14 SCCs. The TMAs were designed and prepared as described previously and profiled by genomic platforms(Sanfrancesco et al., 2016). In brief, TMAs (two 1-mm cores per

case) were constructed with a manual tissue arrayer (Beecher Instruments, Silver Spring, MD). Tissue sections from the TMAs were stained with hematoxylin and eosin to confirm the presence of tumor tissue. Immunohistochemical staining was performed with mouse monoclonal antibodies against human GATA3 (HG3-31 clone, 1:100 dilution; Santa Cruz Biotechnology Inc., Santa Cruz, CA), cytokeratin 5/6 (clone D5/16B4, 1:50 dilution, Dako, Carpinteria, CA), cytokeratin 14 (LL002 clone, 1:50 dilution; BioGenex, Fremont, CA), Synaptophysin, (clone 27G12, 1:600 dilution, Novocastra), Chromogranin (clone LK2H10, 1:600 dilution, Chemicon), and INSM1(clone A-8, 1:300 dilution, Santa Cruz Biotechnology Inc., Santa Cruz, CA). TMAs were also stained with antibodies specific for CD3 (rabbit polyclonal, 1:100 dilution, Dako, Carpinteria, CA), and CD8 Ab-1 (mouse clone C8/144B, 1:50 dilution, Thermo Scientific, Kalamazoo, MI). Immunostaining was performed using the Bond-Max Autostainer (Leica Biosystems, Buffalo Grove, IL). The staining intensities were scored by two pathologists (CCG and BAC) as negative and mildly, moderately, or strongly positive. In addition, the levels of Synaptophysin (D8F6H clone, 1:1000 dilution, Cell Signaling Technology), MSI1 (D46A8 clone, 1:1000 dilution, Cell Signaling Technology), E2F1 (Polyclone, 1:1000 dilution, Cell Signaling Technology), E-Cadherin (4A2 clone, 1:1000 dilution, Cell Signaling Technology), and ADORA2A (7F6-G5-A2 clone, 1:200 dilution, Santa Cruz Biotechnology Inc., Santa Cruz, CA) in SCC were confirmed on selected frozen tumor samples by western blotting.

Locked Nucleic Acid Ablation of miR-17-5p. We first analyzed genomic mRNA profile from 30 bladder cancer cell lines. In brief, total RNA was isolated using the mirVana miRNA isolation kit (Ambion, Inc). RNA purity and integrity were measured by NanoDrop ND-1000 and Agilent Bioanalyzer and only high-quality RNA was used for the cRNA amplification. Direct hybridization assays were performed using the Illumina RNA amplification kit (Ambion, Inc,

Austin, TX) and Illumina HT12 V4 chips (Illumina, Inc., San Diego, CA). Slides were scanned with Bead Station 500X and signal intensities were quantified with GenomeStudio (Illumina, Inc.). Quantile normalization in the Linear Models for Microarray Data (LIMMA) package in the R language environment was used to normalize the data. For the purpose of locked nucleic acid ablation of miR-17-5p experiment we focused our analysis on the expression pattern of luminal, basal, and neural markers. The basal urothelial carcinoma cell line UC6 was selected for the ablation of miR-17-5p. UC6 cells were grown to 70% confluency and were transfected using Lipofectamine 2000 (Thermo Fisher) in triplicates with 20 nM LNA for miR-17-5p (assay ID YI04100215-ADA) and negative control A-LNA (assay ID YI00199006-ADA) (Qiagen). After 24 hours, the cells were washed with PBS and RNA was isolated using Trizol (ThermoFisher). Gene expression profiles of the miR17-5p-LNA and negative control A-LNA in transfected and control UC6 cells were analyzed by RNASeq on an Illumina HiSeq2000. RNA was extracted as before then 76 bp libraries were constructed. mRNA expression profiles were mapped to the hg19 genome using MapSplice (de Matos Simoes et al., 2015). Gene expression was determined by RSEM and normalized with to the upper quartile.

For selected RNAs, expression levels after miR17-5p-LNA transfection were measured by quantitative PCR on an ABI 7900. For miRNA, each sample was reverse transcribed using TaqMan microRNA Reverse Transcription kit with the target specific primer. PCR was then performed to quantify miR-17-5p with miR-125a-3p as control and RNU48 as reference. To determine fold-changes in E2F1, RNA from miR17-5p-LNA and negative control A-LNA transfected UC6 cells were reverse transcribed. PCR was performed for E2F1 and human β actin was used as a reference (Applied Biosystems, Foster City, CA). Software defaults were used to

compute the relative change in expression by the ΔCt method. GraphPad 8.0.0 Prism was used to plot RQ and calculate p value using an unpaired t-test.

General statistical analysis

Survival analyses were performed by Kaplan–Meier analysis and log-rank testing. For genome-wide mRNA and miRNA differential expression analysis, the Benjamini and Hochberg(BH) method was applied to control the false discovery rate(FDR). An adjusted p value with FDR <0.05 was considered statistically significant.

References

Cai, L., Yuan, W., Zhang, Z., He, L., and Chou, K.C. (2016). In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. Sci Rep *6*, 36540.

Callari, M., Sammut, S.J., De Mattos-Arruda, L., Bruna, A., Rueda, O.M., Chin, S.F., and Caldas, C. (2017). Intersect-then-combine approach: improving the performance of somatic variant calling in whole exome sequencing data using multiple aligners and callers. Genome Med *9*, 35.

Castro, M.A., de Santiago, I., Campbell, T.M., Vaughn, C., Hickey, T.E., Ross, E., Tilley, W.D., Markowetz, F., Ponder, B.A., and Meyer, K.B. (2016). Regulators of genetic risk of breast cancer identified by integrative network analysis. Nat Genet *48*, 12-21.

Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., and Alizadeh, A.A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol *1711*, 243-259.

Choi, W., Czerniak, B., Ochoa, A., Su, X., Siefker-Radtke, A., Dinney, C., and McConkey, D.J. (2014). Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. Nat Rev Urol *11*, 400-410.

de Matos Simoes, R., Dalleau, S., Williamson, K.E., and Emmert-Streib, F. (2015). Urothelial cancer gene regulatory networks inferred from large-scale RNAseq, Bead and Oligo gene expression data. BMC Syst Biol *9*, 21.

De Simone, M., Arrigoni, A., Rossetti, G., Gruarin, P., Ranzani, V., Politano, C., Bonnal, R.J.P., Provasi, E., Sarnicola, M.L., Panzeri, I.*, et al.* (2016). Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. Immunity *45*, 1135-1147.

Dong, F., Xu, T., Shen, Y., Zhong, S., Chen, S., Ding, Q., and Shen, Z. (2017). Dysregulation of miRNAs in bladder cancer: altered expression with aberrant biogenesis procedure. Oncotarget *8*, 27547-27568.

Etherington, G.J., Ramirez-Gonzalez, R.H., and MacLean, D. (2015). bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. Bioinformatics *31*, 2565-2567.

Fletcher, M.N., Castro, M.A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S.F., Rueda, O.M., Caldas, C., Ponder, B.A., Markowetz, F.*, et al.* (2013). Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun *4*, 2464.

Garg, M., and Singh, R. (2019). Epithelial-to-mesenchymal transition: Event and core associates in bladder cancer. Front Biosci (Elite Ed) *11*, 150-165.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D.*, et al.* (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med *21*, 938-945.

Guo, C.C., Majewski, T., Zhang, L., Yao, H., Bondaruk, J., Wang, Y., Zhang, S., Wang, Z., Lee, J.G., Lee, S.*, et al.* (2019). Dysregulation of EMT Drives the Progression to Clinically Aggressive Sarcomatoid Bladder Cancer. Cell Rep *27*, 1781-1793 e1784.

Iglesia, M.D., Vincent, B.G., Parker, J.S., Hoadley, K.A., Carey, L.A., Perou, C.M., and Serody, J.S. (2014). Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. Clin Cancer Res *20*, 3818-3829.

Jimenez-Marin, A., Collado-Romero, M., Ramirez-Boo, M., Arce, C., and Garrido, J.J. (2009). Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. BMC Proc *3 Suppl 4*, S6.

Kramer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523-530.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A.*, et al.* (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214-218.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Mak, M.P., Tong, P., Diao, L., Cardnell, R.J., Gibbons, D.L., William, W.N., Skoulidis, F., Parra, E.R., Rodriguez-Canales, J., Wistuba, II*, et al.* (2016). A Patient-Derived, Pan-Cancer EMT Signature Identifies

Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition. Clin Cancer Res *22*, 609-620.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.
Ochoa, A.E., Choi, W., Su, X., Siefker-Radtke, A., Czerniak, B., Dinney, C., and McConkey, D.J. (2016). Specific micro-RNA expression patterns distinguish the basal and luminal subtypes of muscle-invasive bladder cancer. Oncotarget *7*, 80164-80174.
Pan, S., Zhan, Y., Chen, X., Wu, B., and Liu, B. (2019). Bladder Cancer Exhibiting High Immune Infiltration Shows the Lowest Response Rate to Immune Checkpoint Inhibitors. Front Oncol *9*, 1101.
Ramos, A.H., Lichtenstein, L., Gupta, M., Lawrence, M.S., Pugh, T.J., Saksena, G., Meyerson, M., and Getz, G. (2015). Oncotator: cancer variant annotation tool. Hum Mutat *36*, E2423-2429.
Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue, T., Laird, P.W., Hoadley, K.A., Akbani, R.*, et al.* (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell *171*, 540-556 e525.
Sanfrancesco, J., McKenney, J.K., Leivo, M.Z., Gupta, S., Elson, P., and Hansel, D.E. (2016). Sarcomatoid Urothelial Carcinoma of the Bladder: Analysis of 28 Cases With Emphasis on Clinicopathologic Features and Markers of Epithelial-to-Mesenchymal Transition. Arch Pathol Lab Med *140*, 543-551.
Sherman, B.T., Huang da, W., Tan, Q., Guo, Y., Bour, S., Liu, D., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics *8*, 426.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Torri, A., Beretta, O., Ranghetti, A., Granucci, F., Ricciardi-Castagnoli, P., and Foti, M. (2010). Gene expression profiles identify inflammatory signatures in dendritic cells. PLoS One *5*, e9404.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J.*, et al.* (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics *43*, 11 10 11-33.