Supplemental Data

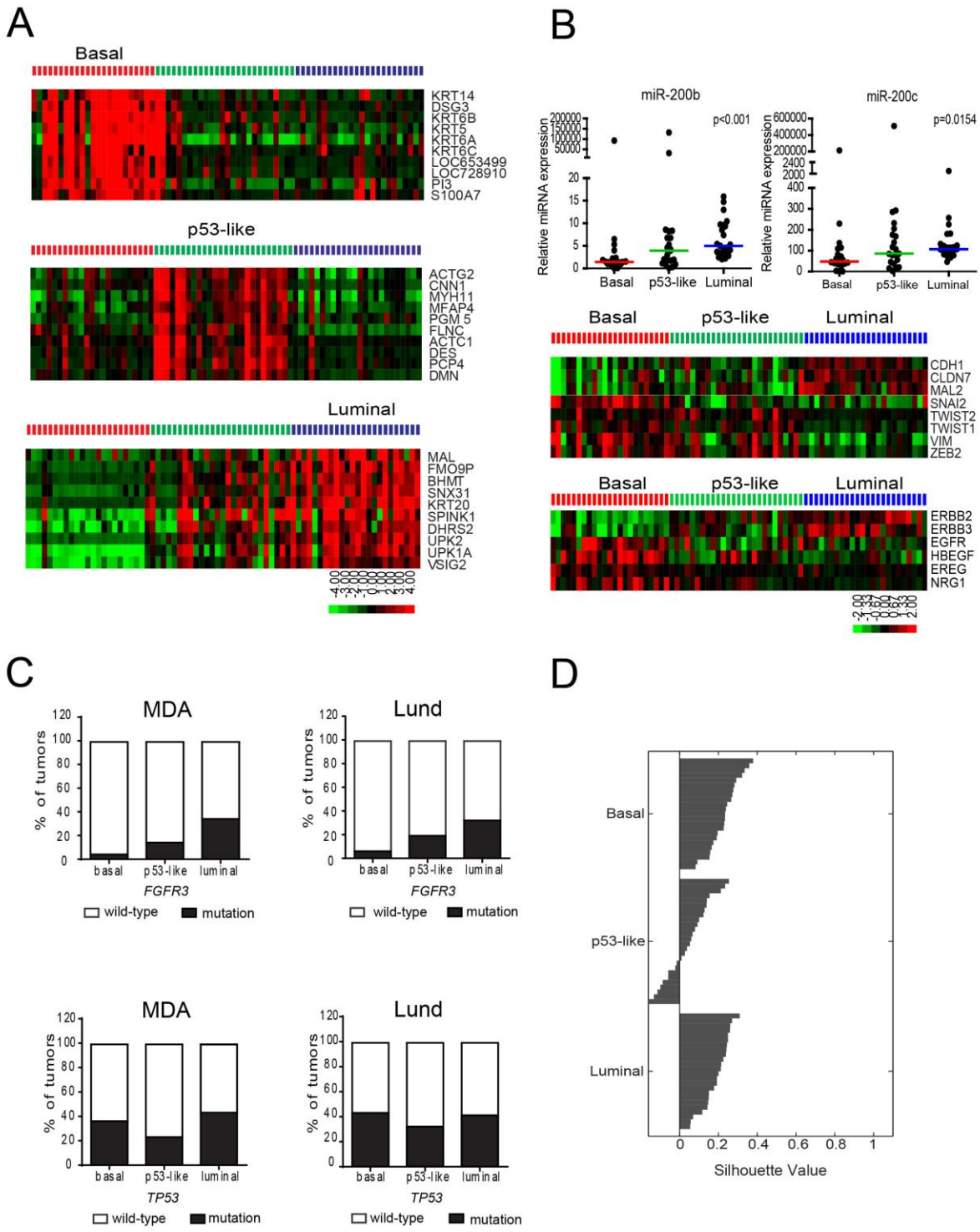


Figure S1. Related to Figure 1.

Biomarkers in the basal and luminal subtypes. A. Top 10 upregulated probes based on fold changes in each subtype ($p<0.001$, FDR<0.1). Heat maps display the top upregualted genes in the basal (top), p53-like (middle), and luminal (bottom) subtypes in the MD Anderson discovery cohort. B. Relative expression of EMT markers and components the EGFR pathway in the 3 subtypes. miR-200b and miR-200c levels were measured by quantitative RT-PCR. The line shows the median expression of miR-200b/c in each subtype. The heat maps display EMT marker (top) and EGFR pathway (bottom) gene expression as a function of subtype in the MD Anderson discovery cohort. C. *FGFR3* and *TP53* mutation distributions in the subtypes. D. Silhouette scores were calculated to determine stability of tumor classification and are displayed as a function of subtype membership.

Table S1. Related to Figure 2.

Clinicopathologic Characteristics of the MDA validation Cohort (n=57)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size	57	13 (23%)	25 (44%)	19 (33%)	
Mean Age (y) ± SD	65.7 ± 10.8	63.3 ± 10.1	66.8 ± 8.6	66.0 ± 13.7	0.649
Gender					
Male	49 (86%)	10 (77%)	22 (88%)	17 (90%)	0.560
Female	8 (14%)	3 (23%)	3 (12%)	2 (11%)	
Race					
Caucasian	48 (84%)	11 (85%)	23 (92%)	15 (74%)	0.554
African American	3 (5%)	1 (8%)	0 (0%)	2 (11%)	
Hispanic	4 (7%)	1 (8%)	1 (4%)	2 (11%)	
Asian	1 (2%)	0 (0%)	1 (4%)	0 (0%)	
Clinical Stage at TUR (N0,M0)					
≤ cT1	11 (19%)	0 (0%)	6 (24%)	5 (26%)	0.178
cT2	36 (63%)	9 (69%)	16 (64%)	11 (58%)	
cT3	7 (12%)	3 (23%)	1 (4%)	3 (16%)	
cT4	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Positive Clinical Lymph Nodes, cN+	1 (2%)	0 (0%)	1 (4%)	0 (0%)	0.521
Positive Clinical Metastasis, cM+	2 (94%)	1 (8%)	1 (4%)	0 (0%)	0.501
Neoadjuvant Chemotherapy	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Pathologic T stage					
pT0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.454
pTa, pT1, pTis	2 (4%)	0 (0%)	1 (4%)	1 (5%)	
pT2	13 (23%)	1 (8%)	9 (36%)	3 (16%)	
pT3	33 (58%)	9 (69%)	12 (48%)	12 (63%)	
pT4	9 (16%)	3 (23%)	3 (12%)	3 (16%)	
Positive Pathologic Lymph Nodes	36 (63%)	9 (69%)	15 (60%)	12 (63%)	0.855
Variant histology at cystectomy					
Squamous Differentiation	6 (11%)	5 (39%)	1 (4%)	0 (0%)	**0.007**
Sarcomatoid Differentiation	3 (5%)	2 (15%)	0 (0%)	1 (5%)	
Other (Micropapillary, Glandular, Adenocarcinoma)	8 (14%)	1 (8%)	3 (12%)	4 (21%)	
Median Overall Survival (m)	79.2	25.0	105.9	79.2	**0.011**
Median Disease Specific Survival (m)	Not Reached	25.3	Not Reached	79.2	**0.004**

The Kruskal-Wallis test was used to compare differences in mean age between groups. Log-rank test was used to compare differences in survival (overall and disease specific) between groups. For the remainder of categorical variables, Fisher's exact test was used to determine differences between subtypes. p-values <0.05 were considered significant.

Table S2. Related to Figure 2.

Clinicopathologic Characteristics of the Chungbuk Cohort (n=55)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size	55	11 (20%)	23 (42%)	21 (38%)	
Mean Age (y) ± SD	67.3 ± 10.1	69.6 ± 8.4	61.5 ± 10.5	72.5 ± 7.1	**0.001**
Gender					
Male	42 (76%)	6 (55%)	21 (91%)	15 (71%)	**0.049**
Female	13 (24%)	5 (46%)	2 (9%)	6 (29%)	
Clinical T Stage (N0,M0)					
≤ cT1	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
cT2	22 (40%)	3 (27%)	10 (44%)	9 (43%)	0.906
cT3	11 (20%)	3 (27%)	4 (17%)	4 (19%)	
cT4	11 (20%)	1 (9%)	3 (13%)	7 (13%)	
Positive Clinical Lymph Nodes - cN+	15 (27%)	4 (36%)	6 (26%)	5 (24%)	0.740
Positive Clinical Metastasis - cM+	7 (13%)	1 (9%)	4 (17%)	2 (10%)	0.679
Systemic Chemotherapy	25 (46%)	5 (46%)	13 (57%)	7 (33%)	0.304
Median Overall Survival (m)	17.1	10.4	26.4	Not Reached	0.058
Median Disease Specific Survival (m)	26.4	11.2	66.3	Not Reached	0.102

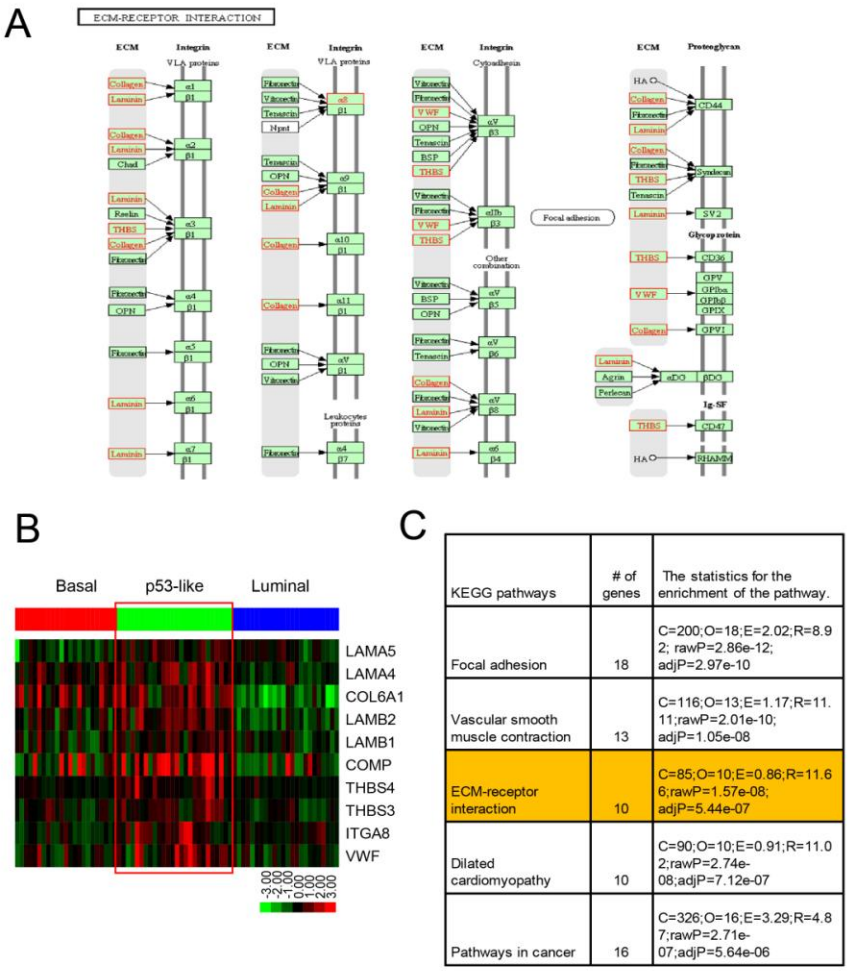
The Kruskal-Wallis test was used to compare differences in mean age between groups. Log-rank test was used to compare differences in survival (overall and disease specific) between groups. For the remainder of categorical variables, Fisher's exact test was used to determine differences between subtypes. p-values <0.05 were considered significant.

Table S3. Related to Figure 3.

Clinicopathologic Characteristics of the Lund Cohort (n=93)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size (n)	93	30 (32%)	32 (35%)	31 (33%)	
Mean Age (y) ± SD	71.8 ± 11.6	76.2 ± 11.0	71.7 ± 11.7	67.6 ± 10.9	**0.012**
Gender (n)					
Male	68 (73%)	17 (57%)	27 (84%)	24 (77%)	**0.039**
Female	25 (27%)	13 (43%)	5 (16%)	7 (23%)	
Clinical Stage at TUR (n)					
cT2	85 (91%)	28 (93%)	30 (94%)	27 (87%)	0.666
cT3	7 (8%)	2 (7%)	2 (6%)	3 (10%)	
cT4	1 (1%)	0 (0%)	0 (0%)	1 (3%)	
Primary Treatment (n)					
Cystectomy	51 (55%)	12 (40%)	18 (56%)	21 (68%)	0.092
Other	42 (45%)	18 (60%)	14 (44%)	10 (32%)	
Positive Pathologic Lymph Nodes (n)	20 (39%)	4 (33%)	9 (50%)	7 (33%)	0.507
Variant histology (squamous)*	17 (18%)	16 (53%)	1 (3%)	0 (0%)	**<0.001**
Median Disease Specific Survival (m)	Not Reached	34.8	36.4	Not Reached	0.860

*squamous differentiation was noted specifically in 17 tumors based on publically available data. The Kruskal-Wallis test was used to compare differences in mean age between groups. Log-rank test was used to compare differences in survival between groups. For the remainder of categorical variables, Fisher's exact test was used to determine differences between subtypes. p-values <0.05 were considered significant



C

KEGG pathways	# of genes	The statistics for the enrichment of the pathway.
Focal adhesion	18	C=200;O=18;E=2.02;R=8.92; rawP=2.86e-12; adjP=2.97e-10
Vascular smooth muscle contraction	13	C=116;O=13;E=1.17;R=11.11;rawP=2.01e-10; adjP=1.05e-08
ECM-receptor interaction	10	C=85;O=10;E=0.86;R=11.66;rawP=1.57e-08; adjP=5.44e-07
Dilated cardiomyopathy	10	C=90;O=10;E=0.91;R=11.02;rawP=2.74e-08;adjP=7.12e-07
Pathways in cancer	16	C=326;O=16;E=3.29;R=4.87;rawP=2.71e-07;adjP=5.64e-06

Figure S2. Related to Figure 3.

Enrichment of ECM-receptor and ECM biomarkers indicative of stromal fibroblast infiltration in the p53-like tumors. A. ECM-receptor interaction in the p53-like tumors as determined using the mRNA expression profiling data from the MD Anderson discovery cohort and KEGG pathway analysis. B. Significantly differentially expressed genes in ECM receptor interaction pathway. C. Top five significant KEGG pathways as determined using the significantly upregulated genes in the p53-like tumors. The analysis was performed using WebGestalt (http://genereg.ornl.gov/webgestalt). The statistics column lists the number of reference genes in the category (C), the number of genes in the gene set that are also in the category (O), the expected number in the category (E), the ratio of enrichment (R), the p value from a hypergeometric test (rawP), and the p value adjusted by the multiple test adjustment (adjP).

Table S4. Related to Figure 3.

Clinicopathologic Characteristics of the UCSF Cohort (n=53)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size	53	15 (28%)	14 (27%)	24 (45%)	
Mean Age (y) ± SD	64.7 ± 12.1	67.5 ± 12.4	61.2 ± 14.2	65.0 ± 10.6	0.407
Gender					
Male	37 (70%)	10 (67%)	9 (64%)	18 (75%)	0.748
Female	16 (30%)	5 (33%)	5 (36%)	6 (25%)	
Pathologic T stage (n)					
pT2	14 (26%)	5 (33%)	3 (22%)	6 (25%)	
pT3	26 (49%)	4 (27%)	9 (64%)	13 (54%)	0.293
pT4	13 (25%)	6 (40%)	2 (14%)	5 (21%)	
Positive Pathologic Lymph Nodes (n)	17 (32%)	3 (20%)	7 (50%)	7 (29%)	0.403
Variant histology (squamous)*	6 (11%)	4 (27%)	2 (14%)	0 (0%)	**0.035**
Median Overall Survival (m)	17.7	11.3	15.5	19.5	0.880

*Squamous differentiation. The Kruskal-Wallis test was used to compare differences in mean age between groups. Log-rank test was used to compare differences in survival between groups. For the remainder of categorical variables, Fisher's exact test was used to determine differences between subtypes. p-values <0.05 were considered significant.

Table S5. Related to Figure 3.

Clinicopathologic Characteristics of the TMA Cohort

	TOTAL	TCC	TCC with squamous differentiation	p-value
Cohort Size	332	267	44	
Mean Age (y) ± SD	65.9 ± 10.2	65.8 ± 10.3	66.9 ± 9.9	0.478
Gender				
Female	75 (23%)	66 (23%)	9 (21%)	0.863
Male	256 (77%)	221 (77%)	35 (80%)	
Preoperative Chemotherapy				
Yes	3 (1%)	2 (1%)	1 (2%)	
No	327 (99%)	284 (99%)	43 (98%)	0.506
Unknown	3 (1%)	3 (1%)	0 (0%)	
Adjuvant Chemotherapy				
Yes	77 (23%)	65 (27%)	12 (23%)	
No	215 (65%)	192 (67%)	23 (52%)	0.102
Unknown	40 (12%)	31 (11%)	9 (21%)	
Pathologic T stage				
pT3 (not specified)	62 (19%)	57 (20%)	5 (11%)	
pT3a	158 (48%)	137 (48%)	21 (48%)	0.329
pT3b	112 (34%)	94 (33%)	18 (41%)	
Pathologic N stage				
pN0	244 (74%)	208 (72%)	36 (82%)	
pN+	81 (24%)	74 (26%)	7 (16%)	0.371
pNx	7 (2%)	6 (2%)	1 (2%)	
Pathologic M stage				
pM0	330 (99%)	286 (99%)	44 (100%)	0.794
pM1	2 (1%)	2 (1%)	0 (0%)	
KRT 5/6 % positive staining (± SEM)	25.1 ± 1.7	20.9 ± 1.6	52.7 ± 5.4	**<0.0001**
KRT 20 % positive staining (± SEM)	2.7 ± 0.5	3.0 ± 0.5	0.10 ± 0.06	

*SD= Standard Deviation, SEM= Standard effect of the mean

Table S6. Related to Figure 4.

Predicted upstream regulators within the 3 subtypes (TF*)

	Predicted Activation State :Activated			Predicted Activation State :Inhibited		
	Upstream Regulator	Activation z-score	p-value of overlap	Upstream Regulator	Activation z-score	p-value of overlap
Basal	STAT3	4.832	6.66E-18	ER*	-3.646	1.18E-11
	NFκB(complex)	6.837	9.35E-15	TRIM24	-4.000	7.28E-09
	IRF7	5.543	1.75E-10	PPARA	-2.815	3.28E-05
	JUN	2.295	5.99E-10	Hdac	-2.088	5.97E-05
	STAT1	4.396	7.46E-10	GATA3	-2.566	1.49E-04
	SP1	2.227	1.39E-09	N-cor	-2.449	4.28E-04
	TP63	3.434	1.95E-08	PIAS4	-2.000	2.57E-03
	RELA	2.793	2.23E-08	KLF2	-2.366	3.48E-03
	HIF1A	3.606	4.92E-07	SPDEF	-2.931	4.92E-03
	IRF3	2.82	5.77E-07	MEOX2	-2.646	1.54E-02
P53-like	TP53	4.814	9.08E-17	TBX2	-4.690	1.92E-13
	CDKN2A	4.748	3.78E-12	FOXM1	-2.797	4.04E-10
	RB1	2.071	5.70E-09	MYC	-4.208	8.37E-06
	MYOCD	3.366	9.94E-09	SMAD7	-2.704	8.55E-05
	MKL1	2.956	7.52E-08	E2F2	-2.236	4.50E-04
	TCF3	3.889	1.14E-07	MYCN	-2.779	5.42E-04
	SMARCB1	3.637	3.75E-06	AHR	-2.850	8.86E-04
	SRF	3.847	5.29E-06	HEY2	-2.168	9.36E-04
	HTT	2.333	2.30E-05	NFE2L2	-2.707	4.29E-02
	Rb	2.425	1.80E-03	SPDEF	-2.236	1.14E-01
Luminal	AHR	2.540	3.65E-12	TP53	-3.296	2.27E-15
	ER**	5.505	9.02E-12	STAT3	-4.084	3.15E-14
	MYC	3.710	1.10E-10	SMARCA4	-2.218	1.46E-11
	SPDEF	3.615	1.19E-08	PGR	-2.175	2.35E-10
	Hdac	2.089	9.77E-08	NFκB(complex)	-5.342	3.03E-10
	SMAD7	3.504	2.40E-07	STAT1	-2.414	7.34E-10
	PPARA	3.246	7.64E-05	HTT	-2.983	1.70E-08
	TRIM24	3.742	5.93E-04	SMAD3	-3.870	5.92E-08
	PPARG	2.768	1.08E-03	SRF	-4.105	7.32E-08
	SREBF2	3.255	6.12E-03	MKL1	-2.960	3.79E-07

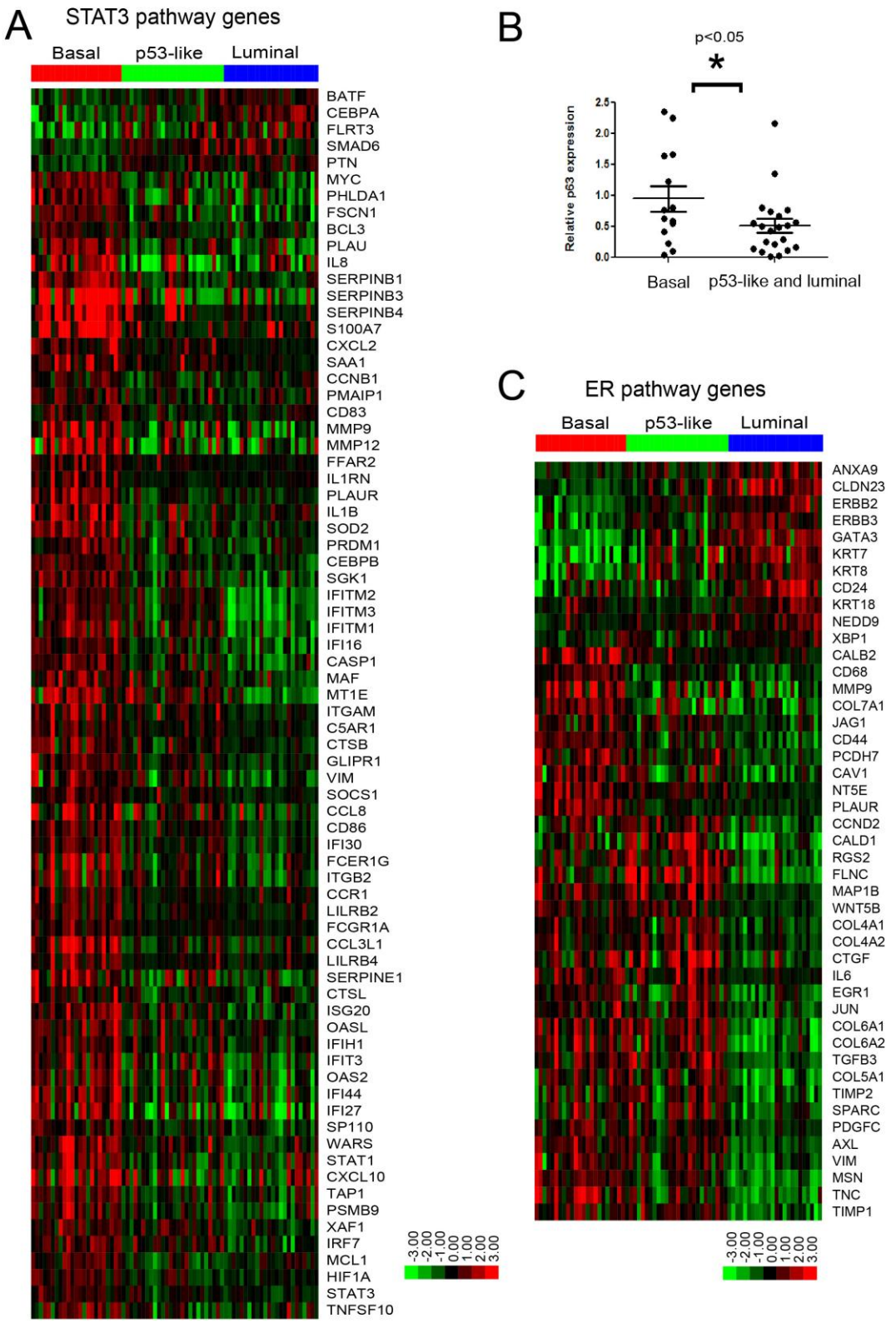
*TF: transcriptional factors in Molecule Type **ER: Estrogen Receptor

Table S7. Related to Figure 4.

Predicted upstream regulators in the p53-like subtype for 73 or 64 tumors (TF*)

		73 tumors		64 tumors	
Upstream Regulator	Predicted Activation State	Activation z-score	p-value of overlap	Activation z-score	p-value of overlap
TP53	Activated	4.814	9.08E-17	5.185	6.13E-26
CDKN2A	Activated	4.748	3.78E-12	3.842	1.53E-12
RB1	Activated	2.071	5.70E-09	2.813	1.42E-09
MYOCD	Activated	3.366	9.94E-09	3.492	5.61E-08
MKL1	Activated	2.956	7.52E-08	2.411	1.31E-04
TCF3	Activated	3.889	1.14E-07	4.455	1.30E-08
SMARCB1	Activated	3.637	3.75E-06	4.469	2.10E-08
SRF	Activated	3.847	5.29E-06	3.867	2.87E-08
Rb	Activated	2.425	1.80E-03	2.97	1.25E-04
TBX2	Inhibited	-4.69	1.92E-13	-5.000	5.15E-13
FOXM1	Inhibited	-2.797	4.04E-10	-3.114	1.73E-11
MYC	Inhibited	-4.208	8.37E-06	-5.014	9.82E-13
SMAD7	Inhibited	-2.704	8.55E-05	-3.307	9.83E-06
E2F2	Inhibited	-2.236	4.50E-04	-1.89	3.80E-04
MYCN	Inhibited	-2.779	5.42E-04	-3.966	6.06E-06
HEY2	Inhibited	-2.168	9.36E-04	-1.939	2.48E-02
NFE2L2	Inhibited	-2.707	4.29E-02	-0.343	1.13E-03
AHR	Inhibited	-2.85	8.86E-04	-3.355	1.64E-05
SPDEF	Inhibited	-2.236	1.14E-01	-3.162	4.25E-03

*TF: transcriptional factors in Molecule Type


A
STAT3 pathway genes
Basal
p53-like
Luminal
BATF
CEBPA
FLRT3
SMAD6
PTN
MYC
PHLDA1
FSCN1
BCL3
PLAU
IL8
SERPINB1
SERPINB3
SERPINB4
S100A7
CXCL2
SAA1
CCNB1
PMAIP1
CD83
MMP9
MMP12
FFAR2
IL1RN
PLAUR
IL1B
SOD2
PRDM1
CEBPB
SGK1
IFITM2
IFITM3
IFITM1
IFI16
CASP1
MAF
MT1E
ITGAM
C5AR1
CTSB
GLIPR1
VIM
SOCS1
CCL8
CD86
IFI30
FCER1G
ITGB2
CCR1
LILRB2
FCGR1A
CCL3L1
LILRB4
SERPINE1
CTSL
ISG20
OASL
IFIH1
IFIT3
OAS2
IFI44
IFI27
SP110
WARS
STAT1
CXCL10
TAP1
PSMB9
XAF1
IRF7
MCL1
HIF1A
STAT3
TNFSF10
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
B
p<0.05
*
Relative p63 expression
2.5
2.0
1.5
1.0
0.5
0.0
Basal
p53-like and luminal
C
ER pathway genes
Basal
p53-like
Luminal
ANXA9
CLDN23
ERBB2
ERBB3
GATA3
KRT7
KRT8
CD24
KRT18
NEDD9
XBP1
CALB2
CD68
MMP9
COL7A1
JAG1
CD44
PCDH7
CAV1
NT5E
PLAUR
CCND2
CALD1
RGS2
FLNC
MAP1B
WNT5B
COL4A1
COL4A2
CTGF
IL6
EGR1
JUN
COL6A1
COL6A2
TGFB3
COL5A1
TIMP2
SPARC
PDGFC
AXL
VIM
MSN
TNC
TIMP1
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00

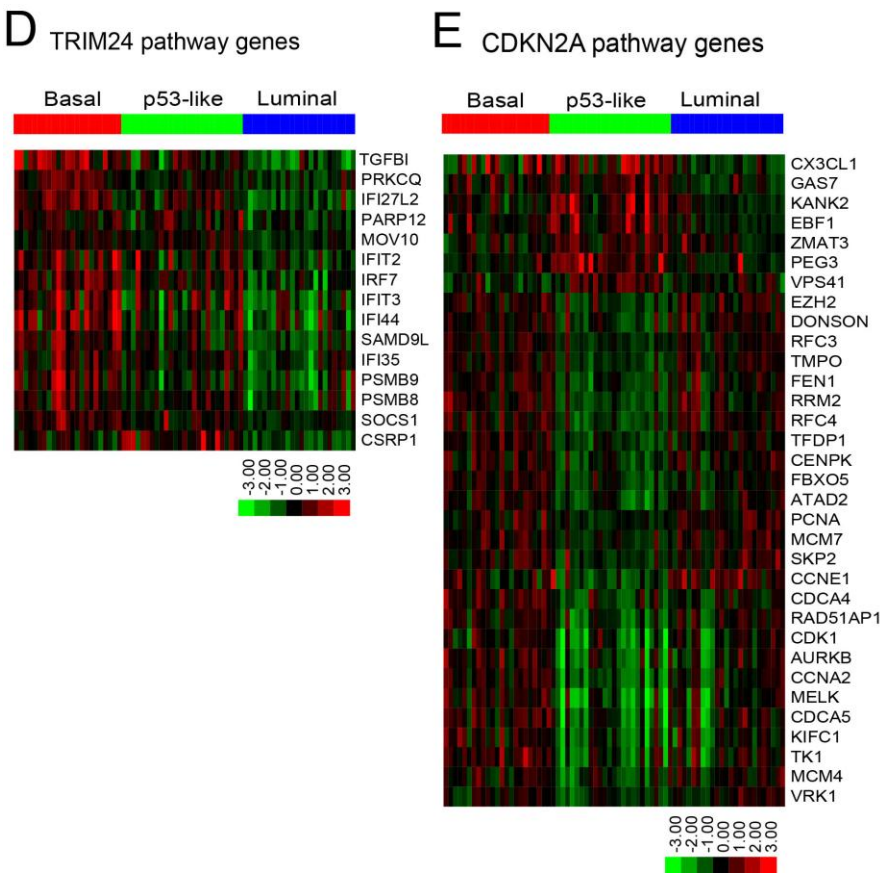
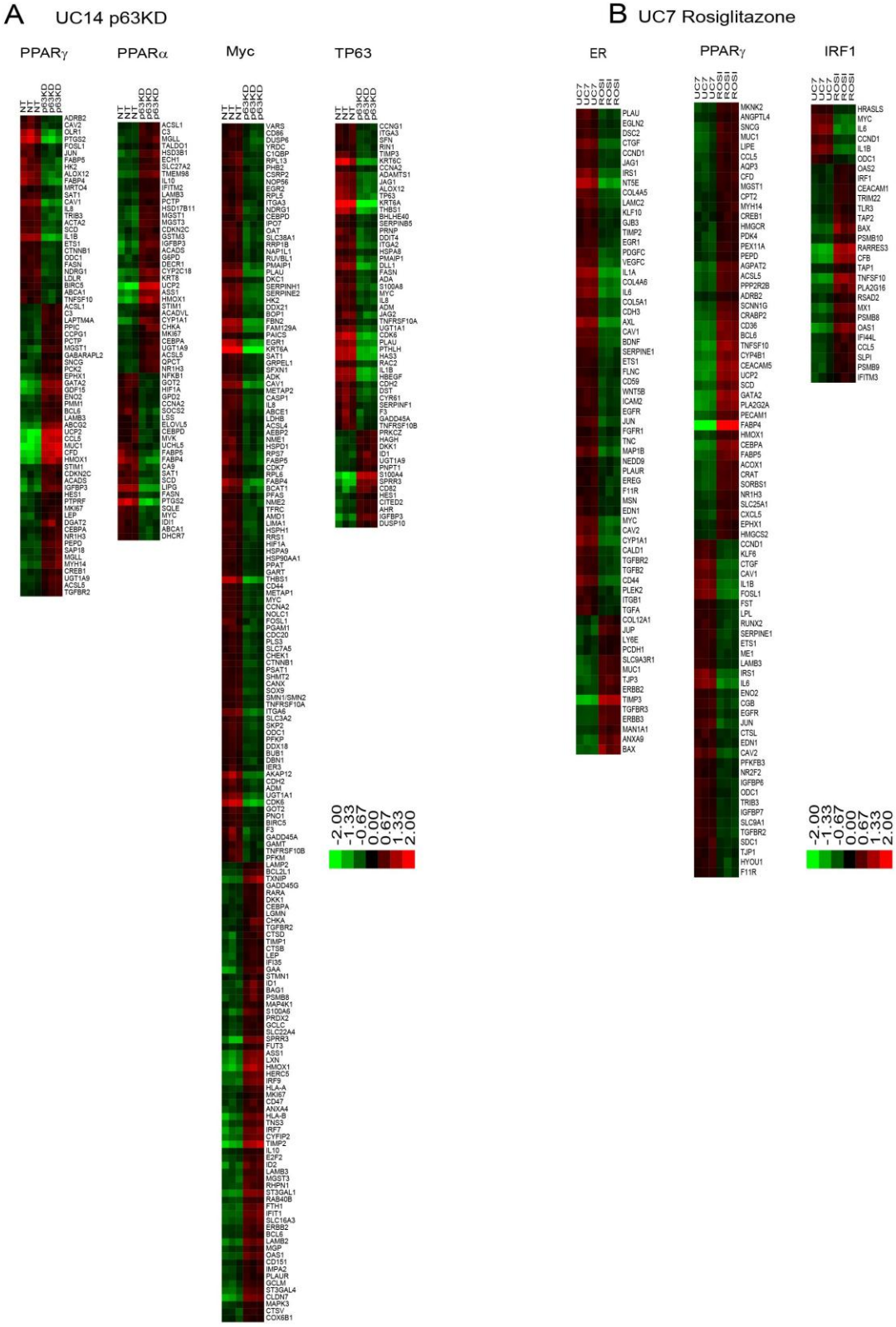


Figure S3. Related to Figure 4.

Expression of transcriptional targets of each upstream regulator in the three subtypes. A. STAT3 pathway genes in the basal subtype as determined by IPA analysis. B. TP63 qPCR data using randomly selected samples from basal and luminal tumors are shown by the mean with SEM. C. ER pathway genes in the luminal subtype. D. TRIM24 pathway genes in the luminal subtype. E. CDKN2A pathway genes in p53-like subtype.


A
UC14 p63KD
PPARγ
PPARα
Myc
TP63
NT
NT
NT
p63KD
p63KD
p63KD
ADRB2
CAV2
OLR1
PTGS2
FOSL1
JUN
FABP5
HK2
ALOX12
FABP4
MRTO4
SAT1
CAV1
IL8
TRIB3
ACTA2
SCD
IL1B
ETS1
CTNNB1
ODC1
FASN
NDRG1
LDLR
BIRC5
ABCA1
TNFSF10
ACSL1
C3
LAPTM4A
PPIC
CCPG1
PCTP
MGST1
GABARAPL2
SNCG
PCK2
EPHX1
GATA2
GOT15
ENO2
PMM1
BCL6
LAMB3
ABCG2
UCP2
CCL5
MUC1
CFD
HMOX1
STIM1
CDKN2C
ACADS
IGFBP3
HES1
PTPRF
MKI67
LEP
DGAT2
CEBPA
NR1H3
PEPD
SAP18
MGLL
MYH14
CREB1
UGT1A9
ACSL5
TGFBR2
ACSL1
C3
MGLL
TALDO1
HSD3B1
ECH1
SLC27A2
TMEM98
IL10
IFITM2
LAMB3
PCTP
HSD17B11
MGST1
MGST3
CDKN2C
GSTM3
IGFBP3
ACADS
G6PD
DECR1
CYP2C18
KRT8
UCP2
ASS1
HMOX1
STIM1
ACADVL
CYP1A1
CHKA
MKI67
CEBPA
UGT1A9
ACSL5
QPCT
NR1H3
NFKB1
GOT2
HIF1A
GPD2
CCNA2
SOCS2
LSS
ELOVL5
CEBPD
MVK
UCHL5
FABP5
FABP4
CA9
SAT1
SCD
LIPG
FASN
PTGS2
SQLE
MYC
ID1
ABCA1
DHCR7
VARS
CD86
DUSP6
YRDC
C1QBP
RPL13
PHB2
CSRP2
NOP56
EGR2
RPL5
ITGA3
NDRG1
CEBPD
IPO7
OAT
SLC38A1
RRP1B
NAP1L1
RUVBL1
PMAIP1
PLAU
DKC1
SERPINH1
SERPINE2
HK2
DDX21
BOP1
FBN2
FAM129A
PAICS
EGR1
KRT6A
SAT1
GRPEL1
SFXN1
ADK
CAV1
METAP2
CASP1
IL8
ABCE1
LDHB
ACSL4
AEBP2
NME1
HSPD1
RPS7
FABP5
CDK7
RPL6
FABP4
BCAT1
PFAS
NME2
TFRC
AMD1
LMNA1
HSPH1
RRS1
HIF1A
HSPA9
HSP90AA1
PPAT
GART
THBS1
CD44
METAP1
MYC
CCNA2
NOLC1
FOSL1
PGAM1
CDC20
PLS3
SLC7A5
CHEK1
CTNNB1
PSAT1
SHMT2
CANX
SOX9
SMN1/SMN2
TNFRSF10A
ITGA5
SLC3A2
SKP2
ODC1
PFKP
DDX18
BUB1
DBN1
IER3
AKAP12
CDH2
ADM
UGT1A1
CDK6
GOT2
PNO1
BIRC5
F3
GADD45A
GAMT
TNFRSF10B
PFKM
LAMP2
BCL2L1
TXNIP
GADD45G
RARA
DKK1
CEBPA
LGMN
CHKA
TGFBR2
CTSD
TIMP1
CTSB
LEP
IFI35
QAA
STMN1
ID1
BAG1
PSMB8
MAP4K1
S100A6
PRDX2
GCLC
SLC22A4
SPRR3
FUT3
ASS1
LXN
HMOX1
HERC5
IRF9
HLA-A
MKI67
CD47
ANXA4
HLA-B
TNS3
IRF7
CYFIP2
TIMP2
IL10
E2F2
ID2
LAMB3
MGST3
RHPN1
ST3GAL1
RAB40B
FTH1
IFIT1
SLC16A3
ERBB2
BCL6
LAMB2
MGP
OAS1
CD151
IMPA2
PLAUR
GCLM
ST3GAL4
CLDN7
MAPK3
CTSV
COX6B1
CCNG1
ITGA3
SFN
RIN1
TIMP3
KRT6C
CCNA2
ADAMTS1
JAG1
ALOX12
TP63
KRT6A
THBS1
BHLHE40
SERPINB5
PRNP
DDIT4
ITGA2
HSPA8
PMAIP1
DLL1
FASN
ADA
S100A8
MYC
IL8
ADM
JAG2
TNFRSF10A
UGT1A1
CDK6
PLAU
PTHLH
HAS3
RAC2
IL1B
HBEGF
CDH2
DST
CYR61
SERPINF1
F3
GADD45A
TNFRSF10B
PRKCZ
HAGH
DKK1
ID1
UGT1A9
PNPT1
S100A4
SPRR3
CD82
HES1
CITED2
AHR
IGFBP3
DUSP10
-2.00
-1.33
-0.67
0.00
0.67
1.33
2.00
B
UC7 Rosiglitazone
ER
PPARγ
IRF1
UC7
UC7
ROSI
ROSI
ROSI
PLAU
EGLN2
DSC2
CTGF
CCND1
JAG1
IRS1
NT5E
COL4A5
LAMC2
KLF10
GJB3
TIMP2
EGR1
PDGFC
VEGFC
IL1A
COL4A6
IL6
COL5A1
CDH3
AXL
CAV1
BDNF
SERPINE1
ETS1
FLNC
CD59
WNT5B
ICAM2
EGFR
JUN
FGFR1
TNC
MAP1B
NEDD9
PLAUR
EREG
F11R
MSN
EDN1
MYC
CAV2
CYP1A1
CALD1
TGFBR2
TGFB2
CD44
PLEK2
ITGB1
TGFA
COL12A1
JUP
LY6E
PCDH1
SLC9A3R1
MUC1
TJP3
ERBB2
TIMP3
TGFBR3
ERBB3
MAN1A1
ANXA9
BAX
MKNK2
ANGPTL4
SNCG
MUC1
LIPE
CCL5
AQP3
CFD
MGST1
CPT2
MYH14
CREB1
HMGCR
PDK4
PEX11A
PEPD
AGPAT2
ACSL5
PPP2R2B
ADRB2
SCNN1G
CRABP2
CD36
BCL6
TNFSF10
CYP4B1
CEACAM5
UCP2
SCD
GATA2
PLA2G2A
PECAM1
FABP4
HMOX1
CEBPA
FABP5
ACOX1
CRAT
SORBS1
NR1H3
SLC25A1
CXCL5
EPHX1
HMGCS2
CCND1
KLF6
CTGF
CAV1
IL1B
FOSL1
FST
LPL
RUNX2
SERPINE1
ETS1
ME1
LAMB3
IRS1
IL6
ENO2
CGB
EGFR
JUN
CTSL
EDN1
CAV2
PFKFB3
NR2F2
IGFBP6
ODC1
TRIB3
IGFBP7
SLC9A1
TGFBR2
SDC1
TJP1
HYOU1
F11R
HRASLS
MYC
IL6
CCND1
IL1B
ODC1
OAS2
IRF1
CEACAM1
TRIM22
TLR3
TAP2
BAX
PSMB10
RARRES3
CFB
TAP1
TNFSF10
PLA2G16
RSAD2
MX1
PSMB8
OAS1
IFI44L
CCL5
SLPI
PSMB9
IFITM3
-2.00
-1.33
-0.67
0.00
0.67
1.33
2.00

C UC9 Rosiglitazone

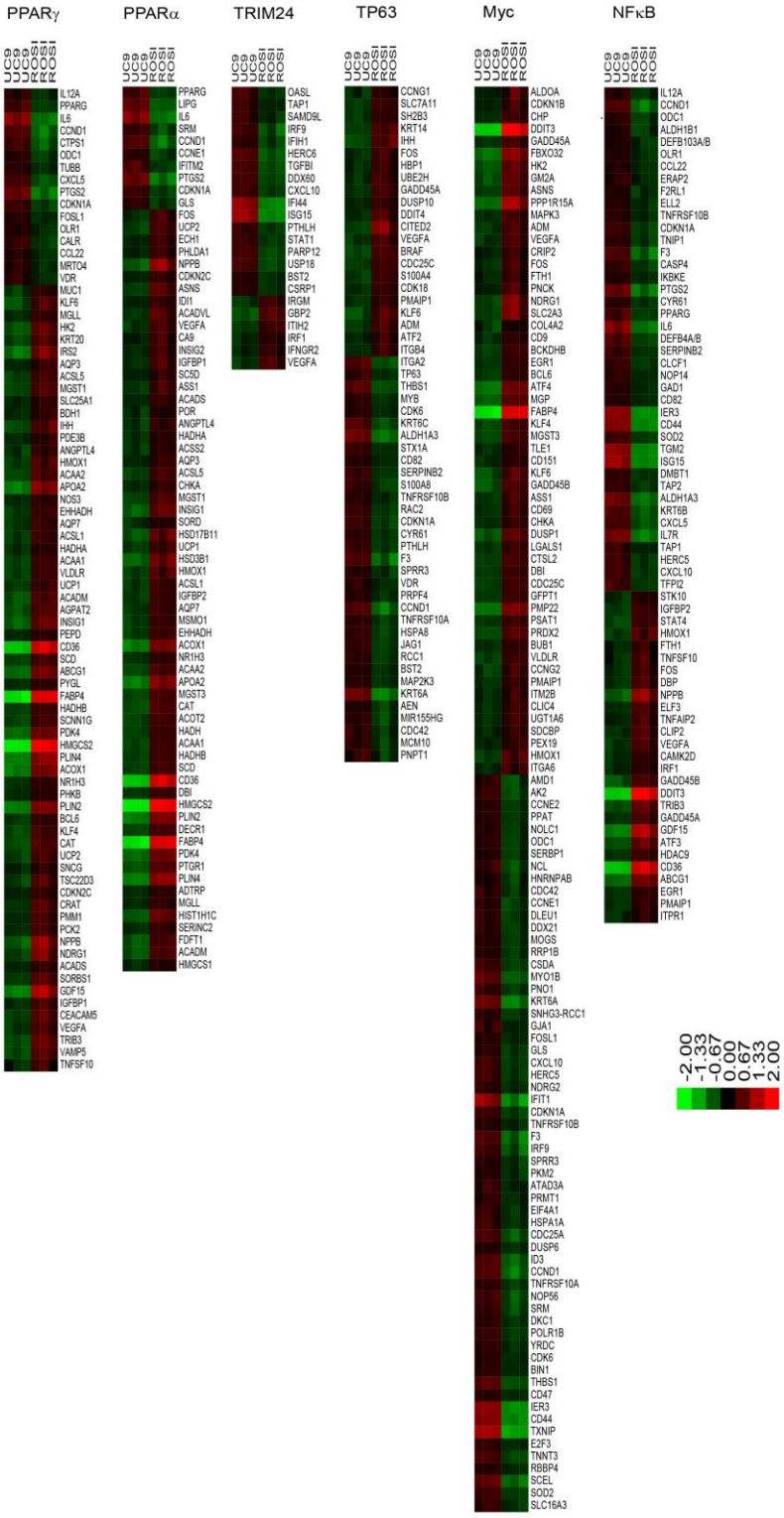


Figure S4. Related to Figure 5.

Heat maps depicting the changes in IPA upstream regulator-associated gene expression resulting from the modulation of basal and luminal transcriptional factors. A. Effects of TP63 knockdown on upstream regulators in UC14. B. Effects of the PPARγ agonist rosiglitazone on upstream regulators in UC7. C. Effects of rosiglitazone on upstream regulators in UC9.

Table S8. Related to Figure 6.

MD Anderson Neoadjuvant Chemotherapy Cohort (n=34*)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size (n)	34	9 (26.5%)	15 (44.1%)	10 (29.4%)	
Mean Age (y) ± SD	65.3 ± 7.3	63.7 ± 5.2	68.1 ± 4.4	62.5 ± 10.7	0.128
Gender (n)					
Male	22 (65%)	5 (56%)	8 (53%)	9 (90%)	0.137
Female	12 (35%)	4 (44%)	7 (47%)	1 (10%)	
Race (n)					
Caucasian	26 (76%)	6 (67%)	13 (86%)	7 (70%)	
African American	5 (15%)	2 (22%)	1 (7%)	2 (20%)	0.788
Hispanic	3 (9%)	1 (11%)	1 (7%)	1 (13%)	
Clinical Stage at TUR (n)					
≤ cT1N0M0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
cT2N0M0	12 (35%)	6 (67%)	4 (27%)	2 (20%)	0.244
cT3N0M0	16 (47%)	2 (22%)	8 (53%)	6 (60%)	
cT4N0M0	6 (18%)	1 (11%)	3 (20%)	2 (20%)	
Positive Clinical Lymph Nodes, cN+ (n)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Positive Clinical Metastasis, cM+ (n)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Pathologic T stage (n)					
pT0	10 (29%)	5 (56%)	2 (13%)	3 (30%)	
pTa, pT1, pTis	5 (15%)	1 (11%)	1 (7%)	3 (30%)	
pT2	2 (6%)	0 (0%)	2 (13%)	0 (0%)	0.267
pT3	10 (29%)	2 (22%)	6 (40%)	2 (20%)	
pT4	7 (21%)	1 (11%)	4 (27%)	2 (20%)	
Positive Pathologic Lymph Nodes (n)	14 (42%)	2 (22%)	10 (67%)	2 (20%)	**0.027**
Variant histology at TUR					
Squamous Differentiation	4 (12%)	2 (22%)	2 (13%)	0 (0%)	
Sarcomatoid Differentiation	1 (3%)	1 (11%)	0 (0%)	0 (0%)	
Other (Micropapillary, Glandular, Adenocarcinoma)	3 (9%)	0 (0%)	2 (13%)	1 (10%)	0.385
Response to NAC (p0 or ≤pT1)^					
Yes	13 (38%)	5 (56%)	2 (13%)	6 (60%)	**0.029**
No	21 (62%)	4 (44%)	13 (87%)	4 (40%)	
Median Overall Survival (m)	42.8	42.8	25.4	65.6	0.147
Median Disease Specific Survival(m)	46.3	42.8	35.4	65.6	0.217

^Response to NAC= Decrease in stage to pT0 or pT1 (for patients with high risk features at TUR: lymphovascular invasion, variant histology, hydronephrosis, or abnormal exam under anesthesia) at cystectomy. * Combination of 18 tumors from Discovery cohort with NAC and 16 tumors from patients treated with NAC on- and off- protocol. The Kruskal-Wallis test was used to compare differences in mean age between groups. Log-rank test was used to compare differences in survival (overall and disease specific) between groups. For the remainder of categorical variables, Fisher's exact test was used to determine differences between subtypes. p-values <0.05 were considered significant.

Table S9. Related to Figure 6.

MD Anderson Phase III MVAC Neoadjuvant Chemotherapy Validation Cohort (n=23)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size (n)	23	10 (43%)	6 (26%)	7 (30%)	
Mean Age (y) ± SD	64.3 ± 10.2	61.1 ± 8.1	69.0 ± 7.7	64.8 ± 13.9	0.129
Gender (n)					
Male	17 (74%)	7 (70%)	5 (83%)	5 (71%)	0.828
Female	6 (26%)	3 (30%)	1 (17%)	2 (29%)	
Race					
Caucasian	20 (87%)	7 (70%)	6 (100%)	7 (100%)	0.106
African American	3 (13%)	3 (30%)	0 (0%)	0 (0%)	
Clinical Stage at TUR (n)					
cT1N0M0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
cT2N0M0	15 (65%)	5 (50%)	4 (67%)	6 (86%)	0.209
cT3N0M0	7 (30%)	5 (50%)	1 (17%)	1 (14%)	
cT4N0M0	1 (4%)	0 (0%)	1 (17%)	0 (0%)	
Positive Clinical Lymph Nodes, cN+	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Positive Clinical Metastasis, cM+	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Treatment: Cystectomy	23 (100%)	10 (100%)	6 (100%)	7 (100%)	
Pathologic T Stage					
pT0	4 (17%)	4 (40%)	0 (0%)	0 (0%)	
pTa, pT1, pTis	2 (9%)	0 (0%)	0 (0%)	2 (29%)	
pT2	8 (35%)	2 (20%)	3 (50%)	3 (43%)	0.084
pT3	8 (35%)	4 (40%)	2 (33%)	2 (29%)	
pT4	1 (4%)	0 (0%)	1 (17%)	0 (0%)	
Positive Pathologic Lymph Nodes (n)	3 (13%)	0 (0%)	1 (17%)	2 (29%)	0.217
Variant histology at TUR					
Squamous Differentiation	4 (17%)	3 (30%)	1 (17%)	0 (0%)	
Sarcomatoid Features^	1 (4%)	1 (10%)	0 (0%)	0 (0%)	0.404
Focal Glandular*	1 (4%)	0 (0%)	0 (0%)	1 (14%)	
Micropapillary	1 (4%)	0 (0%)	0 (0%)	1 (14%)	
Response to NAC (p0 or ≤pT1)					
Yes	6 (26%)	4 (40%)	0 (0%)	2 (29%)	0.208
No	17 (74%)	6 (60%)	6 (100%)	5 (71%)	

^Patient also had focal glandular component on TUR pathology but ultimately had sarcomatoid carcinoma on analysis of cystectomy specimen. *Pathology on cystectomy was adenocarcinoma.

Table S10. Related to Figure 7.

Philadelphia Phase II DDMVAC cohort (n=43)

	TOTAL	Basal	p53-like	Luminal	p-value
Cohort Size (n)	43	14 (33%)	9 (21%)	20 (46%)	
Gender (n)					
Male	29 (67%)	9 (64%)	7 (78%)	13 (65%)	0.757
Female	14 (33%)	5 (36%)	2 (22%)	7 (35%)	
Race (n)					
Caucasian	38 (88%)	11 (79%)	9 (100%)	18 (90%)	
African American	4 (9%)	2 (14%)	0 (0%)	2 (10%)	0.465
Asian	1 (2%)	1 (7%)	0 (0%)	0 (0%)	
Clinical Stage at TUR (n)					
≤ cT1N0M0*	1 (2%)	0 (0%)	0 (0%)	1 (5%)	
cT2N0M0	15 (35%)	4 (29%)	1 (11%)	10 (53%)	
cT3N0M0	18 (42%)	6 (43%)	8 (89%)	4 (21%)	
cT4N0M0	6 (14%)	4 (29%)	0 (0%)	2 (5%)	
Positive Clinical Lymph Nodes (n)	3 (7%)	0 (0%)	0 (0%)	3 (16%)	0.059
Pathologic T stage (n)					
pT0	15 (35%)	7 (50%)	1 (11%)	7 (35%)	
pTa, pT1, pTis	8 (19%)	2 (14%)	0 (0%)	6 (30%)	
pT2	9 (21%)	2 (14%)	3 (33%)	4 (20%)	**0.033**
pT3	8 (19%)	1 (7%)	5 (56%)	2 (10%)	
pT4	3 (7%)	2 (14%)	0 (0%)	1 (5%)	
Positive Pathologic Lymph Nodes (n)	5 (12%)	2 (14%)	2 (22%)	1 (5%)	0.380
Variant histology at TUR					
Squamous Differentiation	4 (9%)	4 (29%)	0 (0%)	0 (0%)	**0.012**
Nested Variant	1 (2%)	0 (0%)	1 (11%)	0 (0%)	
Response to NAC (p0 or ≤pT1)^					
Yes	20 (47%)	7 (50%)	1 (11%)	12 (60%)	**0.048**
No	23 (53%)	7 (50%)	8 (89%)	8 (40%)	

^Response to NAC= Decrease in stage to pT0 or pT1 (for patients with abnormal exam under anesthesia at TUR denoting cT3/T4) at cystectomy. *Patients with upper tract tumors and high grade disease are considered candidates for NAC despite non-muscle invasive disease

Table S11. Related to Figure 7. Provided as an Excel file.

Probes that defined p53-ness in the p53-like subtype.

Probes that were shared between the p53-like and chemoresistant tumors.

Probes that defined p53-ness in the chemoresistant tumors .

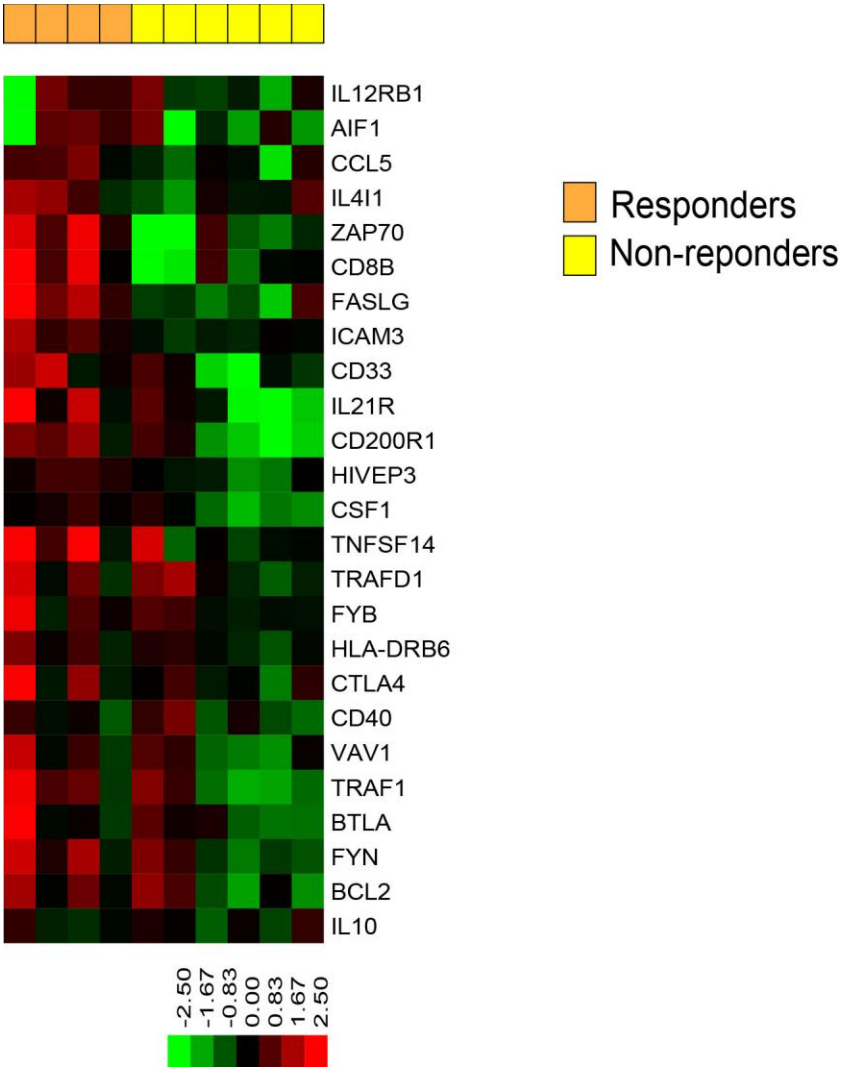


Figure S5. Related to Figure 7.

Heat map depicting the expression of the immune signature identified in the Philadelphia basal tumors in the sensitive basal tumors from the MD Anderson Phase III MVAC cohort.

Supplemental Experimental Procedures

Microarray Experiments and Data Processing: Total RNA from fresh frozen and formalin-fixed, paraffin-embedded human specimens was isolated using the mirVana^{TM}miRNA isolation kit (Ambion, Inc) and the High Pure miRNA isolation kit (Roche), respectively. RNA purity and integrity were measured by NanoDrop ND-1000 and Agilent Bioanalyzer and only high quality RNA was used for the cRNA amplification. For the fresh frozen specimens, direct hybridization assays were performed using the Illumina RNA amplification kit (Ambion, Inc, Austin, TX) and Illumina HT12 V3 chips (Illumina, Inc., San Diego, CA). For the formalin-fixed, paraffin-embedded specimens, DASL (Illumina) was employed using WG-DASL HT12 V4 chips. 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. Gene expression profiling data were uploaded to Gene Expression Omnibus with accession numbers GSE48277 and GSE47993.

BRB ArrayTools version 4.2 developed by National Cancer Institute was used to analyze the data. To identify molecular subtypes, we subjected the data obtained with the fresh frozen ("discovery") cohort to unsupervised hierarchical cluster analysis using the 6700 probes that exhibited expression ratios of at least 2-fold relative to the median gene expression level across all samples in at least 7 samples. The significantly differentially expressed genes for each subset (basal vs. the rest, p53-like vs. the rest and luminal vs. the rest) in the discovery cohort ($p<0.001$ with FDR <0.1, 1.5 fold cut-off) were then extracted and combined to yield 2,507 differentially expressed genes (2,998 probes). Since the discovery cohort, the validation cohort (57 tumors) and the Chungbuk cohort used different versions of the array platform (HT12V3, WG-DASL HT12 V4, and human 6V2 chips, respectively), we used only the 2,709 out of 2,998 probes that were shared between the discovery and validation platforms and the 2,409 out of

2507 genes that were shared between the discovery and Chungbuk validation platforms, respectively. These feature sets were further refined by subjecting them to an F-test ($p < 0.001$), yielding 2,446 probes for the analysis of the validation cohort and 2,160 genes for analysis of the Chungbuk cohort. Each refined feature set was independently centralized and then used to form a oneNN classifier using the discovery cohort, and the prediction accuracy of resulting classifier was tested using leave one out cross validation (LOOCV)(Dudoit et al., 2002; Simon et al., 2003). For LOOCV in the training set, the entire model building process was repeated, including the gene selection process. We also examined whether the cross-validated error rate estimate for a model was significantly less than would be expected from random prediction.

We correlated the presence of squamous features with basal subtype membership in the Lund and UCSF gene expression profiling datasets. Since the MD Anderson discovery and Lund cohorts were analyzed using the same array platform (Illumina HT12v3 chip), the 2998 probes that were significantly differentially expressed in each MIBC subtype in the MD Anderson discovery cohort were used to perform an F-test ($p<0.001$), yielding 2,697 probe sets. Because the UCSF tumors were analyzed using an in-house custom array platform, only the 1,058 genes (out of 2,507) that were common to the UCSF and Illumina platforms were used to perform an F-test to refine the UCSF feature sets (964 genes). Each refined feature set was independently centralized and then used to form a classifier using the MD Anderson discovery cohort. To examine the relative chemotherapy sensitivities of the 3 different MIBC molecular subtypes, we performed prediction analyses using DASL data from the discovery cohort plus 16 additional tumors from patients treated with NAC on- and off-protocol, 23 tumors from a Phase III trial of conventional MVAC (Millikan et al., 2001), 43 TUR tumors and 43 TUR tumors plus all available (n = 20) matched cystectomy tumors from a Phase II clinical trial of dose dense MVAC (DDMVAC). The tumors in each were combined with the MD Anderson validation cohort (n = 57 tumors) and prediction analyses were performed using same probe IDs identified previously

(2,446 probes). After subtype membership was assigned, significantly expressed genes within the basal or luminal tumor subtypes that did or did not respond to chemotherapy were extracted using class comparison tools in BRB ArrayTools (responders vs non-responders in each subsets, p<0.001 with FDR<0.2). Using the same tools, the differentially expressed genes between UC14 cells transduced with non-targeting (NT) or p63-specific shRNAs, or vehicle- or rosiglitazone-exposed UC7 and UC9 cells, were extracted (p<0.001 with FDR<0.1) to perform IPA (1.5-fold cut-off) yielding 2473 probes for UC14, 1546 probes for UC7 and 1673 probes for UC9, and GSEA (2-fold cut-off) yielding 893 probes for UC14, 443 probes for UC7 and 353 probes for UC9. In order to identify a chemoresistance-associated gene signature, a paired t-test (using class comparison tools) was performed on the 20 matched pairs of pre- and post-treatment DDMVAC tumors, yielding 2469 probes (p<0.001 with FDR<0.1, 1.5-fold cut-off) that were used for subsequent IPA. To visualize gene expression patterns, specific gene expression values, adjusted to a median of zero, were used for clustering using Cluster 3.0 and TreeView (Eisen et al., 1998). The probe with larger standard deviation was used for heat map if there are multiple probes for the same gene.

Silhouette score analyses: Silhouette scores were calculated to determine the accuracy of subset membership assignments (Rousseeuw, 1987). The silhouette score for each sample is a measure of how similar that sample is to all other samples in its own cluster compared to the samples in other clusters. More specifically, it is defined as:

$$S(i) = \frac{a_i - b_i}{\min(a_i, b_i)}$$

where a_i is the average distance from the i^{th} sample to the other samples in the same cluster as i, and b_i is the minimum average distance from the i^{th} sample to samples in a different cluster, minimized over clusters.

Gene Set Enrichment Analyses: For GSEA, selected breast basal/luminal markers and an immune signature that was defined by the literature were tested for enrichment in basal and luminal tumors from the MD Anderson discovery cohort and the chemotherapy responders and non-responders within the Philadelphia basal subtype of tumors, respectively (Subramanian et al., 2005). Genes were sorted by the value of the signal to noise ratio against "basal vs luminal and p53-like" phenotype or "responders in basal vs non-responders in basal" phenotype, respectively.

Pathway Analyses: Functional and pathway analyses were performed using Ingenuity Pathway Analysis (IPA) software (Ingenuity® Systems, CA), which contains a database for identifying networks and pathways of interest in genomic data. "Transcriptional factors as molecule type in upstream regulator" categories within IPA were used to interpret the biological properties of the bladder tumor subtypes. For upstream regulator analyses, IPA performs statistical analyses for overlap p values and an activation Z-score. Based on the IPA knowledge database, p values and Z-scores can be calculated based on how many targets of each transcriptional factor were overwrapped (p values) and the extent of concordance of the known effects (activation or inhibition) of the targets in the gene lists (Z-score). Enrichment of KEGG pathway ECM/stromal infiltration in the p53-like tumors was investigated using WebGestalt (http://genereg.ornl.gov/webgestalt) (Wang et al., 2013).

Sequencing Analyses. Sequencing analyses of *FGFR3*, *TP53*, *RB1* and *H/K/N/RAS* on the available tissues was performed on all available tumors within the MD Anderson discovery cohort (66/73). Sequencing was carried out by the ABI Big Dye terminator method. PCR products were purified from unincorporated primers and dNTPs by using exonuclease I and shrimp alkaline phosphatase and subsequent sequencing reactions were analyzed with an ABI 3730 sequencer (Applied Biosystems). Sequence was obtained from both strands and potential

mutations were identified using Mutation Surveyor software (Softgenetics, State College, PA, USA) and confirmed by visual inspection. For selected samples, the presence of mutations was confirmed by sequencing of subcloned amplified exons.

Sequence analysis of *TP53* status in 28 human bladder cancer cell lines was performed in the CCSG-supported genomics core at MD Anderson Cancer Center. Mutation detection was performed by amplifying purified DNA with primers designed to amplify the *TP53* coding regions. Primers were designed using a variety of software applications including, but not limited to, Primer Express v3.0 (Applied Biosystems). The PCR products were purified with ExoSAP-IT (USB) and sequenced in both directions using BigDye Terminator chemistry (Applied Biosystems) and run on a 3730 DNA Analyzer (Applied Biosystems). The sequence data files were aligned and compared to a reference sequence (from Ensembl) in SeqScape v2.5 Software (Applied Biosystems) and mutations analyzed.

Tissue Microarrays and Immunohistochemistry: Basal and luminal cytokeratins (KRT5/6 and KRT20) were analyzed on a tissue microarray consisting of 332 stage-matched (pT3) muscle invasive bladder cancer tissues. Immunohistochemical staining was performed in the MD Anderson Pathology Core using anti-KRT20 (clone Ks20.8), anti-KRT5/6 (clone D5/16 B4)(both from Dako, Carpinteria, CA) and CD44 (HCAM;DF1485, Leica Biosystems, Buffalo Grove, IL) using established clinical protocols. For the TMA analyses, percentages of positive tumor cells were quantified using an automated digital image analyzer, GenoMx^{TM} (Bio Genex, San Ramon, CA).

Cell lines: Cell lines were obtained from the MD Anderson Bladder SPORE Tissue Bank, and their identities were validated by DNA fingerprinting using AmpFlSTR® Identifiler® Amplification kit (Applied Biosystems, Foster City, CA), performed by the MD Anderson Characterized Cell Line Core. Cell lines were cultured in modified Eagle's MEM supplemented with 10% fetal

bovine serum, vitamins, sodium pyruvate, L-glutamine, penicillin, streptomycin, and nonessential amino acids at 37 °C in 5% CO_2 incubator. To generate p63 stable knockdown cells, pan p63 targeting lentiviral shRNA constructs (Open Biosystems, V3LHS_397885) and the pGIPZ lentiviral empty vector (Open Biosystems, RHS4339) were transfected into 293T cells to propagate lentiviral particles. Bladder cancer cells were plated in 6-well plates (12×10^4 cells/well), and medium containing lentiviral particles was added 24 h later. Cells were incubated with lentivirus for 16 h and were washed and cultured in fresh medium. Fluorescence-activated cell sorting (FACS) was performed after 4–5 d to isolate GFP positive cells, and these cells were then cultured in medium containing puromycin (4 μg/ml).

Chemicals. The PPARγ agonist rosiglitazone was purchased from Cayman Chemical (Ann Arbor, MI).

Real-time Quantitative Reverse Transcriptase PCR analyses: p63, cytokeratins (KRT5, KRT6A, KRT20), FOXA1 and CD44 were analyzed by real-time PCR (StepOne; Applied Biosystems, Foster City, CA) using TaqMan primers (Applied Biosystems). All calculations and analyses were performed using StepOne™ Software (Applied Biosystems) that uses the $2^{-\Delta\Delta Ct}$ method with a relative quantification (RQ)min/RQmax (95% confidence level) (Livak and Schmittgen, 2001). The cyclophilin A gene was used as an internal control to normalize for the amount of amplifiable RNA in each reaction. Levels of miR-200b and -c in primary tumors were also quantified by RT-PCR. Total RNA including miRNA was extracted using the mirVana™ miRNA isolation kit (Ambion, Inc) and 10 ng total RNA along with miR-specific primers were used for expression analysis based on the TaqMan MicroRNA Assay system (Applied Biosystems). miRNA U6 expression was used as an internal standard and the $2^{-\Delta\Delta Ct}$ method was used to generate relative expression values. TaqMan primers for hsa-miR-200b (002251) and hsa-miR-200c (002300) were purchased from Applied Biosystems.

Supplemental References

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. . J Comput Appl Math *20*, 53-65.
Simon, R., Radmacher, M. D., Dobbin, K., and McShane, L. M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. Journal of the National Cancer Institute *95*, 14-18.
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., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.
Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic acids research *41*, W77-83.