

Claudin-low bladder tumors are immune infiltrated and actively immune suppressed

Jordan Kardos,^{1,2} Shengjie Chai,^{1,3,4} Lisle E. Mose,¹ Sara R. Selitsky,¹ Bhavani Krishnan,¹ Ryoichi Saito,¹ Michael D. Iglesia,^{1,2} Matthew I. Milowsky,^{1,5,6} Joel S. Parker,^{1,2} William Y. Kim,^{1,2,5,6} and Benjamin G. Vincent^{1,5}

¹Lineberger Comprehensive Cancer Center, ²Department of Genetics, ³Department of Microbiology/Immunology, ⁴Curriculum in Bioinformatics and Computational Biology, ⁵Department of Medicine, Division of Hematology/Oncology, and ⁶Department of Urology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

We report the discovery of a claudin-low molecular subtype of high-grade bladder cancer that shares characteristics with the homonymous subtype of breast cancer. Claudin-low bladder tumors were enriched for multiple genetic features including increased rates of *RB1*, *EP300*, and *NCOR1* mutations; increased frequency of *EGFR* amplification; decreased rates of *FGFR3*, *ELF3*, and *KDM6A* mutations; and decreased frequency of *PPARG* amplification. While claudin-low tumors showed the highest expression of immune gene signatures, they also demonstrated gene expression patterns consistent with those observed in active immunosuppression. This did not appear to be due to differences in predicted neoantigen burden, but rather was associated with broad upregulation of cytokine and chemokine levels from low *PPARG* activity, allowing unopposed *NFKB* activity. Taken together, these results define a molecular subtype of bladder cancer with distinct molecular features and an immunologic profile that would, in theory, be primed for immunotherapeutic response.

Introduction

In the United States, bladder cancer is the fourth most common malignancy in men, with approximately 74,000 new cases and 16,000 deaths expected in 2015. Bladder cancer is histologically divided into low-grade or high-grade tumors that are associated with distinct genomic alterations and differences in prognosis (1). Low-grade tumors are almost uniformly noninvasive (Ta) and have a 5-year survival rate of 96%. In contrast, high-grade tumors can become muscle-invasive and metastatic and are associated with a 5-year survival rate ranging from 70% (muscle-invasive) to 5% (metastatic).

Multiple studies have now identified distinct RNA expression subtypes within both low- and high-grade bladder cancer (2–8). Building upon the work of our colleagues, we and others have recently described distinct subtypes of high-grade muscle-invasive urothelial carcinoma (UC), luminal and basal, that reflect attributes of their corresponding breast cancer subtypes. These studies highlight the similarities in the underlying biology between breast and bladder cancer (2, 8). In addition to the originally reported molecular subtypes of breast cancer (luminal A, luminal B, her2-enriched, and basal-like), a claudin-low subtype of breast cancer has been more recently identified and is characterized by a stromal phenotype, lack of luminal differentiation marker expression, enrichment for epithelial-to-mesenchymal transition (EMT) markers, cancer stem cell-like features, and immune response genes (9).

Clinical trials using immune checkpoint Abs targeting the PD1/PD-L1 axis have recently shown promise in a portion of patients with advanced UC, with the premise that activation of immune checkpoint pathways, including PD-L1, results in active immunosuppression (10). Despite the excitement surrounding PD1/PD-L1 axis inhibition in treating advanced UC, only approximately 30% of patients respond. Therefore, the majority of patients display intrinsic resistance to PD1/PD-L1 inhibition, and a priori identification of these patients would clearly be beneficial.

We report here the discovery of a claudin-low subtype of high-grade, muscle-invasive UC defined by biologic characteristics of the claudin-low subtype of breast cancer. Claudin-low tumors were uniformly enriched for immune gene signatures but simultaneously expressed immune checkpoint molecules, demonstrating that, despite being immune infiltrated, claudin-low tumors are also actively immunosuppressed. Interestingly, the predicted neoantigen burden was not significantly increased in claudin-low tumors.

Authorship note: J. Kardos and S. Chai contributed equally to this work.

Conflict of interest: W.Y. Kim is an inventor on the BASE47 gene classifier.

Submitted: December 7, 2015

Accepted: February 5, 2016

Published: March 17, 2016

Reference information:

JCI Insight. 2016;1(3):e85902.
doi:10.1172/jci.insight.85902.

Instead, they highly expressed cytokines and chemokines associated with leukocyte chemotaxis into the tumor immune microenvironment as a result of an imbalance between PPAR γ and NF- κ B signaling. These results highlight the association between molecular subtype and the degree of immune infiltration and immune suppression and suggest that mechanisms other than neoantigen burden can drive the development of immune infiltrated tumors and also that claudin-low tumors are poised to respond to immune checkpoint inhibition.

Results

Identification of a claudin-low subtype in bladder cancer. Previous studies have identified a claudin-low molecular subtype of breast cancer (11). Given the previously documented similarities in gene expression patterns between breast and bladder cancer (2, 8), we asked whether a claudin-low subtype also exists in bladder cancer. To this end, we performed unsupervised hierarchical clustering on 408 high-grade, muscle-invasive bladder tumors from the The Cancer Genome Atlas (TCGA) urothelial bladder carcinoma (BLCA) data set using gene signatures representative of biologic characteristics that are known to define breast cancer claudin-low tumors such as an enrichment for tumor-initiating cells (TICs) and an EMT (Figure 1A and refs. 9, 11). Specifically, we used gene lists of the tight-junction claudins (*CLDN3*, *CLDN4*, and *CLDN7*) and a previously published bladder cancer–derived TIC signature (12). In addition, we derived a bidirectional (EMT_UP and EMT_DOWN), bladder cancer–specific, notch-dependent EMT gene signature from the publicly available Gene Expression Omnibus (GEO) gene expression data set (GEO GSE60564) (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/jci.insight.85902DS1). Unsupervised hierarchical clustering with these gene signatures revealed a distinct cluster that had characteristics of a claudin-low subtype (Figure 1A, highlighted in green).

To ensure that the set of tumors within the presumed claudin-low cluster were homogeneous and distinct from adjacent clusters of tumors, we performed a Gaussian distribution analysis, starting with the smallest cluster and iteratively repeated the analysis with the addition of adjacent clusters using SigClust R software (Supplemental Figure 1A and ref. 13). This method identified a conserved node of 48 tumors that had consensus enrichment for claudin-low features, and these tumors were therefore defined as claudin-low. All 48 claudin-low tumors were classified as basal by our BASE47 subtype classifier (Fisher's exact $P = 1.18 \times 10^{-16}$) (8), and when examined for their correlation to the BASE47 basal or luminal centroid, they were found to be highly basal (Figure 1B). Further supporting the notion that these tumors exhibit features of claudin-low breast cancer, we applied the previously defined breast cancer–specific claudin-low classifier to the TCGA BLCA tumors and found a significant enrichment (Fisher's exact $P = 1.10 \times 10^{-18}$) of the breast cancer–defined claudin-low tumors within the bladder claudin-low cluster (Supplemental Table 2). Given these findings, we propose a 3-subtype classification of bladder cancer consisting of basal (~40%), luminal (~50%), and claudin-low (~10%) tumors. While basal-like bladder cancer consistently has a worse clinical outcome (2, 3, 7, 8), consistent with previous work on breast cancer (9), we did not find an observable significant difference in overall survival rates between patients with claudin-low tumors and those with basal tumors (Figure 1C).

A 40-gene classifier, bladder claudin-low 40, accurately predicts claudin-low tumors. To define a minimal set of genes that could accurately classify claudin-low bladder tumors, we applied prediction analysis of microarrays (PAMs) to the TCGA BLCA tumors and derived a 40-gene signature, bladder claudin-low 40 (BCL40) (Supplemental Table 3), which accurately classifies bladder tumors into claudin-low and non-claudin-low subtypes, with a training error rate of 0.23 and 0.13, respectively. When combined with the previously validated bladder cancer analysis of subtypes by gene expression (BASE47) predictor (8), this provides a 3-class predictor that can accurately classify bladder tumors as claudin-low, basal, or luminal.

In order to validate the predictor, we compiled a 130-tumor metadata set from 2 previously compiled published data sets (GEO GSE48277; ref. 2). The BASE47 and BCL40 predictors identified 36 claudin-low tumors (~30%), 27 basal tumors (~20%), and 67 luminal tumors (~50%). We found that these subtypes were phenotypically similar to the initially derived subtypes in our discovery set of TCGA bladder tumors as measured by expression of the EMT, TIC, and claudin gene signatures (Supplemental Figure 1, B–E). Furthermore, we ran a transcriptome-wide correlation analysis between the basal, luminal, and claudin-low tumors identified in the discovery (TCGA BLCA) and validation data sets (GEO GSE48277) and found a strong correlation in gene expression between the subtypes identified in the discovery and validation data sets (basal [Pearson's $R = 0.459$, $P < 2.2 \times 10^{-16}$], claudin-low [Pearson's $R = 0.805$, $P < 2.2 \times 10^{-16}$], and

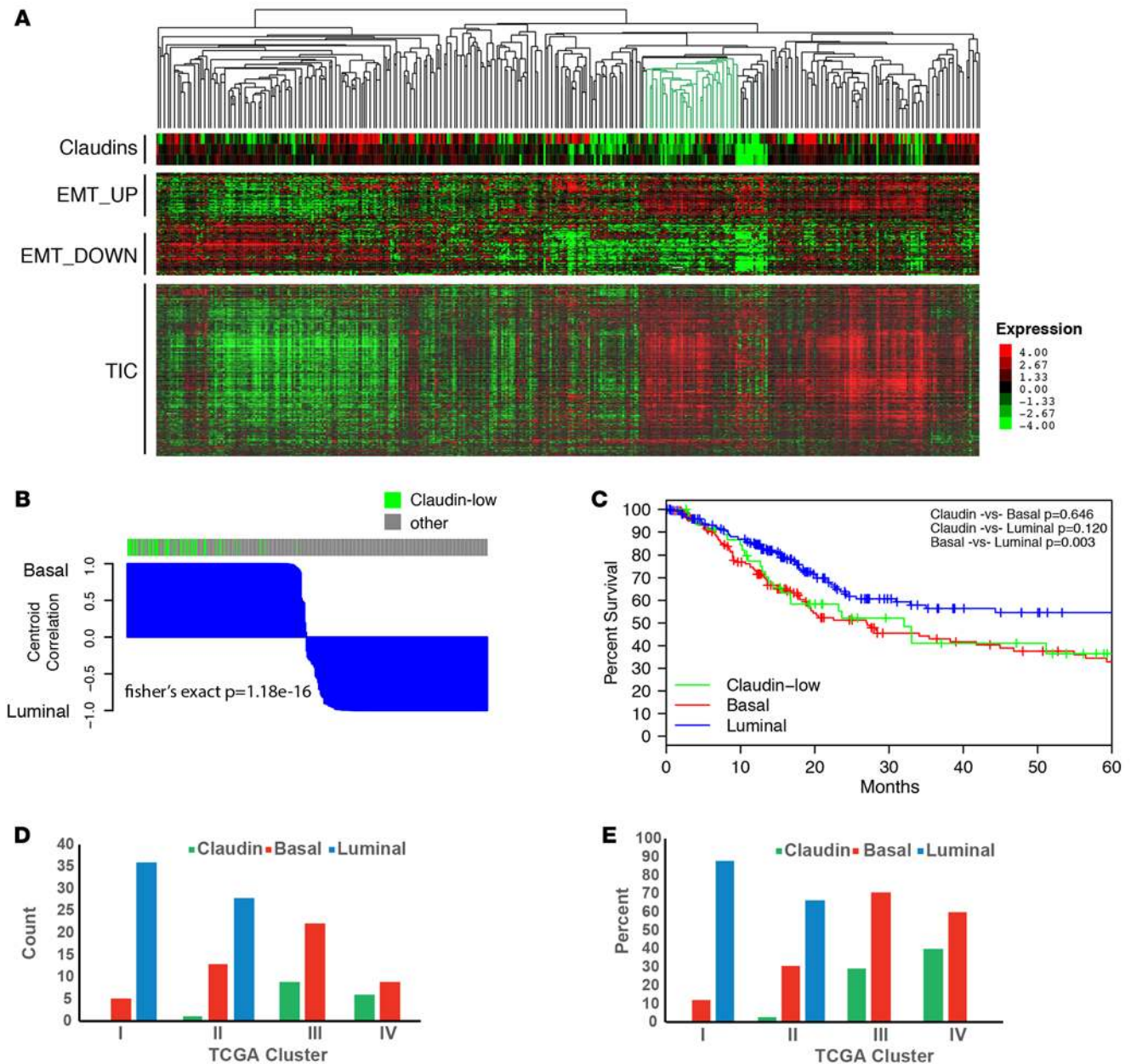


Figure 1. Identification of a claudin-low subtype in bladder cancer. (A) Unsupervised clustering of TCGA muscle-invasive UC samples. Samples were clustered on the basis of expression of tight-junction claudins, a bidirectional EMT signature, and a TIC signature. The tumors identified as claudin-low are highlighted in green on the dendrogram. $n = 408$. (B) Waterfall plot showing correlation with the basal and luminal centroids as defined by BASE47 classification; claudin-low tumors are highlighted in green. Claudin-low tumors were significantly enriched in the BASE47 basal subtype (Fisher's exact test $P = 1.18 \times 10^{-16}$) and were highly correlated with the basal centroid (Pearson's correlation $P = 9.33 \times 10^{-15}$). $n = 408$. (C) Kaplan-Meier plot showing overall survival of bladder cancer by molecular subtype. Significance was determined by log-rank testing with a Bonferroni correction. $n = 408$. (D and E) Bar graphs showing the classification of TCGA UC tumors by TCGA mRNA cluster subtype (x axis) and our subtype classifications (y axis) by count and percentage. $n = 129$. EMT, epithelial-to-mesenchymal transition; TCGA, The Cancer Genome Atlas; TIC, tumor-initiating cell; UC, urothelial carcinoma.

luminal [Pearson's $R = 0.809$, $P < 2.2 \times 10^{-16}$] (data not shown). This further confirmed that the subtypes identified across separate data sets had consistent genome-wide RNA expression profiles.

Comparison with MD Anderson and TCGA UC intrinsic molecular subtype classifications. We next examined whether our claudin-low subtype merely recapitulated any of the existing molecular subtypes of UC published by MD Anderson or TCGA. A comparison of our claudin-low, basal, and luminal predictions on the 408 provisional TCGA BLCA tumors with the MD Anderson oneNN classification system (p53-like, basal, and luminal) (2) re-demonstrated the high concordance of luminal subtype designations (14) as well

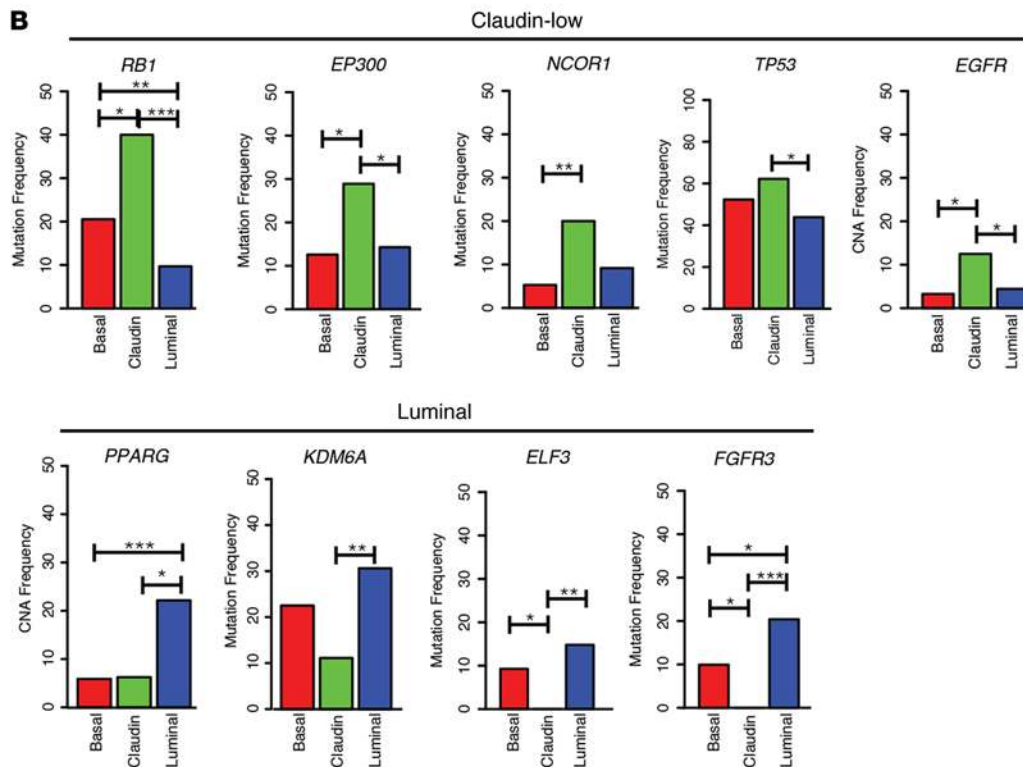
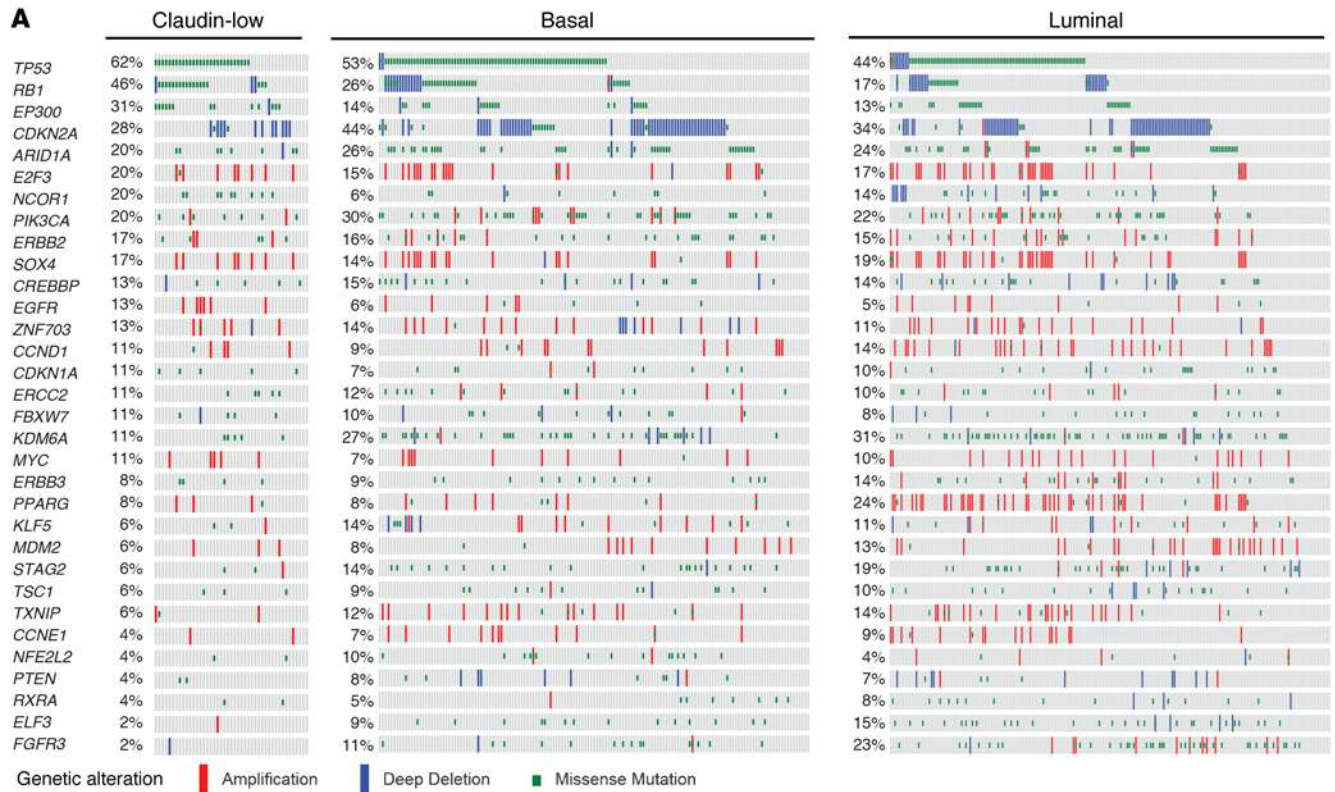


Figure 2. Genomic characterization of bladder cancer subtypes. (A) Oncoprint of genomic copy number alterations and mutations by bladder cancer subtype for genes previously identified as significantly mutated or copy number altered in more than 5% of bladder tumors. $n = 408$. (B) Bar plots of genes that were identified to have a significant ($P < 0.05$) difference in either gene mutation or copy number alteration (CNA) between the claudin-low and basal and/or luminal subtypes. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$, by Fisher's exact test.

as the notion that claudin-low tumors arise primarily from basal tumors (Supplemental Figure 2, A and B). We further compared our claudin-low, basal, and luminal predictions on the 129 published TCGA BLCA tumors with TCGA 4-subtype classification (clusters I, II, III, and IV) (6). Our claudin-low tumors were primarily found in TCGA clusters III and IV (Figure 1, D and E). These comparisons further strengthen the notion that claudin-low tumors do not merely recapitulate a previously described molecular subtype of bladder cancer.

The claudin-low subtype displays unique, intrinsic genomic alterations and gene expression patterns. We next examined the association between molecular subtype and genomic events within significantly mutated or copy number–altered genes identified as being altered at a greater than 5% frequency within TCGA BLCA data set (6). A comparison of claudin-low and basal subtypes revealed that claudin-low tumors had significantly increased rates of *RBI*, *EP300*, and *NCOR1* mutations, an increased percentage of tumors with *EGFR* amplification, as well as decreased rates of mutations in *FGFR3* and *ELF3* (Figure 2, A and B). Relative to the luminal subtype, claudin-low tumors revealed a significantly higher rate of mutation of *TP53*, *RBI*, and *EP300* and an increased percentage of tumors with *EGFR* amplification. Conversely, luminal tumors (compared with claudin-low tumors) had a significantly higher rate of *PPARG* amplification and mutation of *KDM6A*, *ELF3*, and *FGFR3*. These results are in keeping with the notion that genomic alterations and their subsequent effects on signal transduction and transcription may be partially responsible for differences in gene expression subtypes.

To further understand the gene expression patterns that differentiate claudin-low tumors, we performed 2-class significance analysis of microarrays (SAMs), comparing each subtype against all of the other tumors (e.g., claudin-low vs. basal plus luminal). We detected a significant number of differentially expressed genes (FDR = 0.05) (Supplemental Figure 3A and Supplemental Table 4) by this comparison as well as by a direct comparison of each subtype with another (e.g., claudin-low vs. basal) (Figure 3A and Supplemental Table 5). Ingenuity Pathway Analysis (IPA) revealed that, compared with both basal and luminal tumors, claudin-low tumors had significant enrichment in the upstream regulators *IFNG*, *TNF*, and *TGFB1*, which are well-known proinflammatory cytokines (IFN- γ and TNF- α) and pro-EMT (TGF- β) growth factors (Supplemental Table 6). Additionally, claudin-low tumors had higher levels of *IL4* and *IL13* signaling relative to signaling levels in basal and luminal tumors, respectively. Further IPA analysis demonstrated enrichment of other immune-associated pathways in claudin-low tumors (Supplemental Figure 3B). These observations are in keeping with the EMT phenotype, which is a defining characteristic of claudin-low tumors, but are also strongly suggest that these tumors are heavily immune infiltrated.

Claudin-low tumors are enriched in immune gene signature expression. To better characterize the immune cell populations present within claudin-low tumors, we used previously defined gene signatures indicative of specific cellular immune populations (15) and examined their expression by molecular subtype. All examined signatures appeared to be and were statistically enriched in the claudin-low subtype when each signature was collapsed into a single value per tumor (*z* score) (Figure 3B and Supplemental Figure 4). To assess the level of immunosuppression, we examined the expression of a panel of immune checkpoint molecules (immunosuppression score) derived from the literature and found that they were uniformly highly expressed in claudin-low tumors compared with expression levels in both basal and luminal tumors, respectively (Figure 3, C and D).

Bladder cancer as a whole expressed moderate levels of *PD-L1* and our immunosuppression score relative to the spectrum of 12 tumors in TCGA Pan-Cancer analysis (Supplemental Figure 5, A and B, and ref. 16). When broken down by subtype, however, claudin-low tumors in particular had very high levels of *PD-L1* expression (Figure 3E) and high expression of the immunosuppression score (Figure 3F). In aggregate, these findings indicate that claudin-low tumors consistently harbor a high level of immune infiltration that is matched by a high level of active immune suppression. Basal tumors, in contrast, have a more heterogeneous phenotype, while luminal tumors appear to have a paucity of immune cells or immune checkpoint expression. In keeping with this notion, there was a strong correlation between the immune signatures and the immunosuppression signature (Supplemental Figure 5C) across all tumors.

The presence of an immune infiltrate has been shown to be prognostic in other cancers (17). In muscle-invasive bladder cancer, specifically, the presence of CD8⁺ tumor-infiltrating lymphocytes (TILs) (18) and a low ratio of FOXP3 to CD4 or CD8 expression on TILs (19) have been associated with improved disease-free and overall survival rates. In keeping with the work by Sharma et al. (18), Cox proportional hazards (Cox PH) modeling for each immune gene signature across all tumors in TCGA BLCA data set

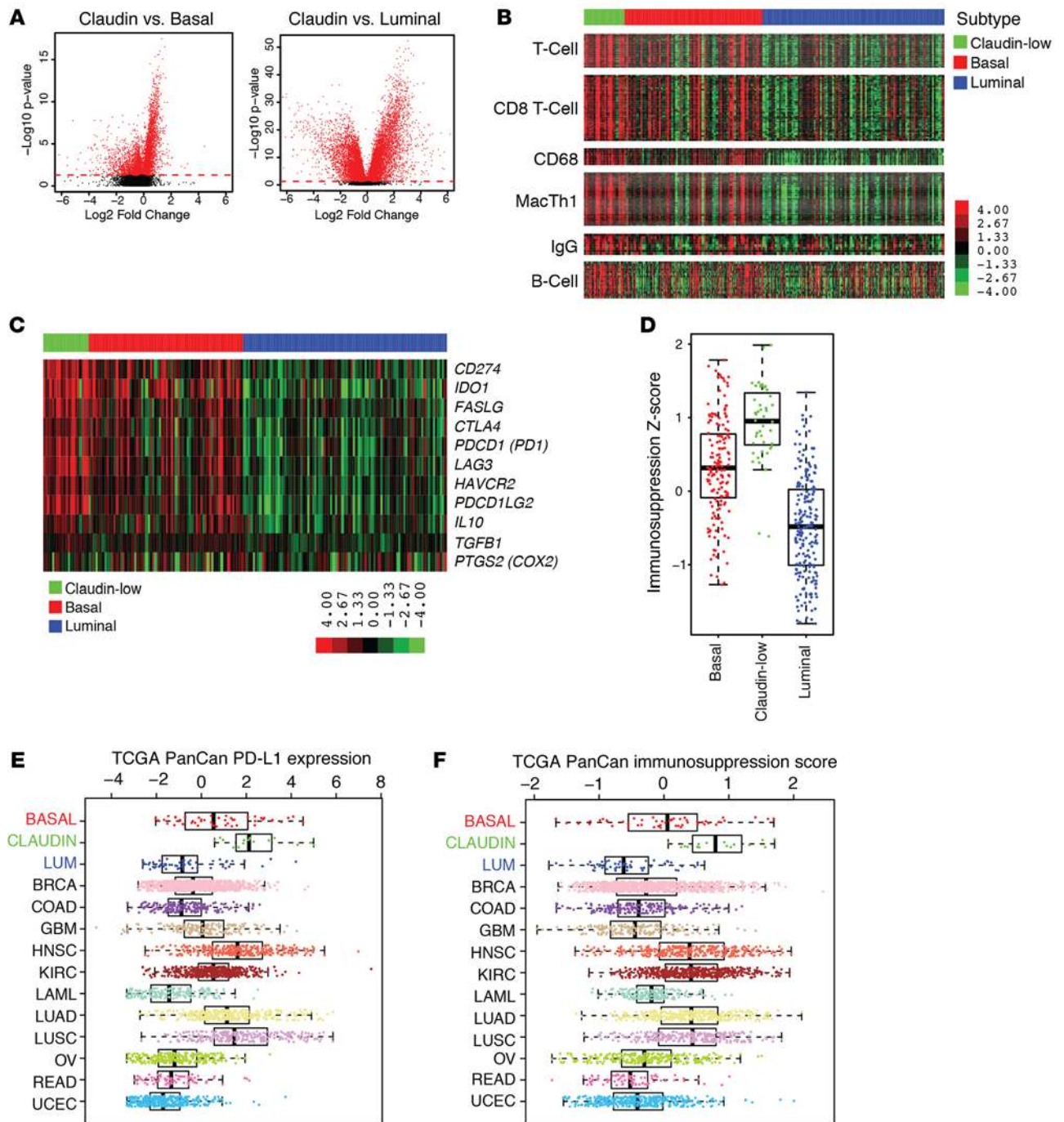


Figure 3. Immune characterization of bladder cancer subtypes. (A) Volcano plot of \log_2 fold change of median gene expression and $-\log_{10} P$ value of gene expression across bladder tumor subtypes. Dashed line across the plots corresponds to a significance threshold of $P = 0.05$. $n = 408$. Significance was calculated using Student's t test with a Bonferroni correction. (B) Heatmaps of supervised clustering of bladder tumor subtypes across previously identified immune signatures. $n = 408$. (C) Heatmap of supervised clustering of bladder tumor subtypes across an immune suppression gene signature. $n = 408$. (D) Box plot of immune suppression gene signature z score across bladder tumor subtypes. $n = 408$. (E) Box plot of PD-L1 gene expression across the Pan-Cancer tumor types. $n = 3,602$. (F) Box plot of immune suppression gene signature z scores across the Pan-Cancer tumor types. $n = 3,602$. The box plots denote the interquartile range (IQR), with the box representing Q1 to Q3, the line denoting Q2, and the whiskers extending an additional 1.5 times the IQR beyond Q1 and Q3. The dots represent data points. BLCA, bladder urothelial carcinoma; BRCA, breast cancer; COAD, colon adenocarcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LAML, acute myeloid leukemia; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; READ, rectum adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; LUM, luminal; TCGA, The Cancer Genome Atlas; PanCan, Pan-Cancer.

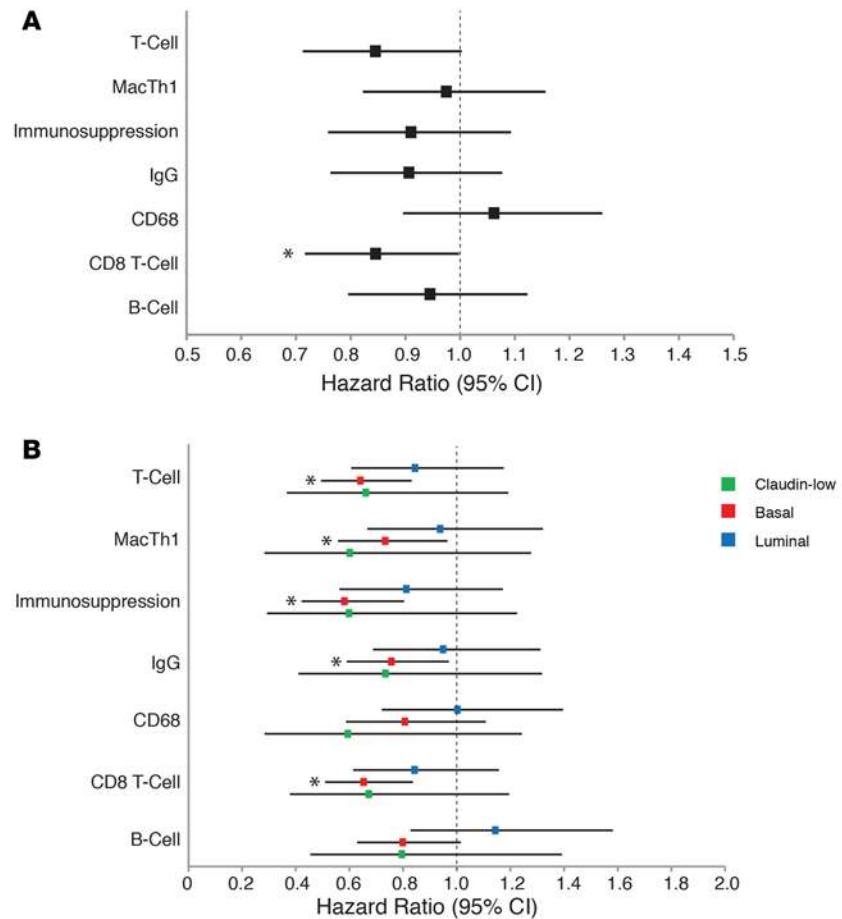


Figure 4. Immune gene signatures have prognostic value across bladder cancer subtypes. (A) Forest plot of Cox PH ratios of the immune gene signatures across all tumors, with a 95% CI indicated around the values. $n = 408$. **(B)** Forest plot of Cox PH ratios of the immune gene signatures within defined tumor subtypes, with a 95% CI indicated around the values. $n = 408$. * $P < 0.05$, prognostically significant signatures by Cox PH modeling. Cox PH, Cox proportional hazard.

showed that only the CD8_T_Cell signature was prognostic (Cox PH = 0.846, $P = 0.047$) (Figure 4A), further supporting the unique importance of CD8⁺ TILs. When Cox PH modeling was performed within each subtype, none of the signatures were prognostic within the claudin-low and luminal subtypes. However, within the basal subtype, numerous signatures were prognostic, including the Ig signature, macrophage signature, T cell signature, CD8⁺ T cell signature, and immunosuppression signature (Figure 4B). We believe these findings are consistent with immune gene signatures being consistently upregulated in the claudin-low subtype and downregulated in the luminal subtype, respectively, while the basal subtype has a more heterogeneous range of gene signature expression, allowing for a more dynamic range across which these subtypes can be prognostic. Supporting this, the basal subtype had the largest SD of immune signature expression across all signatures (basal vs. claudin: $P = 0.007$; basal vs. luminal: $P = 0.097$, Bonferroni-corrected Student's t test).

Specific T cell receptor and B cell receptor gene segment expression levels are prognostic in bladder cancer subtypes. An antigen-driven T cell and/or B cell response would be expected to drive clonal expansion of T cells and/or B cells, resulting in decreased diversity of T cell receptor (TCR) and/or B cell receptor (BCR) repertoires. In addition, if a clonally expanded immune response was active intratumorally, this should be reflected in associations of specific TCR and/or BCR gene segment expression with improved survival. For example, decreased TCR diversity has been associated with response to immune checkpoint inhibition in melanoma (20) and has been shown to be prognostic in bladder cancer (21). To evaluate this concept in TCGA bladder samples, we fit Cox PH models to test the association of expression of each TCR or BCR gene segment with survival and calculated the number of prognostic gene segments by subtype. To establish null distributions for the number of gene segments expected in each subtype, we used the bootstrap

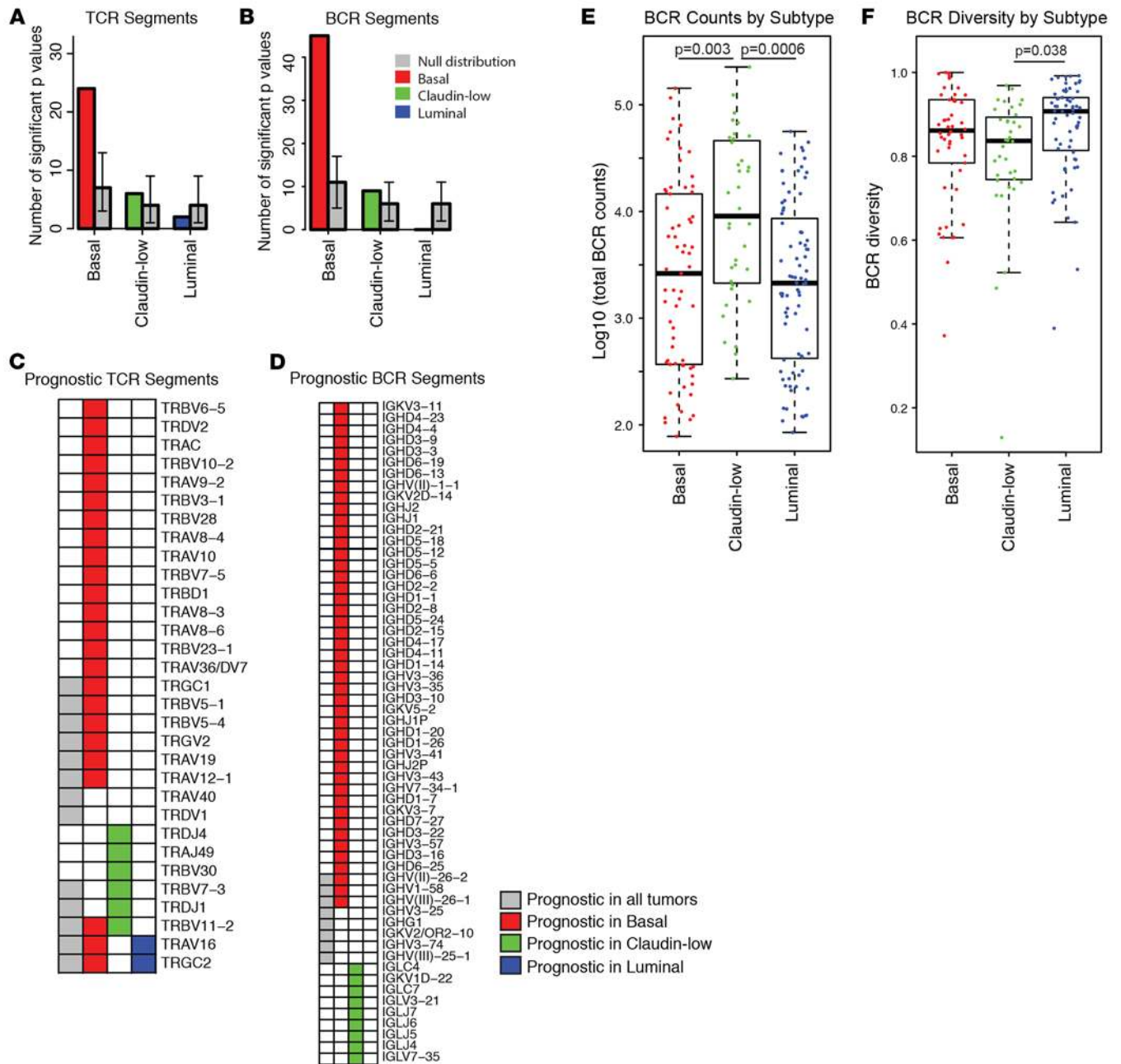


Figure 5. BCR and TCR segment expression is prognostic. (A) Number of TCR gene segments by subtype in which increased expression was significantly associated with improved survival by Cox PH model fit. Null distributions (gray bars) with 95% CIs were generated for each by bootstrap resampling of non-TCR genes and calculation of the number of significant *P* values that were similarly associated with prolonged survival. *n* = 292. (B) Number of BCR gene segments by subtype in which increased expression was significantly associated with improved survival by Cox PH model fit. Null distributions (gray bars) with 95% CIs were generated for each by bootstrap resampling of non-TCR genes and calculation of the number of significant *P* values that were similarly associated with prolonged survival. *n* = 292. (C) Specific TCR gene segments in which increased expression was significantly associated with improved survival by Cox PH model fit for all tumors (gray boxes), basal tumors (red boxes), claudin-low tumors (green boxes), and luminal tumors (blue boxes). (D) Specific BCR gene segments in which increased expression was significantly associated with improved survival by Cox PH model fit for all tumors (gray boxes), basal tumors (red boxes), claudin-low tumors (green boxes), and luminal tumors (blue boxes). (E) Log base 10 number of reads supporting any BCR V(D)J rearrangement are shown by subtype. *n* = 181. Mann-Whitney *U*-Wilcoxon test with an FDR multiple testing correction was used to determine significance. (F) Repertoire diversity by subtype. The box plots in E and F denote the interquartile range (IQR), with the box representing Q1 to Q3, the line denoting Q2, and the whiskers extending an additional 1.5 times the IQR beyond Q1 and Q3. The dots represent data points. *n* = 150. Mann-Whitney *U*-Wilcoxon test with an FDR multiple testing correction was used to determine significance. BCR, B cell receptor; Cox PH, Cox proportional hazard; TCR, T cell receptor.

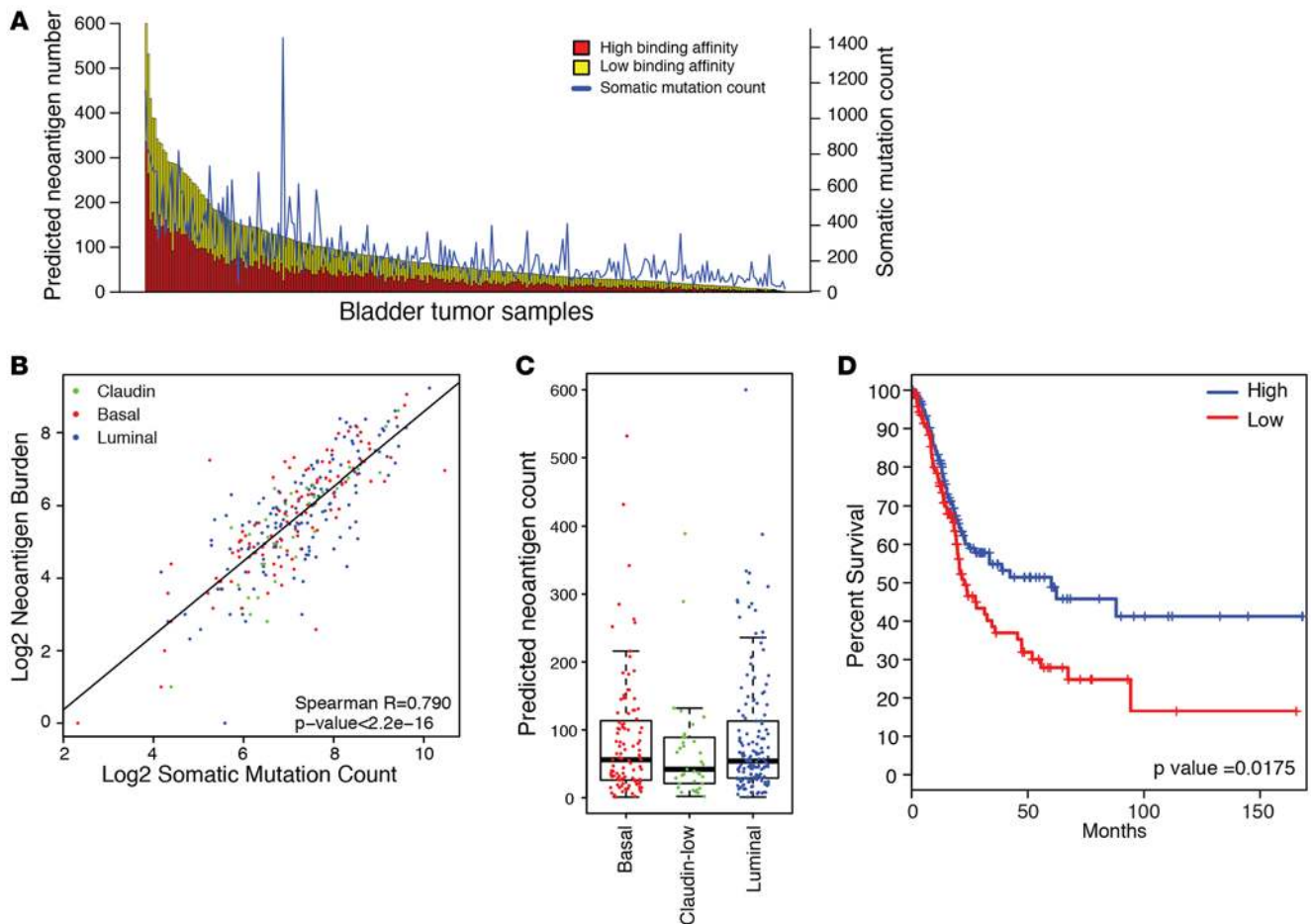


Figure 6. Predicted neoantigen burden by bladder cancer subtype. (A) Stacked bar plot showing the number of predicted neoantigens in each bladder tumor with a predicted IC_{50} of less than 50 nm (red bars) and less than 150 nm (yellow bars). Numbers of predicted neoantigens are shown in the left y axis. Blue line and right y axis show the number of missense mutations per tumor. $n = 289$. (B) Scatter plot of somatic missense mutations (\log_2) versus predicted neoantigen burden (\log_2) across TCGA data set. Significance and correlation were determined using Spearman's rank test. $n = 289$. (C) Box plot showing the number of predicted neoantigens with an IC_{50} of less than 50 nm by tumor molecular subtype. Subtypes were not significantly different ($P > 0.05$). Significance was determined by 1-way ANOVA. $n = 289$. The box plots denote the interquartile range (IQR), with the box representing Q1 to Q3, the line denoting Q2, and the whiskers extending an additional 1.5 times the IQR beyond Q1 and Q3. The dots represent data points. (D) Kaplan-Meier plot showing survival of bladder cancer patients with high (greater than median value, blue line) versus low (less than median value, red line) predicted numbers of neoantigens. Vertical hash marks indicate censored data. Significance was determined by log-rank test. $n = 289$. TCGA, The Cancer Genome Atlas.

resampling method previously published by our group (15). For both TCR gene segments (Figure 5A) and BCR gene segments (Figure 5B), a significantly higher number of gene segments than expected by chance were prognostic in the basal subtype, but not in the claudin-low or luminal subtype. Figure 5, C and D show the specific gene segments that were prognostic in each subtype. Prognostic segments were found in multiple TCR and BCR families, with a small number of gene segments discovered in multiple subtypes (i.e., TRBV11-2). This suggests that adaptive immune responses important in endogenous antitumor immunity are not uniform in TCR and BCR usage between the subtypes.

Despite the presumed importance of assessing T cell and B cell clonality in tumor immunology, at present, this can only be done by direct TCR or BCR sequencing. Our group has developed a bioinformatics method (VDJician) to accurately and efficiently reconstruct rearranged BCR V(D)J sequence repertoires from short-read RNA-sequencing data. We applied this to TCGA bladder data to evaluate whether overall BCR expression (Figure 5E) and/or repertoire diversity (Figure 5F) varied by subtype. BCR expression was higher and repertoire diversity lower (indicative of clonal expansion) in the claudin-low subtype relative to that observed in the luminal subtype, which is consistent with the presence of a selective antigen-directed response in claudin-low tumors. These results, in conjunction with our previous findings, indicate that claudin-low tumors are immune infiltrated and have an active immune response within the tumor microenvironment.

Predicted neoantigen burden does not vary significantly by bladder cancer subtype but is selectively associated with survival in basal tumors. Neoantigens are altered peptides derived from tumor-intrinsic mutant proteins that are presented by MHC molecules and can drive robust antitumor T cell responses (22). This is in contrast to self-antigens that may be overexpressed in tumors but have been subjected to central immune tolerance (23). Neoantigens derived from tumor-specific genomic aberrations can be predicted using whole-exome sequencing of paired tumor and matched normal samples, and expression is confirmed by incorporation of RNA expression data. The predicted neoantigen number has been positively associated with favorable clinical outcomes in multiple tumor types (24) as well as with response to immune checkpoint inhibition in melanoma (25, 26) and non-small-cell lung cancer (27). These results suggest an important protective role for the endogenous repertoire of T cells able to target tumor cells. In order to determine whether neoantigen burden varied by bladder tumor subtype, we implemented an informatics pipeline based on the approach published by Rajasagi et al. (28) and applied this to TCGA bladder data (Figure 6). There was a noisy but clear correlation between predicted neoantigen burden and the number of somatic mutations (Figure 6A: left y axis and right y axis, respectively) (Spearman's $R = 0.79$, $P < 2 \times 10^{-16}$, Figure 6B). Interestingly, claudin-low tumors, despite having a high level of immune infiltration and active immunosuppression, did not have a significantly different level of predicted neoantigens compared with that of basal or luminal subtypes (Figure 6C).

To assess the association between predicted neoantigen burden and subtype, we performed Cox PH analysis with the predicted neoantigen count as the potential explanatory variable. In the basal but not claudin-low or luminal subtypes, an increased number of predicted neoantigens was associated with prolonged survival ($P = 0.025$). For all bladder tumors taken together, the association was significant as well ($P = 0.005$). Figure 6D shows survival curves for all bladder tumors divided by the median predicted neoantigen count into high versus low neoantigen burden. Analyzed in this way as well, high neoantigen burden was associated with prolonged overall survival. Therefore, while there is a high correlation between bladder cancer molecular subtype and immune signature expression, this does not appear to be explained by the predicted neoantigen number.

Claudin-low tumors express high levels of cytokines and chemokines normally repressed by PPARG. Given that predicted neoantigen burden was relatively similar across molecular subtypes, we explored the possibility that claudin-low tumors harbor an immune infiltrate because of increased production of proinflammatory cytokines and chemokines. To this end, we examined the relative expression of a panel of cytokines and chemokines (Supplemental Table 7) and their receptors among bladder subtypes and found that the majority of them were significantly upregulated in claudin-low tumors relative to expression levels in both basal and luminal tumors (Figure 7, A and B). We noted that NF- κ B target genes in particular were significantly upregulated in the claudin-low subtype compared with expression in both the basal and luminal subtypes (Fisher's exact P value = 1.885×10^{-8} , data not shown).

A defining transcriptional program of urothelial differentiation and of luminal bladder tumors is activation of peroxisome proliferator-activated receptor γ (*PPARG*) signaling (29). Consistent with this, we noted that *PPARG* was significantly amplified in luminal relative to claudin-low tumors (Figure 2B). Because *PPARG* is known to directly inhibit NF- κ B signaling (30), we hypothesized that heightened *PPARG* activity might play a role in restraining the proinflammatory effects of NF- κ B. Using a publicly available gene expression data set (GEO GSE48124), we noted that the expression changes induced by treatment with rosiglitazone, a PPAR γ agonist, in UMUC7 and UMUC9 bladder cancer cells predicted suppression of the upstream regulator *NFKB1* as well as a number of genes known to be activated by NF- κ B (*STAT5A*, *IL6*, *TNF*, *CCL5*) (Supplemental Table 8). Furthermore, rosiglitazone-treated UMUC7 and UMUC9 cells had downregulation in gene signatures of NF- κ B activation as assessed by gene set enrichment analysis (GSEA) (Figure 7C). Interestingly, we saw that rosiglitazone treatment resulted in significant downregulation of immune checkpoint molecules (such as *PDL1*, *PDL2*, *IL12*, and *PGSL2*) found in our immunosuppression signature (Figure 7D). In aggregate, these data support the notion that downregulation of PPAR γ activity results in unopposed NF- κ B signaling, which contributes to the proinflammatory milieu of claudin-low tumors as well as to their high level of active immune suppression.

Finally, in keeping with recent work demonstrating that EMT is associated with immune checkpoint molecule expression (31, 32), we observed a strong correlation between our bladder cancer-derived EMT signatures and multiple immune signatures including our immunosuppression score: $R = 0.462$ [EMT (Up)] and $R = -0.471$ [EMT (Down)]; $P < 2.2 \times 10^{-16}$ (both "Up" and "Down") (Figure 7E). Further-

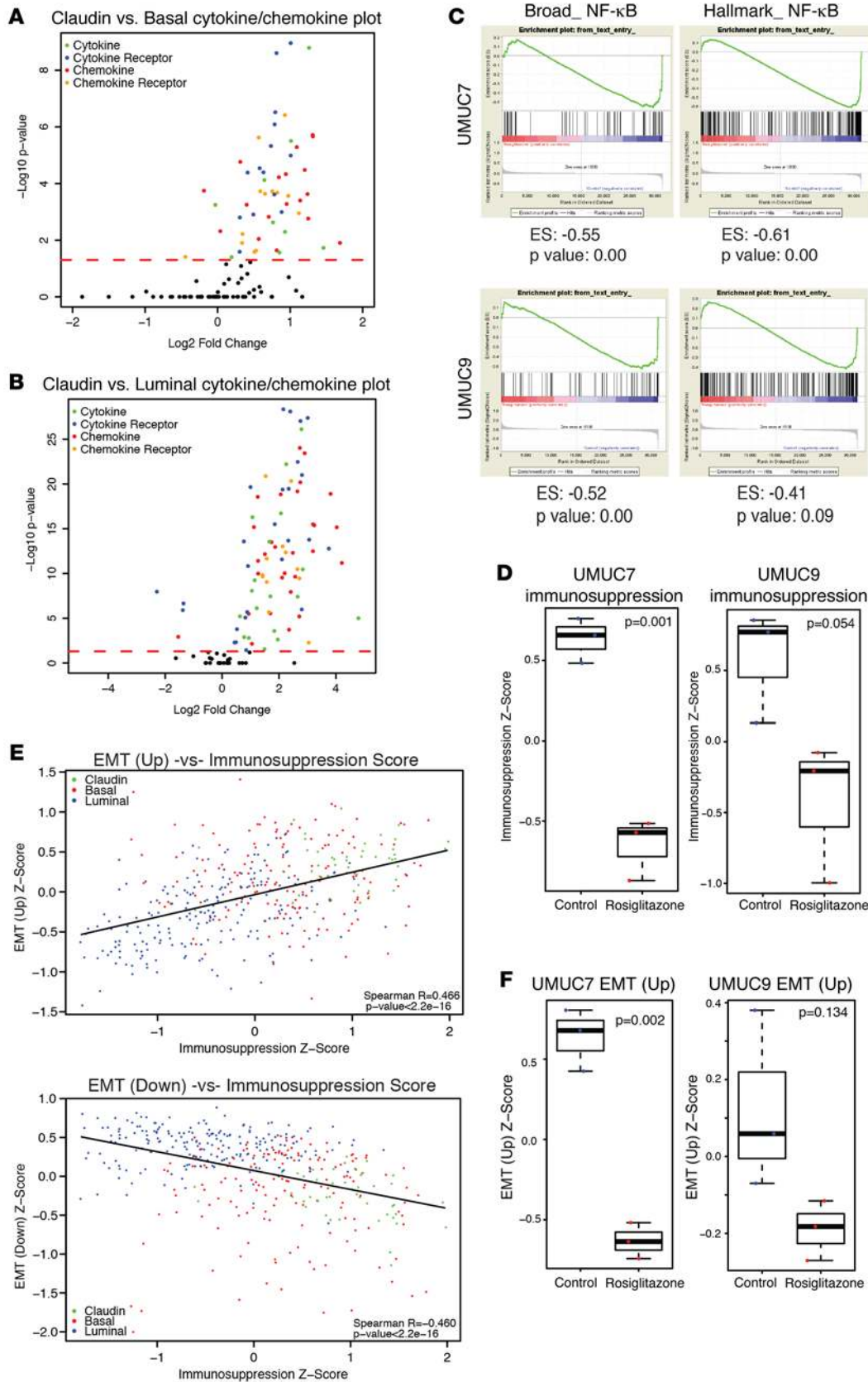


Figure 7. Cytokine and chemokine regulation across bladder cancer subtypes. (A and B) Volcano plots of \log_2 fold change of median gene expression and $-\log_{10} P$ value of gene expression for cytokines and chemokines across claudin-low/basal and claudin-low/luminal subtypes. Dashed lines across plots correspond to $P = 0.05$. Significance was calculated using Student's t test with a Bonferroni correction. $n = 408$. (C) GSEA enrichment plots indicating that NF- κ B signatures were decreased in rosiglitazone-treated UMUC7 and UMUC9 bladder cancer cell lines. Significance was determined using GSEA software. (D) Box plots showing that immunosuppression gene signature expression was significantly decreased across UMUC7 and UMUC9 cell lines after rosiglitazone treatment. Significance was determined using Student's t test. $n = 6$. (E) Correlation plot of immunosuppression and EMT gene signature expression. $n = 408$. Significance and correlation were calculated using a Spearman's rank test. (F) Box plots showing that EMT gene signature expression was decreased across UMUC7 and UMUC9 cell lines after rosiglitazone treatment. Significance was determined using Student's t test. $n = 6$. The box plots in D and F denote the interquartile range (IQR), with the box representing Q1 to Q3, the line denoting Q2, and the whiskers extending an additional 1.5 times the IQR beyond Q1 and Q3. The dots represent data points; EMT, epithelial-to-mesenchymal transition; GSEA, gene set enrichment analysis.

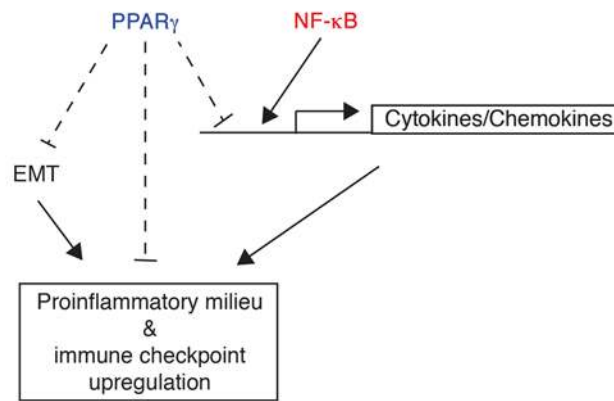


Figure 8. Model of immune infiltration across bladder cancer subtype. Proposed model of immune response regulation through PPAR γ and NF- κ B signaling.

more, given the important role of PPAR γ in terminal urothelial differentiation (30), we hypothesized that it may be a critical regulator of epithelial-mesenchymal balance in urothelial cancers. Indeed, we found that PPAR γ activation (by rosiglitazone) in UMUC7 and UMUC9 cells decreased levels of our EMT (Up) signature (Figure 7F).

Discussion

Herein, we characterize the claudin-low, molecular subtype of high-grade UC. Claudin-low bladder tumors are defined by high levels of EMT, enrichment for TIC signatures, and low expression levels of tight-junction claudins. In addition, claudin-low tumors are enriched in specific genomic alterations (e.g., mutations in *EP300* and *NCOR1* as well as amplification in *EGFR*) and have a distinct transcriptional profile. Furthermore, we found that claudin-low tumors are highly enriched in all immune gene signatures examined, but also express high levels of immune checkpoint molecules. In contrast to melanoma and non-small-cell lung cancer, the predicted neoantigen burden did not appear to correlate with immune infiltration in bladder cancer. Instead, claudin-low tumors appeared to downregulate PPAR γ signaling, resulting in unopposed NF- κ B activity and contributing to a proinflammatory milieu (Figure 8).

In our study, as in previous studies, expression of the various immune gene signatures was highly correlated, including high correlations between gene signatures associated with specific cellular subpopulations (CD8⁺ T cells, B cell lineage, Th1-polarizing macrophages) and the immunosuppression gene signature. This supports the claim that tumors growing in the presence of immune cell influx must adaptively suppress the antitumor response in order to survive. Immune gene signature expression levels, the prognostic value of immune gene signatures, and TCR and BCR gene segment expression divide the bladder cancer subtypes into 3 groups: (a) low infiltrate with nonsignificant prognostic value (luminal); (b) heterogeneous infiltrate with significant prognostic value (basal); and (c) high infiltrate with nonsignificant prognostic value (claudin-low). We hypothesize that the lack of prognostic benefit in claudin-low and luminal tumors is driven by different mechanisms. Luminal tumors were sparsely infiltrated and showed low expression levels of molecules associated with immunosuppression (Supplemental Table 9). In contrast, claudin-low tumors showed a substantial but ineffective infiltrate in the context of high expression levels of immunosuppression markers. Immune features may fail to be prognostic in luminal tumors because no infiltrate is present, whereas they fail in claudin-low tumors because, despite a dense infiltrate, the level of immunosuppression overwhelms active antitumor immunity. Basal tumors have the highest degree of variability in immune gene signature expression, and in this model, some basal tumors will have generated an immune response that is competing more effectively (though ultimately insufficiently to clear tumor) with tumor-driven immune suppression. While additional studies are required to test this hypothesis, our data suggest that claudin-low tumors as a whole, as well as a subset of basal tumors, are poised for response to immune checkpoint blockade.

The different molecular aberrations that characterize the bladder cancer subtypes may yield differential exposure of antigens to the immune system, resulting in skewing of the tumor-infiltrating TCR and/or BCR repertoires in predictable ways should the antigens be public (i.e., shared between multiple patients). Though our study was not designed to formally test this, we report here a high degree of variability, in which adaptive immune gene segments were prognostic among the bladder cancer subtypes, an effect that would

be expected if TCR/BCR repertoire features associated with tumor targeting were to vary by tumor subtype. Interestingly, in the basal subtype, multiple TCR gene segments associated with $\gamma\delta$ T cells were found to be significantly prognostic ($P < 0.05$ by Cox PH). As this specific subset of T cells is involved in adaptive immunity at mucosal surfaces and able to respond to mycobacteria, $\gamma\delta$ T cells may be involved in antitumor immunity and an attractive target for the development of biomarkers of response to bladder cancer immunotherapy, including bacille Calmette-Guérin (BCG), which is commonly given for nonmuscle invasive disease.

We report here the VDJcian algorithm that performs de novo assembly of repertoires of fully rearranged BCR VDJ sequences. When analyzed, the claudin-low subtype showed the highest expression levels but the lowest repertoire diversity compared with basal and luminal subtypes. This is consistent with the presence of an antigen-driven response in the claudin-low tumors, leading to clonal expansion of antigen-reactive B cell–lineage cells. Plasma cells are known to express high levels of BCR mRNA, and these results would also be consistent with a restricted plasma cell infiltrate. In addition, as plasma cells represent a terminal differentiation in the B cell lineage in response to antigenic stimulation, their presence would also be expected in an antigen-driven response. Future experiments will be necessary to confirm these findings and attempt to map immunogenic epitopes in claudin-low tumors.

In melanoma and a subset of solid tumors, neoantigen burden correlates with expression of perforin and granzyme A (a measure of cytolytic activity) (26, 33) and tumors with these attributes appear to be more responsive to CTLA4 checkpoint blockade. In bladder cancers examined in that study, there was a trend toward increased cytolytic activity, with increased predicted neoantigen burden ($P = 0.096$, data not shown) (33). In contrast, we did not see significant correlations between neoantigen burden and predicted features such as T cell or CD8⁺ T cell gene signatures, immunosuppression score, or molecular subtype, suggesting that alternate etiologies exist to explain the proinflammatory state of claudin-low and basal tumors relative to that of luminal tumors. In this regard, we observed significant upregulation of cytokines and chemokines in claudin-low tumors and hypothesize that this cytokine milieu is favorable to a proinflammatory state and immune cell influx. We propose that PPAR γ activity, through its ability to repress NF- κ B, is inversely correlated with this proinflammatory milieu and, therefore, that luminal tumors, which are enriched in *PPARG* amplification and activation of *PPARG* gene signatures, have very little inflammation. Conversely, we found that claudin-low tumors, which have relatively low levels of *PPARG* pathway activation, have high levels of immune infiltration. Therefore, in contrast to the inflamed tumors found in melanoma and non–small-cell lung cancer, which appear driven by neoantigen expression, inflamed bladder cancers have a proinflammatory state induced by an enhanced cytokine/chemokine milieu. It will be important to determine whether altering the balance between PPAR γ and NF- κ B activity can be used to alter the immune milieu toward a more favorable response to immune therapy and whether other transcriptional programs can be harnessed as well.

Finally, while immune checkpoint inhibition holds great promise, the response rates of various solid tumors remain approximately 20% to 30%, suggesting that many patients will not derive benefit. Our *BASE47* and *BCL40* gene classifiers, which can accurately subtype high-grade bladder tumors, may serve to identify useful predictive biomarkers of response (i.e., claudin-low) or lack of response (i.e., luminal) to PD1 axis inhibition. Moreover, our studies further validate the notion of subtype-specific therapy in bladder cancer (i.e., basal = chemotherapy; claudin-low = immune checkpoint blockade) and advance the possibility that claudin-low breast tumors may have similar immune features.

Methods

TCGA data set manipulation. TCGA Bladder Urothelial Carcinoma RNA Expression data set was downloaded from the Broad Institute Firehose Pipeline (<http://gdac.broadinstitute.org>) on August, 27, 2015. RNA expression was downloaded in a normalized RSEM file. Expression values were \log_2 transformed, and genes with less than 80% expression across all samples were filtered out. Missing values were imputed using the *K*-nearest neighbor imputation method. Tumor-adjacent normal samples were removed, and gene expression values were median centered across each gene. TCGA Pan-Cancer data set was downloaded from the Synapse website (<https://www.synapse.org>) from data set syn 2468297 (16). Genes with less than 80% expression across all samples were filtered out. Missing values were imputed using *K*-nearest neighbor imputation.

Gene signatures. Bladder TIC, EMT, and tight-junction claudin gene signatures were used in the classification of a claudin-low subtype. The TIC signature was derived by Chan et al. (12). The set of claudins used was identified by Prat et al. (9). The EMT signature is a bidirectional signature derived on the GEO

(GEO GSE60564) data set of Notch2 overexpression in a urinary bladder RT4V6 cell line. The data set was mean collapsed onto genes. Genes were filtered for a significant difference (Student's *t* test, $P < 0.05$) between the control and Notch2-overexpressed (EMT-induced) cell lines and also for their presence in TCGA bladder UC data set. Genes were then ranked on the basis of median difference between the 2 groups. The top 50 genes with the most increased expression in the EMT-induced cells and the top 50 genes with the most decreased expression in the EMT-induced cells were used to create the bladder cancer-specific EMT_UP and EMT_DOWN signatures, respectively. Immune gene signatures used to describe immune cell processes were derived by Iglesia et al. (15). *Z* scores were calculated for each claudin, basal, and luminal subtype and box plots made of the distributions. Gene signature *z* scores were obtained by calculating the *z* score of each gene within a signature across all samples and taking the median of all gene *z* scores within a gene signature as the *z* score of the gene signature.

Identification of a claudin-low class. Bladder basal and luminal predictions and centroid distances were made using the BASE47 PAM Classifier derived by Damrauer et al. (8). Breast cancer claudin predictions were made using the Distance-Weighted Discrimination (DWD) Claudin Classifier provided by Prat et al. (9).

Data were clustered on the TIC/EMT (Up and Down)/claudin gene sets using average linkage clustering with a centered correlation similarity metric on the Cluster 3.0 platform. Each gene set was individually clustered across genes using average linkage clustering. Gene sets were collapsed down to *z* scores, and a conservative node with high TIC/high EMT UP/low EMT DOWN/low claudin gene set was selected. SigClust was run on the node, expanding out to the entire gene set for each increasing node. Differences in gene expression subtypes were determined using SAMs run on R, with an FDR of 0.05. A PAM predictor (BCL40) was derived on the 408 tumor TCGA data set for a claudin/other subtype classifier. A threshold of 6.4 was selected, giving a 40-gene predictor with an overall error rate of 0.14

A validation data set of 130 muscle-invasive UC samples was compiled from 73-sample and 57-sample data sets from GEO (GEO GSE48277) (2). Each data set was mean collapsed onto genes. The data set was combined and batch effect adjusted using parametric empirical Bayesian adjustments through the ComBat function in the sva R package and was then median centered. Genome-wide correlations and significance were calculated using a Pearson's correlation test.

Clinical, mutation, and copy number alteration analysis. Mutation, copy number, and clinical data were downloaded as mutation packager calls through the Broad Institute Firehose Pipeline (<http://gdac.broadinstitute.org>) on September 3, 2015. Survival status and overall survival were determined on the basis of the data provided. Oncoprint figures were produced using the downloaded TCGA mutation and copy number alteration (CNA) data. Genes were selected on the basis of previously being identified as having significant mutations or CNAs within the gene (6). Significance in CNA and mutation across subtypes was determined using Fisher's exact test. Cox PH ratios and CIs were derived using the survival package on the R platform.

Pathway analysis. Cellular pathway analysis across subtypes was performed using QIAGEN's IPA (www.qiagen.com/ingenuity). Comparison across subtypes was done using the gene list with an FDR of 0.00 as determined by SAM analysis across subtypes.

Gene signature expression analysis. Supervised clustering of samples was performed across all tumor samples by claudin, basal, and luminal subtypes. Genes within each signature were clustered using average linkage on Cluster 3.0. Significance across gene signature *z* scores was calculated using Student's *t* test. Cytokines and chemokines were identified using a RegEx search to capture all members of the molecular families (Supplemental Table 7). Volcano plots were produced using Bonferroni-adjusted Student's *t* test *P* values, and fold change was calculated using normalized RSEM expression values. NF- κ B gene signatures were accessed through Molecular Signatures Database (MSigDB) or compiled by the Broad Institute. GSEA software was used to produce enrichment plots (<http://www.broad.mit.edu/gsea/>) (34). UMUC7 and UMUC9 cell line data were accessed through GEO data sets GSE48124 and GSE47993, respectively. Expression values were mean collapsed onto genes. Gene signatures were compiled on the basis of existing gene lists. Significance was calculated by collapsing gene signatures into *z* scores as described above, and 2-tailed Student's *t* tests were performed across gene signatures.

TCR and BCR gene segment expression analysis. Expression levels of 353 BCR gene segments and 240 TCR gene segments were determined for TCGA bladder tumor samples with available TCGA mRNA-sequencing data and survival data using bedtools (version 2.17.0). Gene expression values were normalized to the upper quartile of total reads within a sample as previously described (35). Survival analyses were performed using a Cox PH model to derive *P* values and coefficients for each gene segment using the Cox

PH function in the survival package in R. The number of gene segments that were significantly associated with improved survival ($P < 0.05$ and coefficient < 0) was calculated for each bladder tumor subtype. Null distributions describing the expected number of prognostic gene segments for each subtype were estimated with 95% CIs according to the bootstrap method previously published by our group (15).

Fisher's exact test was used to compare the number of BCR segments and TCR segments significantly associated with improved survival among all subtypes.

Analysis of rearranged BCR repertoires using VDJician. The VDJician software accepts mRNA-sequencing data mapped to the genome as input and builds a deBruijn graph of read pairs that map to IgH loci or have similarity with germline IgH alleles as well as all unmapped reads. The graph is traversed exhaustively, resulting in a set of putative contigs. Anchor sequences near the 3' end of V segments and the 5' end of J segments are identified in an up-front indexing step. If a contig contains a sequence within a configurable distance of a V anchor and a J anchor, the anchors are a reasonable distance apart, and conserved amino acids that typically bind a CDR3 segment are present (cysteine and tryptophan for IgH), the contig is considered a candidate. The original set of reads is mapped to candidate contigs, which are then further filtered on the basis of coverage and read pair information. VDJician outputs a final set of contigs along with alignments of the original reads mapped to these contigs. This output was passed to RSEM for transcript quantification. The total BCR count was calculated by summing the read count values for all predicted BCR sequences for each sample. Evenness was calculated by dividing the Shannon-Wiener diversity index by the number of BCR sequences for each sample (example expression in R): $-\text{sum}(\text{read count}/\text{sum}(\text{read count})) * \log(\text{read count}/\text{sum}(\text{read count})) / \log(\text{number of BCR sequences})$. P values were determined using a Mann-Whitney U -Wilcoxon test.

Neoantigen prediction. The bladder cancer data set used for neoantigen prediction consisted of 289 samples with available TCGA mRNA-sequencing data, exome-sequencing data, and tumor-specific mutation annotation data (6). Neoantigens were predicted using a bioinformatics pipeline similar to that developed by Rajasagi et al. (28). Tumor-specific single nucleotide variant annotation data were downloaded from the Broad Institute Firehose Pipeline (<http://gdac.broadinstitute.org>). Pysam was used to determine RNA-sequencing read coverage of missense mutations, and bedtools (version 2.17.0) was used to determine the exome-sequencing read coverage of missense mutations. Nine- and ten-mer peptides derived from 3 ORFs with all possible combinations of missense mutations that overlap the genomic location of peptide in the ENCODE reference transcript set were considered in the peptide generation pipeline. DNA sequences corresponding to peptides were retrieved and translated in silico into protein sequences. The expression levels of each peptide generated were determined by the lowest missense mutation RNA-sequencing read coverage. PHLAT was used to identify the HLA class I (HLA-A, HLA-B, HLA-C) type of each tumor sample (36). Binding affinity to MHC molecules expressed by the tumor for all possible 9- and 10-mer peptides generated from missense mutations was predicted using NetMHCpan (version 2.8). Binding affinity of peptides to null alleles, alternatively expressed alleles, and alleles not supported by NetMHCpan were not predicted. Peptides were then filtered by their binding affinities (IC_{50} nM) to each class I allele in the tumor sample's HLA type and RNA expression level of the predicted source transcript(s). Peptides with an IC_{50} value of less than 150 nM for at least 1 class I allele and RNA read support of at least 2 reads were considered predicted neoantigens.

Statistics. A P value of less than 0.05 was considered significant across all analyses performed. SigClust statistical analysis software was used to determine significance in Figure 1A and Supplemental Figure 1A. A Fisher's exact test was used in Figure 1B; Figure 2, A and B; and Supplemental Table 2. A Pearson's correlation was performed in Figure 1B. A log-rank test of survival difference was performed in Figure 1C (Bonferroni-corrected) and Figure 6B. A Bonferroni-corrected 2-tailed Student's t test was performed in Figure 3A; Figure 7, A, B, D, and F; Supplemental Figure 1, B-E; and Supplemental Figure 4. Cox PH modeling was performed in Figure 4, A and B, and Figure 5, A-D. A Mann-Whitney U -Wilcoxon test with an FDR multiple testing correction was performed in Figure 5, E and F. A Spearman's rank correlation was used in Figure 6B, Figure 7E, and Supplemental Figure 5C. A 1-way ANOVA was used in Figure 6C. GSEA significance testing was used in Figure 7C. SAM significance testing was performed in Supplemental Figure 3A and Supplemental Tables 4 and 5. IPA significance testing was used in Supplemental Figure 3B and Supplemental Tables 6 and 8.

Study approval. No experiments included in the manuscript used animal or human subjects and, as such, did not require IRB approval.

Author contributions

JK, SC, MIM, LEM, JSP, RS, BGV, and WYK designed the research questions and studies. JK, SC, LEM, SRS, BK, and BGV conducted the experiments. JK, SC, MIM, LEM, SRS, MDI, and BGV acquired and analyzed the data. LEM and JSP provided reagents. JK, SC, MIM, BGV, and WYK wrote the manuscript.

Acknowledgments

We acknowledge the members of the Kim, Vincent, and Serody laboratories for their useful discussions, the McConkey laboratory for providing their bladder cancer subtype calls, as well as the Lineberger Bioinformatics Core for their technical support. This work was supported by the Bladder Cancer Advocacy Network (BCAN); the James Family Foundation and a Partner Fund Management Bladder Cancer Research Innovation Award (to W.Y. Kim); the University Cancer Research Fund (UCRF) (to W.Y. Kim and B.G. Vincent); and the UNC Oncology Clinical Translational Research Training Program (5K12CA120780, to B.G. Vincent).

Address correspondence to: William Y. Kim or Benjamin G. Vincent, Lineberger Comprehensive Cancer Center, University of North Carolina, CB# 7295, Chapel Hill, North Carolina 27599-7295, USA. Phone: 919.966.4765; E-mail: wykim@med.unc.edu (W.Y. Kim). Phone: 919.966.8412; E-mail: benjamin_vincent@med.unc.edu (B.G. Vincent).

- Goebell PJ, Knowles MA. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urol Oncol*. 2010;28(4):409–428.
- Choi W, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell*. 2014;25(2):152–165.
- Rebouissou S, et al. EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med*. 2014;6(244):244ra91.
- Sjödahl G, et al. A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res*. 2012;18(12):3377–3386.
- Dyrskjöt L, et al. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet*. 2003;33(1):90–96.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315–322.
- Volkmer JP, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc Natl Acad Sci U S A*. 2012;109(6):2078–2083.
- Damrauer JS, et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A*. 2014;111(8):3110–3115.
- Prat A, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010;12(5):R68.
- Powles T, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature*. 2014;515(7528):558–562.
- Prat A, et al. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat*. 2013;142(2):237–255.
- Chan KS, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc Natl Acad Sci U S A*. 2009;106(33):14016–14021.
- Liu Y, Hayes DN, Nobel A, Marron JS. Statistical significance of clustering for high-dimension, low-sample size data. *J Amer Statist Assoc*. 2008;103(483):1281–1293.
- Choi W, et al. Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol*. 2014;11(7):400–410.
- Iglesia MD, et al. Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. *Clin Cancer Res*. 2014;20(14):3818–3829.
- Hoadley KA, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929–944.
- Fridman WH, et al. Immune infiltration in human cancer: prognostic significance and disease control. *Curr Top Microbiol Immunol*. 2011;344:1–24.
- Sharma P, et al. CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc Natl Acad Sci U S A*. 2007;104(10):3967–3972.
- Horn T, et al. The prognostic effect of tumour-infiltrating lymphocytic subpopulations in bladder cancer. *World J Urol*. 2016;34(2):181–187.
- Tumeh PC, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515(7528):568–571.
- Choudhury NJ, et al. Low T-cell receptor diversity, high somatic mutation burden, and high neoantigen load as predictors of clinical outcome in muscle-invasive bladder cancer [published online ahead of print October 9, 2015]. *Eur Urol Focus*. doi: 10.1016/j.euf.2015.09.007.
- Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69–74.
- Snook AE, Magee MS, Schulz S, Waldman SA. Selective antigen-specific CD4(+) T-cell, but not CD8(+) T- or B-cell, tolerance corrupts cancer immunotherapy. *Eur J Immunol*. 2014;44(7):1956–1966.

24. Brown SD, et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res.* 2014;24(5):743–750.
25. Snyder A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med.* 2014;371(23):2189–2199.
26. Van Allen EM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015;350(6257):207–211.
27. Rizvi NA, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* 2015;348(6230):124–128.
28. Rajasagi M, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood.* 2014;124(3):453–462.
29. Choi N, Zhang B, Zhang L, Ittmann M, Xin L. Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer Cell.* 2012;21(2):253–265.
30. Peters JM, Shah YM, Gonzalez FJ. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nat Rev Cancer.* 2012;12(3):181–195.
31. Chen L, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun.* 2014;5:5241.
32. Mak M, et al. A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial to mesenchymal transition. *Clin Cancer Res.* 2016;22(3):609–620.
33. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1):48–61.
34. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–15550.
35. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519–525.
36. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics.* 2014;15(1):325.