Non-muscle-invasive micropapillary bladder cancer has a distinct lncRNA profile associated with unfavorable prognosis

Joep J. de Jong[1], Begoña P. Valderrama[2], Julia Perera[3], Nuria Juanpere[4], Paloma Cejas[5], Henry Long[5], Mar Alba[3], Ewan A. Gibb[6#], Joaquim Bellmunt[7#]

[1]Department of Urology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
[2] Hospital Universitario Virgen del Rocio Sevilla
[3]IMIM-Hospital del Mar and GRIB
[4]Department of Pathology, Parc de Salut Mar, Universitat Pompeu Fabra, Barcelona
[5]Center for functional Cancer Epigenetics, Dana Farber Cancer Institute, Boston, Massachusetts, USA
[6]Veracyte Inc., Vancouver, Canada
[7]Beth Israel Deaconess Medical Center, PSMAR-IMIM Lab, Harvard Medical School, Boston, Massachusetts, USA

[#]Contributed equally

Keywords: non-muscle-invasive bladder cancer, micropapillary, gene expression profiling, variant histology, biomarkers, long non-coding RNA

Running title: lncRNA profiling for HGT1 micropapillary bladder cancer

Additional information:

***Corresponding authors:**
Ewan A. Gibb
Joaquim Bellmunt

Target Journal: British Journal of Cancer

Disclaimers:
Ewan Gibb is an employee of Veracyte, Inc. The remaining authors have no direct or indirect commercial incentive associated with publishing the article.

ABSTRACT: (378 words)

Background: Molecular subtyping of bladder cancer has revealed luminal tumors generally have a more favorable prognosis. However, some aggressive forms of variant histology, including micropapillary, are often classified luminal. In previous work, we found long non-coding RNA (lncRNA) expression profiles could identify a subgroup of luminal bladder tumors with less aggressive biology and better outcomes.

Objective: In the present study, we aimed to investigate whether lncRNA expression profiles could identify high-grade T1 micropapillary bladder cancer with differential outcome.

Design, setting and participants:
LncRNAs were quantified from RNA-seq data from a HGT1 bladder cancer cohort that was enriched for primary micropapillary cases (15/84). Unsupervised consensus clustering of variant lncRNAs identified a three-cluster solution, which was further characterized using a panel of micropapillary-associated biomarkers, molecular subtypes, gene signatures and survival analysis. A single-sample genomic signature was trained using lasso-penalized logistic regression to classify micropapillary-like gene-expression, as characterized by lncRNA clustering. The genomic classifier (GC) was tested on luminal tumors derived from the TCGA cohort (N=202).

Outcome measurements and statistical analysis: Patient and tumor characteristics were compared between subgroups by using X^2 tests and two-sided Wilcoxon rank-sum tests. Primary endpoints were overall, progression-free and high-grade recurrence-free survival, calculated as the date of high-grade T1 disease at TURBT till date of death from any cause, progression or recurrence, respectively. Survival rates were estimated using weighted Kaplan Meier (KM) curves.

Results and limitations: Primary micropapillary HGT1 showed decreased FGFR3, SHH and p53 pathway activity relative to tumors with conventional urothelial carcinoma. Many bladder cancer-associated lncRNAs were downregulated in micropapillary tumors, including *UCA1, LINC00152, and MALAT1*. Unsupervised consensus clustering resulted in a lncRNA cluster 1 (LC1) with worse prognosis that was enriched for primary micropapillary histology and the Luminal Unstable (LumU) molecular subtype. Interestingly, LC1 appeared to better identify aggressive HGT1 disease, compared to stratifying outcomes using primary histologic characteristics. A signature trained to identify LC1 cases showed good performance in the testing cohort, identifying seven cases with significantly worse survival ($p<0.001$). Limitations include the retrospective nature of the study and the lack of a validation cohort.

Conclusions
Using the lncRNA transcriptome we identified a subgroup of aggressive HGT1 bladder cancer that was enriched with micropapillary histology. These data suggest that lncRNAs can facilitate the identification of aggressive micropapillary-like tumors, potentially improving patient management.

Patient Summary
The expression patterns of a unique set of genes called non-coding RNAs helped to identify patients with aggressive non-muscle invasive bladder cancer.

INTRODUCTION

Micropapillary bladder cancer is an aggressive urothelial variant that frequently co-occurs with conventional urothelial carcinoma of the bladder[1]. When present, micropapillary histology has been associated with aggressive clinical behaviour, often presenting as more advanced disease[2]. While the clinical characteristics of micropapillary bladder cancer are well-established, the molecular features of micropapillary bladder cancer are less well-characterized[3, 4].

Molecular subtyping has advanced our understanding of muscle-invasive bladder cancer (MIBC), showing potential utility in predicting prognosis and treatment response[5-7]. At the highest level, bladder cancer can be divided along an intrinsic basal and luminal axis, with additional classifications providing refined granularity[8-10]. More recently, efforts have been made to apply molecular subtyping to non-muscle-invasive bladder cancer (NMIBC), where most cases are found to have a luminal character[11, 12]. In general, luminal tumors tend to have a more favorable prognosis, although the luminal class does present with significant heterogeneity at the molecular level[13]. Unexpectedly, certain aggressive forms of variant histology, including micropapillary, tend to be classified as luminal[3].

While most studies to date have focused on messenger RNA (mRNA) expression to differentiate molecular subtypes of NMIBC, the mammalian transcriptome is comprised of a diverse range of coding (mRNA) and non-coding RNAs[14, 15]. Long non-coding RNAs (lncRNAs) are mRNA-like transcripts that range in length from 200 nucleotides to over 100 kilobases and lack open reading frames. While it is unclear how many lncRNAs have biological functions, their expression patterns can be specific to a particular biological or disease state, including cancer[16]. In previous work, we adopted lncRNA expression profiles to identify a subgroup of luminal muscle-invasive bladder cancer (MIBC) with favourable tumor biology[17]. These data suggest that there might be an inherent utility in lncRNA expression profiles for classifying subsets of bladder cancer with unique biology in other disease stages or in variant histologies.

In the present study, we aimed to investigate whether lncRNA expression profiles could be adopted for the identification and molecular characterization of high-grade T1 micropapillary bladder cancer.

METHODS

Patient populations
For the present study, we selected cases with complete clinical data from a previously published retrospective cohort[4], resulting in a total of 84 high-grade T1 bladder tumors, including 15 cases with high micropapillary component (primary micropapillary) and 7 cases with secondary micropapillary histology. Archival FFPE tissue specimens were collected at the time of TURBT confirming the diagnosis of high-grade T1 bladder cancer. Informed consent was provided by each subject and use of the tissue was approved by the Ethics Committee of the Dana Farber Cancer Institute, Hospital del Mar and Hospital Vall d'Hebron and all research complied with local ethics guidelines.[4] mRNA data of the T1 cohort has been uploaded at GEO (GEO: GSE136401).

Pathology
All specimens were reviewed by two pathologists, who assessed percentage of micropapillary histology, depth of tumor invasion, lymphovascular invasion and presence of carcinoma-in-situ (*CIS*). A two-tier system was used to assess the depth of lamina propria invasion; T1a when tumor involved the subepithelial connective tissue superficial to muscularis mucosae (MM); T1b when tumor was found at the level or beyond the MM. Of note, patients with T1b substaging underwent re-TURBT after BCG therapy as per institutional approved protocol. Primary micropapillary cases contained >50% micropapillary component. For the secondary micropapillary cases, only tumors with at least a >10% of micropapillary component, measured by a semiquantitative (visual) estimation of the micropapillary component percentage, were included. However, when coring, the micropapillary component could not have been included. Only high-grade T1 tumors with a visible, clearly identifiable and disease-free muscularis propria and micropapillary high-grade T1 on which both pathologists concur on the presence of micropapillary component (>10%) were included in this study. Pathology review of tumor regions of interest (ROI), where tumor cellularity was in excess of 70%, were annotated by the two pathologists. Up to 5x0.6mm cores were punched from FFPE tissue blocks within the tumor-rich ROI. Every effort was made when sampling micropapillary areas to ensure that micropapillary histology was present throughout the thickness of the tissue core.

Gene expression profiling
RNA extraction, quality control, quantification using the Quant-iT RiboGreen assay and RNA-Seq library preparation have been previously described for this cohort. For the current study, gene expression profiles were matched to the Decipher array platform using quantile normalization (R package preprocessCore). To facilitate future applicability of our results to multiple RNA expression platforms, we pre-selected genes that were present at the Quant-iT RiboGreen assay, the Decipher bladder assay and the Illumina HiSeq assay (The Cancer Genome Atlas), as the initial gene list for our analyses (19970 genes). The Cancer Genome Atlas cohort has been previously described[17].

Unsupervised clustering using lncRNAs
For unsupervised clustering analysis (R package ConsensusClusterPlus), the normalized gene expression data for N=84 samples (T1 cohort), including expression of 1271 available lncRNAs, was pre-processed by filtering low-variance genes, selecting the 160 lncRNAs with the highest median absolute deviation (R package MADS) (**Table S1**). The expression clustering analysis was performed by a consensus partitioning around medoids (PAM) approach, using Pearson correlations as the similarity metric, the Ward algorithm for clustering, running 5,000 iterations with 0.95 random fraction of samples used in each iteration.

Classification of tumors into molecular mRNA subtypes:
To assign tumors to the consensus bladder cancer (basal/squamous (Ba/Sq), luminal papillary (LumP), luminal non-specified (LumNS), luminal unstable (LumU), stroma-rich, and neuroendocrine-like (NE-like)) molecular subtypes[10], we downloaded and applied the centroid-based models as described[10].

Gene expression analyses
We created heatmaps and boxplots to visualize differences between tumors from lncRNA and mRNA subtypes, regarding the expression of I) individual genes, II) gene signatures, and III) hallmark gene sets

from the molecular signature database hallmark gene set collection (MSigDB). We applied our lncRNA-based genomic classifier that was developed to identify a subgroup of luminal MIBC tumors characterized by distinct biologic activity, including high FGFR3 activity. FGFR3 activity, hedgehog signaling, basal and luminal signature score calculations have been described previously[17][8]. We employed R package Limma to perform differential gene expression analyses and created volcano plots using R package EnhancedVolcano.

Discovery of a micropapillary-like classifier

The current cohort was used to train a genomic signature to predict aggressive bladder cancer that is associated with micropapillary biology, as was identified by lncRNA consensus cluster 1 (LC1). The initial gene list (19970 features) was used for differential gene expression analysis within the training cohort, comparing LC1 (N=16) to the rest of the cohort (N=68). Using R package Limma, we first selected the top 500 differentially expressed features (genes). Next, these 500 genes were further filtered based on an expression level standard deviation of greater than 0.3, resulting in the final input set of 69 genes. These genes were used for the training of a 100x10-fold cross-validated, lasso (α=0) penalized logistic regression model (R package GLMNET) in the training cohort. For this model, input gene features were grouped into coefficients by cutting a dendrogram that was constructed by clustering all input genes, making the model more generalizable for external dataset testing. This signature was evaluated using distribution density plots where a probability threshold of 0.44 was selected for classifying positive predicted model cases (MP+). As a last step, we applied the signature to the testing (N=202, TCGA luminal) cohort. The TCGA RNA-seq data was available as part of the TCGA research network (http://cancergenome.nih.gov/).

Statistical analyses

Statistical analyses were performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria). Patient and tumor characteristics were compared between subgroups by using X^2 tests and two-sided Wilcoxon rank-sum tests. P values in boxplot figures represent results of Kruskal-Wallis rank sum tests when comparing multiple groups, and Wilcoxon rank sum tests when comparing two groups. Primary endpoints were overall, progression-free and high-grade recurrence-free survival, calculated as the date of high-grade T1 disease at TURBT till date of death from any cause, progression or recurrence, respectively. Disease progression was defined as progression to MIBC or development of metastasis. Disease recurrence was defined as recurrence of high-grade NMIBC. Patients who were lost to follow-up were censored at the date of last contact. The Kaplan-Meier method was used to estimate the statistical significance of differences between survival curves for patient subgroups, using the log-rank test. Statistical code was either trivial or is available based upon reasonable request (partner@decipherbio.com).

RESULTS

The clinical characteristics of study cohort are reported in **Table 1**. Patient age, sex, lamina propria invasion, and presence of carcinoma-in-situ (*CIS)*, were similar for both micropapillary and conventional high-grade T1 bladder tumors. P-values in **Table 1** represent comparisons between micropapillary and conventional histology, using X^2 tests for categorical variables and two-sided Wilcoxon rank-sum tests for continuous variables.

Micropapillary tumors have unique gene expression profiles

Application of the consensus molecular subtyping model resulted in most of the tumors being classified as a 'luminal' subtype (LumP, LumU, or LumNS). Significant differences in the distribution of molecular subtypes between urothelial and micropapillary tumors ($p<0.001$) were found, with micropapillary tumors being enriched for the 'LumU' class (9/15) (**Table 1**). Both urothelial and micropapillary tumors showed higher luminal signature scores, with lower scores for the basal signature (**Figure S1**). However, no significant differences were observed for either signature between groups. Further, we observed significant differences in progression-free (PFS) or overall survival (OS) when stratifying by subtype (Ba/Sq, LumP, LumNS, LumU, Stroma-rich; **Figure S2**).

Differential gene expression analyses comparing micropapillary and urothelial tumors found many luminal markers were upregulated in primary micropapillary cases, including *UPK1B, UPK3A* and *KRT20*. Genes associated with micropapillary disease (i.e., *MUC1*, *MUC4*) were also upregulated in the primary micropapillary cases (**Figure 1A**). Interestingly, many of the downregulated genes were lncRNAs. Comparing only the lncRNA profiles, we found several lncRNAs previously associated with bladder cancer were significantly downregulated in micropapillary tumors, including *UCA1[18], LINC00152[19], and MALAT1[20]* (**Figure 1B**).

A long non-coding RNA-based genomic classifier identifies FGFR3 active tumors

In previous work, we developed a lncRNA-based single-sample genomic classifier to identify muscle-invasive bladder tumors with activated FGFR3 (FGFR3+) and less aggressive tumor biology[17]. Application of the lncRNA classifier to the study cohort revealed seven FGFR3+ cases with enhanced SHH activity, higher p53 pathway activity, lower FGFR3 pathway activity and lower EMT scores. Notably, none of the micropapillary cases were classified as FGFR3+, but showed decreased SHH and p53 signature scores, with higher EMT activity (**Figure 2**). The OS of patients with FGFR3+ tumors were similar to the rest of the cohort ($p=0.94$, **Figure S3**).

LncRNA-based unsupervised clustering identifies aggressive micropapillary cases

Unsupervised consensus clustering of 160 highly variant lncRNAs resulted in a three-cluster consensus solution (LC1-3) (**Figure S4A**). Comparing the clustering solution with the cohort histology, we found LC1 contained 16 cases, of which 12 were micropapillary histology, meaning LC1 was significantly enriched for primary micropapillary histology cases ($p<0.001$). The remaining four LC1 cases were all T1b disease. The remaining three micropapillary histology cases were among LC2 (**Figure S4B**). Only one of the seven conventional urothelial carcinoma cases with secondary micropapillary histology was clustered in LC1. Given the enrichment of primary micropapillary histology in LC1, we decided to characterize this cluster in detail, which we named the "Micropapillary LncRNA Consensus Cluster (MP-LncCC)".

Survival analysis of the cohort stratified by histological subtype (micropapillary vs urothelial), found no significant difference in outcome for either recurrence- or progression-free survival ($p=0.95$ and $p=0.4$, respectively), but significant differences in overall survival ($p=0.04$, **Figure 3A-C**). When stratified by MP-LncCC differences in recurrence-free survival were not observed ($p=0.49$), but MP-LncCC cases had significantly worse PFS and OS ($p<0.001$ for both; **Figure 3D-F**). Unsupervised lncRNA-based clustering showed results that were superior to primary histological characteristics with respect to progression free survival but were similar with respect to overall survival. **Table S2** lists Cox proportional hazard models

for recurrence-free, progression-free, and overall survival outcomes, revealing MP-LncCC as a significant predictor for progression-free survival on multivariable analyses.

To characterize the MP-LncCC tumors, we used a set of genes previously associated with micropapillary disease[3], finding similar expression patterns (**Figure 4**). Genes such as *MUC1*, *MESP1* and *CLDN3* were upregulated in the MP-LncCC group, but down-regulated in the remaining tumors. Other genes, such as *KRT5*, were downregulated in the MP-LncCC cases, but were highly upregulated in the remaining tumors, particularly in those classified as Ba/Sq by the consensus classifier (**Figure 4**).

A micropapillary-like gene expression signature

To provide clinical utility to our findings, we developed a single-sample genomic signature to identify aggressive bladder cancer that is associated with micropapillary biology (MP+). To evaluate the performance of this micropapillary signature, we downloaded gene expression data for the TCGA cohort (MIBC, Radical Cystectomy, n=405). As most of the cases in the training cohort had a luminal expression profile, we selected only the luminal tumors (n=202) from the TCGA cohort for testing the model. Here, we identified 7/202 (3.4%) MP+ cases, which were all classified as LumU. The gene expression patterns of micropapillary-associated genes for the MP+ cases were consistent with our previous findings (**Figure 5A**). The MP+ cases also had lower P53-, SHH- and FGFR3 signature scores, but higher EMT activity (**Figure 5B-E**). Importantly, patients with MP+ tumors showed significantly worse survival than other luminal cancer patients (Rest) or FGFR3+ cases ($p<0.001$; **Figure 5F**).

DISCUSSION

Molecular subtyping of bladder cancer using transcriptome profiling has shown promise for predicting outcomes and response to therapy[5-7]. In 2019, the Bladder Cancer Molecular Taxonomy Group put significant effort to harmonize existing subtyping models, reporting on a consensus classifier comprised of six subtypes of MIBC[10]. Molecular subtyping of NMIBC is also underway, where these tumors appear to be enriched for the luminal subtype[11, 12]. However, the majority of these efforts have invested in mRNA-based subtyping, which actually represents only a minority of the full transcriptome[15]. In 2017, the TCGA demonstrated the non-coding transcriptome could be used to stratify the luminal subtype into subgroups with distinct prognoses[6]. This finding was subsequently confirmed, resulting in the development of a lncRNA-based single-sample classifier which identifies luminal tumors with improved outcomes[17]. Taken together, these studies suggest the lncRNA transcriptome may provide a sensitive read on specific luminal disease states.

Reports describing the molecular characterization of variant tumors, including micropapillary, are somewhat limited[21, 22]. In 2016, Guo *et al.* reported on a set of unique molecular features associated with micropapillary MIBC, including differences in cellular pathways such as transformation, cell cycle regulation, DNA damage repair and signal transduction[3]. Likewise, HGT1 micropapillary bladder cancer has a unique transcriptome profile, with increased immune, metabolic and cell cycle pathway activity[4]. Moreover, molecular subtyping, using either RNA- or IHC-based methods, has shown the majority of micropapillary bladder tumors are classified as a luminal subtype[4, 21, 22].

In the present study we focused on the lncRNA transcriptome, building on our previous work in identifying subsets of luminal bladder tumors using a unique lncRNA-based FGFR3+ classifier[17]. This classifier identifies a subset of luminal MIBC tumors (called FGFR3+ by the model) with improved outcomes that have enhanced FGFR3 activity and improved outcomes. When this classifier was applied to the current cohort, we did not identify any micropapillary cases as FGFR3+, as was anticipated given the unique molecular profile for micropapillary tumors. Using differential gene expression analysis, we found that selected lncRNAs were downregulated in micropapillary compared to urothelial tumors. These data suggested that micropapillary tumors may have a unique lncRNA expression profile with potential to serve as a novel biomarker.

Using consensus clustering we identified a group of tumors (LC1) showing gene expression patterns consistent with micropapillary disease, including upregulation of *MUC1* and *MESP1[3]*. Most of these tumors had primary micropapillary histology with four only being conventional urothelial carcinoma. Survival analysis comparing LC1 revealed poor outcomes suggesting lncRNA expression patterns may identify tumors with aggressive clinical characteristics, independent of the histological presentation. Other studies on bladder cancer with variant histology have reported similar findings. The most well characterized of these is the NE-like subtype, where the NE-like tumors have genomic and clinical characteristics that are nearly identical to histological small cell tumors, yet typically present as conventional urothelial carcinoma[23, 24]. Taken together, these data may support the hypothesis that a variant specific expression profile may precede the presence of the variant morphology itself[25].

As a proof of concept, we adopted the lncRNA transcriptome to develop a genomic classifier for identifying aggressive MP-like tumors (MP+). This classifier was tested on a subset of luminal (LumP, LumNS, LumU) tumors from the TCGA cohort[6], finding 7/201 cases were MP+. These MP+ tumors had micropapillary-like gene expression patterns, lower FGFR3, SHH and p53 signature scores, and extremely poor outcomes. While further development and testing are necessary for clinical application, these initial findings demonstrate the potential utility of the lncRNA transcriptome for the development of novel biomarkers identifying aggressive micropapillary bladder cancer.

The retrospective nature and lack of a validation cohort are among the limitations of this study. Tumors with micropapillary histology are also rare, limiting the sample size for development of a validated classifier, which limits the immediate clinical utility of this model for prospective validation studies.

CONCLUSIONS

In summary, using the lncRNA transcriptome, we identified a subgroup aggressive T1 bladder cancer that was associated with micropapillary histology. Genomic analyses revealed significant downregulation of lncRNAs for micropapillary bladder cancer, a variant which is consistently classified as a luminal molecular subtype. These data demonstrate further utility in lncRNA transcriptome profiling for molecular bladder cancer classification.

REFERENCES

[1] Abufaraj M, Foerster B, Schernhammer E, Moschini M, Kimura S, Hassler MR, et al. Micropapillary Urothelial Carcinoma of the Bladder: A Systematic Review and Meta-analysis of Disease Characteristics and Treatment Outcomes. Eur Urol. 2019;75:649-58.
[2] Witjes JA, Babjuk M, Bellmunt J, Bruins HM, De Reijke TM, De Santis M, et al. EAU-ESMO Consensus Statements on the Management of Advanced and Variant Bladder Cancer-An International Collaborative Multistakeholder Effort(dagger): Under the Auspices of the EAU-ESMO Guidelines Committees. Eur Urol. 2020;77:223-50.
[3] Guo CC, Dadhania V, Zhang L, Majewski T, Bondaruk J, Sykulski M, et al. Gene Expression Profile of the Clinically Aggressive Micropapillary Variant of Bladder Cancer. Eur Urol. 2016;70:611-20.
[4] Bowden M, Nadal R, Zhou CW, Werner L, Barletta J, Juanpere N, et al. Transcriptomic analysis of micropapillary high grade T1 urothelial bladder cancer. Sci Rep. 2020;10:20135.
[5] Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell. 2014;25:152-65.
[6] Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;171:540-56 e25.
[7] Necchi A, Raggi D, Gallina A, Ross JS, Fare E, Giannatempo P, et al. Impact of Molecular Subtyping and Immune Infiltration on Pathological Response and Outcome Following Neoadjuvant Pembrolizumab in Muscle-invasive Bladder Cancer. Eur Urol. 2020;77:701-10.
[8] Seiler R, Ashab HAD, Erho N, van Rhijn BWG, Winters B, Douglas J, et al. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. Eur Urol. 2017;72:544-54.
[9] Sjodahl G, Eriksson P, Liedberg F, Hoglund M. Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. J Pathol. 2017;242:113-25.
[10] Kamoun A, de Reynies A, Allory Y, Sjodahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020;77:420-33.
[11] Robertson AG, Groeneveld CS, Jordan B, Lin X, McLaughlin KA, Das A, et al. Identification of Differential Tumor Subtypes of T1 Bladder Cancer. Eur Urol. 2020;78:533-7.
[12] Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtroder K, et al. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. Nat Commun. 2021;12:2301.
[13] de Jong JJ, Zwarthoff EC. Molecular and clinical heterogeneity within the luminal subtype. Nat Rev Urol. 2020;17:69-70.
[14] Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. 2012;9:703-19.
[15] Gibb EA, Vucic EA, Enfield KS, Stewart GL, Lonergan KM, Kennett JY, et al. Human cancer long non-coding RNA transcriptomes. PLoS One. 2011;6:e25915.
[16] Nguyen Q, Carninci P. Expression Specificity of Disease-Associated lncRNAs: Toward Personalized Medicine. Curr Top Microbiol Immunol. 2016;394:237-58.
[17] de Jong JJ, Liu Y, Robertson AG, Seiler R, Groeneveld CS, van der Heijden MS, et al. Long non-coding RNAs identify a subset of luminal muscle-invasive bladder cancer patients with favorable prognosis. Genome Med. 2019;11:60.
[18] Ghafouri-Fard S, Taheri M. UCA1 long non-coding RNA: An update on its roles in malignant behavior of cancers. Biomed Pharmacother. 2019;120:109459.
[19] Xian-Li T, Hong L, Hong Z, Yuan L, Jun-Yong D, Peng X, et al. Higher Expression of Linc00152 Promotes Bladder Cancer Proliferation and Metastasis by Activating the Wnt/beta-Catenin Signaling Pathway. Med Sci Monit. 2019;25:3221-30.
[20] Fan Y, Shen B, Tan M, Mu X, Qin Y, Zhang F, et al. TGF-beta-induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. Clin Cancer Res. 2014;20:1531-41.
[21] Yang Y, Kaimakliotis HZ, Williamson SR, Koch MO, Huang K, Barboza MP, et al. Micropapillary urothelial carcinoma of urinary bladder displays immunophenotypic features of luminal and p53-like subtypes and is not a variant of adenocarcinoma. Urol Oncol. 2020;38:449-58.

[22] Zinnall U, Weyerer V, Comperat E, Camparo P, Gaisa NT, Knuechel-Clarke R, et al. Micropapillary urothelial carcinoma: evaluation of HER2 status and immunohistochemical characterization of the molecular subtype. Hum Pathol. 2018;80:55-64.
[23] Batista da Costa J, Gibb EA, Bivalacqua TJ, Liu Y, Oo HZ, Miyamoto DT, et al. Molecular Characterization of Neuroendocrine-like Bladder Cancer. Clin Cancer Res. 2019;25:3908-20.
[24] Grivas P, Bismar TA, Alva AS, Huang HC, Liu Y, Seiler R, et al. Validation of a neuroendocrine-like classifier confirms poor outcomes in patients with bladder cancer treated with cisplatin-based neoadjuvant chemotherapy. Urol Oncol. 2020;38:262-8.
[25] Genitsch V, Kollar A, Vandekerkhove G, Blarer J, Furrer M, Annala M, et al. Morphologic and genomic characterization of urothelial to sarcomatoid transition in muscle-invasive bladder cancer. Urol Oncol. 2019;37:826-36.

FIGURE AND TABLE LEGENDS

Figure 1: Differential gene expression differences between tumors with micropapillary histology and conventional urothelial carcinoma. (A) Protein-coding genes and lncRNAs and (B) LncRNAs only. Dark grey circles (NS) indicate non-significant genes, yellow circles (LogFC) indicate genes with a logFC lower than -0.35 or higher than 0.35 but with an adjusted p>0.05. Light grey circles are genes with a logFC between -0.35 and 0.35 with an adjusted p<0.05. Orange circles indicate genes with a logFC <-0.35 or >0.35 with an adjusted p of<0.05. FC = fold-change.

Figure 2: Biological pathways differentially regulated between FGFR3 active (FGFR3+) urothelial (N=7), micropapillary (N=15) and remaining urothelial tumors (N=62). (A) SHH-BMP pathway activity, (B) p53 hallmark activity, (C) FGFR3 signature score, and (D) EMT hallmark activity. Median values [IQR] for (A) SHH-BMP pathway activity were 0.32 [0.18, 0.38] for FGFR3 active, 0.01 [-0.03, 0.11] for micropapillary and 0.18 [0.09, 0.32] for urothelial tumors. Median values [IQR] for (B) p53 hallmark activity were 0.44 [0.21, 0.48] for FGFR3 active, 0.13 [-0.02, 0.26] for micropapillary and 0.42 [0.02, 0.73] for urothelial tumors. Median values [IQR] for (C) FGFR3 signature score were 0.42 [0.34, 0.51] for FGFR3 active, 0.07 [0.05, 0.14] for micropapillary and 0.32 [0.11, 0.43] for urothelial tumors. Median values [IQR] for (D) EMT hallmark activity were -0.39 [-0.69, -0.29] for FGFR3 active, 0.36 [0.02, 0.56] for micropapillary and 0.14 [-0.40, 0.56] for urothelial tumors. P values in boxplot figures represent results of comparing three groups by Kruskal-Wallis rank sum tests.

Figure 3: Survival analysis stratified by micropapillary histology for (A) recurrence-free survival, (B) progression-free survival, and (C) overall survival. Panels D-F follow the same order but are stratified by MP-LncCC. P-values represent comparisons between survival curves using log-rank tests.

Figure 4: Biological characterization of the MP-lncRNA cluster (MP-LncCC) using selected micropapillary marker genes. Both the consensus (Ba/Sq, LumNS, LumP, LumU, Stroma-rich) and TCGA subtypes (basal squamous, luminal papillary, luminal, luminal infiltrated) are indicated in the covariate tracks. Additional covariate tracks include micropapillary patterns, microstaging, and micropapillary histology.

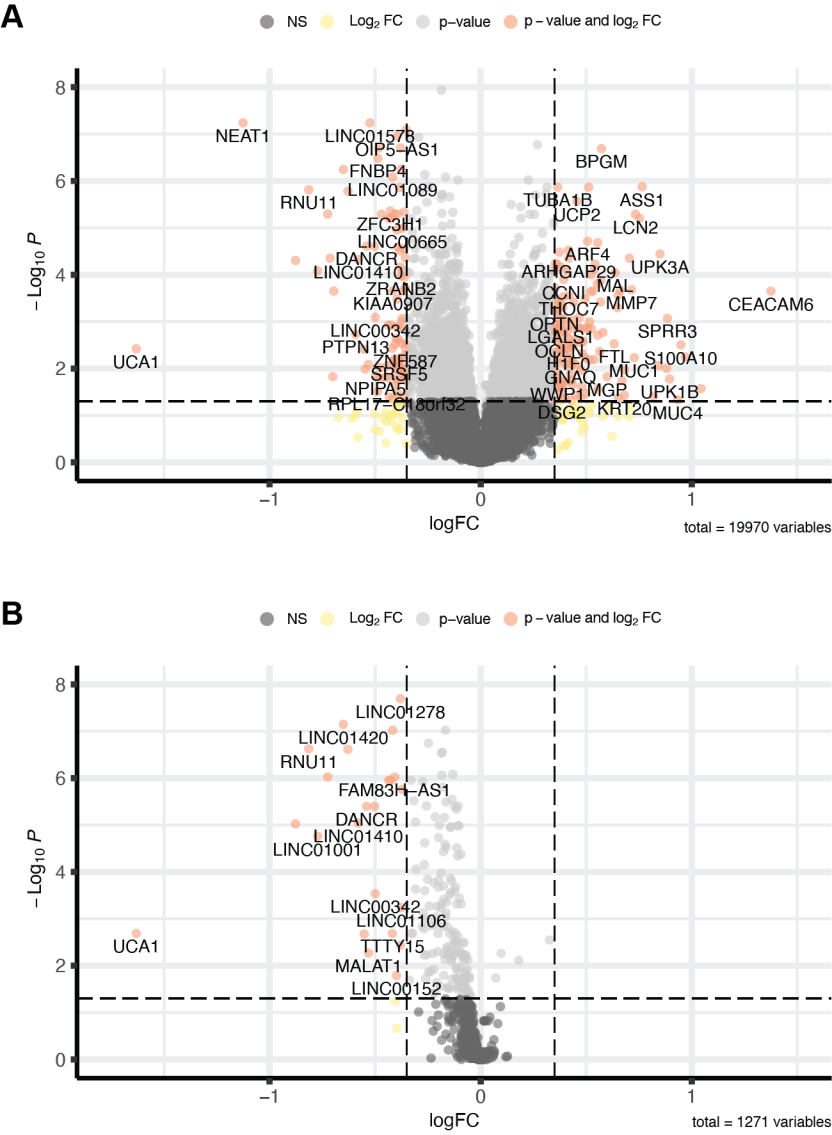
Figure 5: Biological and clinical characterization of the FGFR3+ and MP+ cases in the subset of N=202 luminal tumors from the TCGA cohort. (A) Expression patterns of selected micropapillary genes. The consensus subtypes (Ba/Sq, LumNS, LumP, LumU, Stroma-rich), FGFR3 mutations and fusions, and TP53 mutations are indicated in the covariate tracks. Signature scores for FGFR3+ (N=54), MP+ (N=7) and the rest of the cohort (N=141) (B) SHH-BMP pathway, (C) p53 pathway, (D) FGFR3 pathway, and (E) epithelial-mesenchymal transition (EMT). (F) Overall survival KM plots.

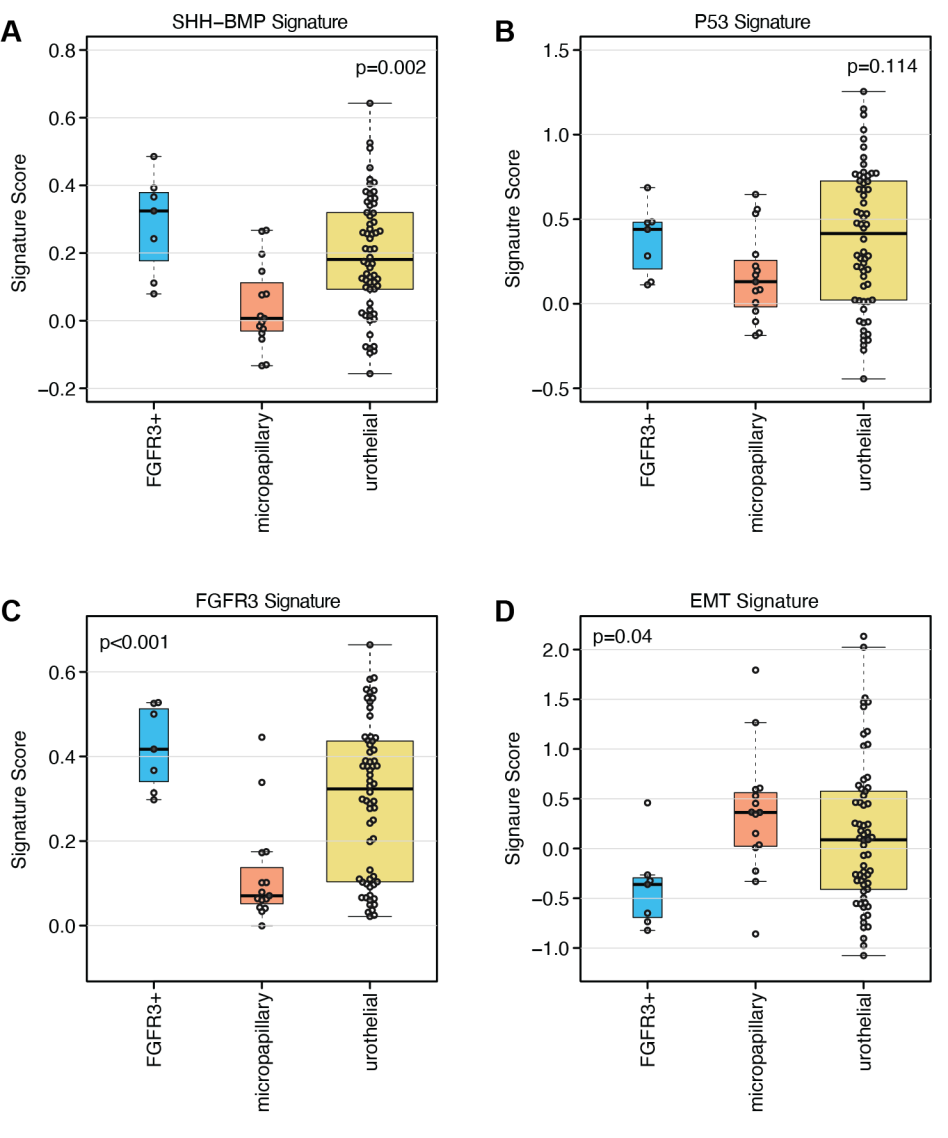
Table 1. Clinicopathological characteristics of the high-grade T1 study cohort. P-values represent comparisons between micropapillary and conventional histology, using X^2 tests for categorical variables and two-sided Wilcoxon rank-sum tests for continuous variables.

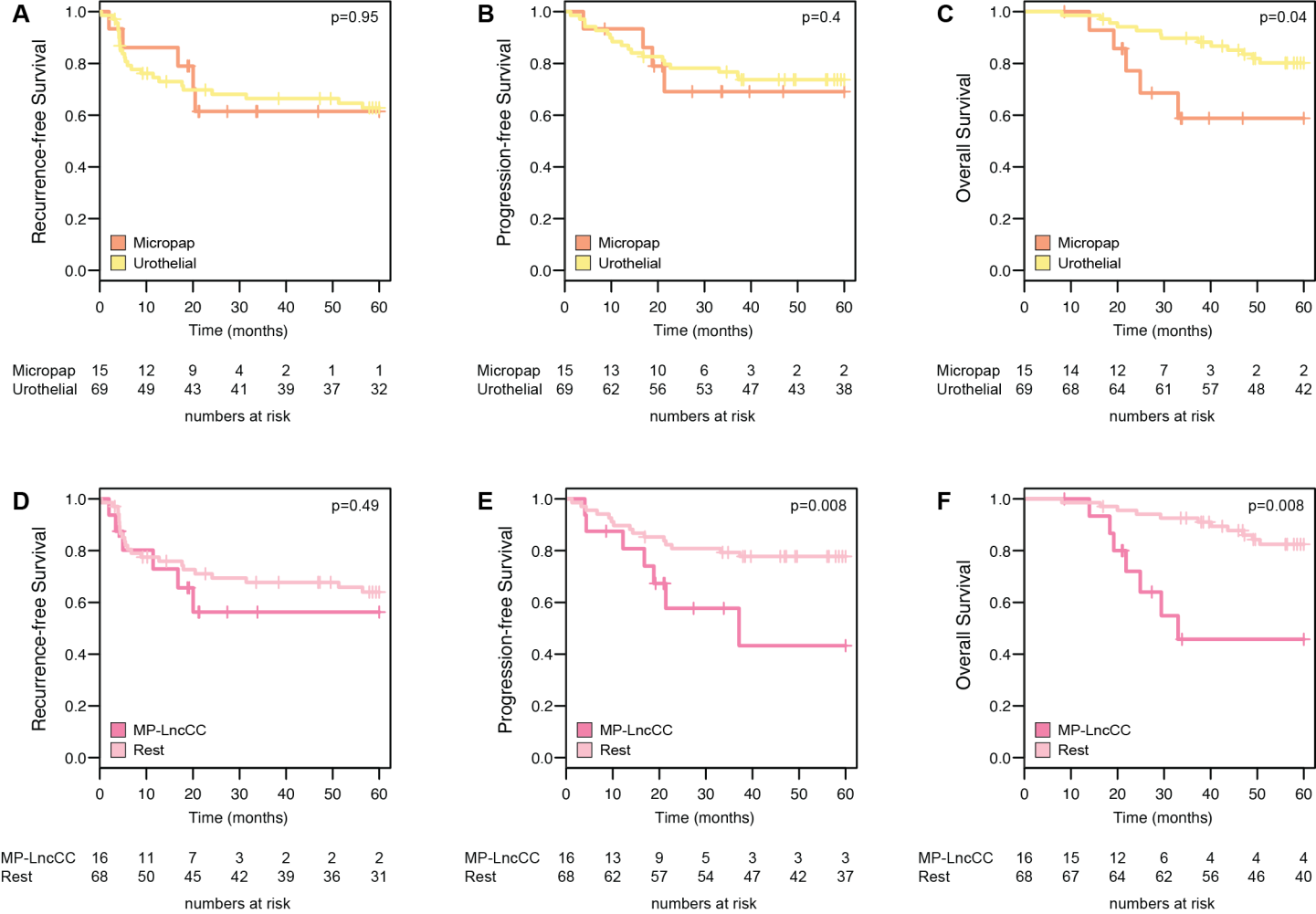
TABLES

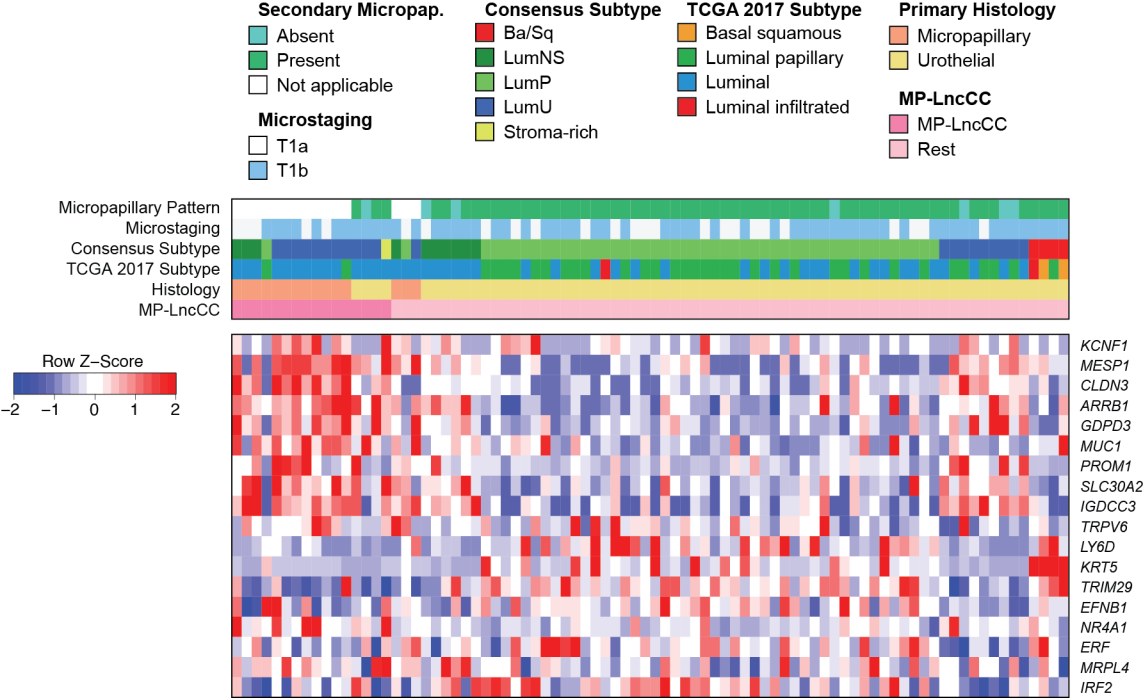
Table 1. Clinicopathological characteristics of the high-grade T1 study cohort. P-values represent comparisons between micropapillary and conventional histology, using X^2 tests for categorical variables and two-sided Wilcoxon rank-sum tests for continuous variables.

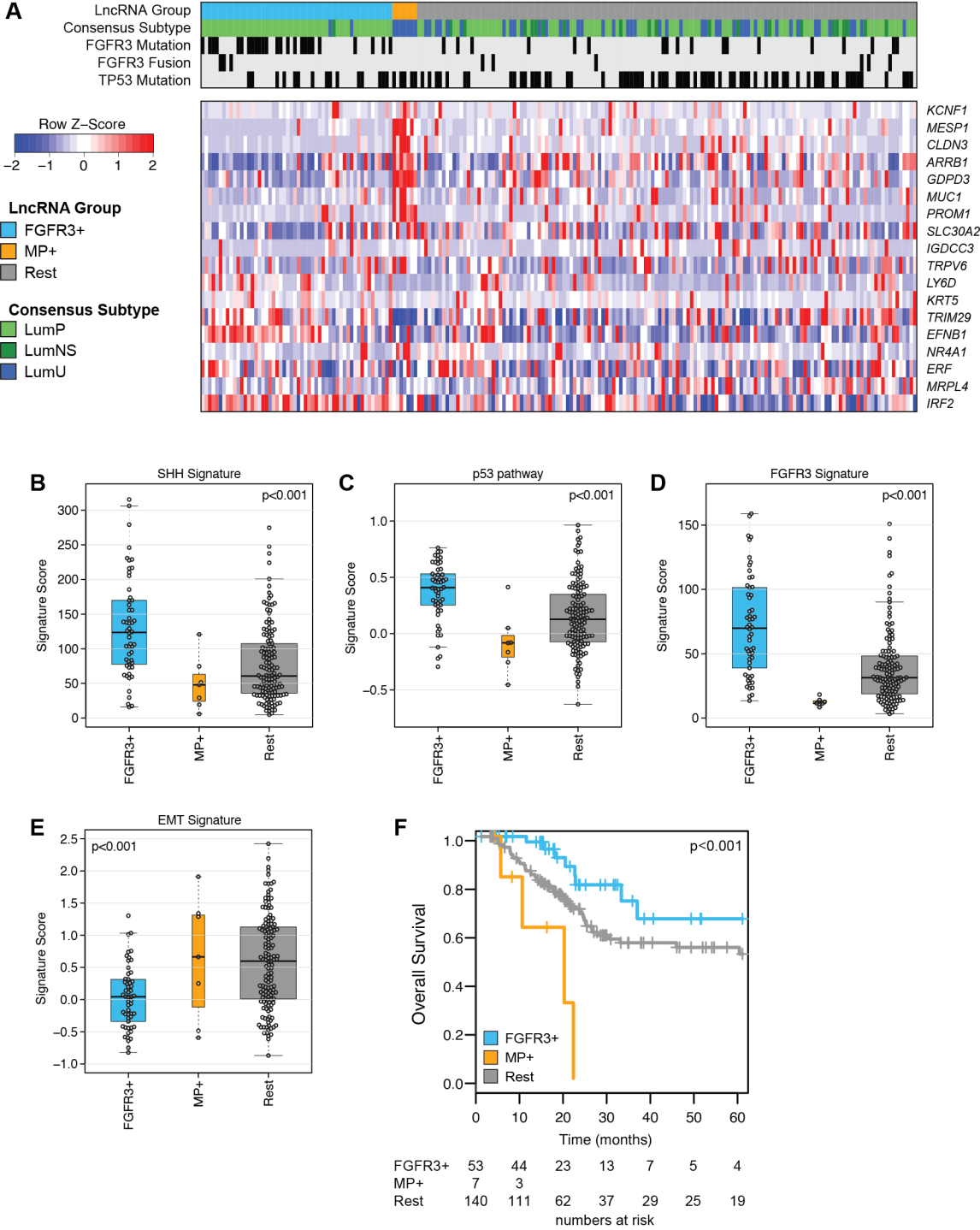
Variables		**Conventional n (%)**	**Micropapillary n (%)**	*p-value*
Total		69 (100%)	15 (100%)	
Age	Median [IQR]	69 [61-74]	70 [61-78]	0.54
Sex	Female	10	2	
	Male	59	13	1
Lamina propria invasion	T1a	19	6	
	T1b	50	9	0.52
Secondary MP histology	Present	7	.	
	Absent	62	.	
Carcinoma In Situ	Absent	49	8	
	Present	20	6	0.48
	Unknown	*0*	*1*	
Consensus subtype	Basal/squamous	4	0	
	Luminal-papillary	46	2	
	Luminal non-specified	6	4	
	Luminal unstable	12	9	
	Stroma-rich	1	0	<0.001


A
NS
Log2 FC
p−value
p − value and log2 FC
NEAT1
LINC01578
OIP5−AS1
FNBP4
LINC01089
RNU11
ZFC3H1
LINC00665
DANCR
LINC01410
ZRANB2
KIAA0907
LINC00342
PTPN13
ZNF587
SRSF5
NPIPA5
RPL17−C18orf32
UCA1
BPGM
TUBA1B
ASS1
UCP2
LCN2
ARF4
ARHGAP29
UPK3A
MAL
CCNI
MMP7
CEACAM6
THOC7
OPTN
SPRR3
LGALS1
OCLN
FTL
S100A10
HIF0
MUC1
GNAQ
WWP1
MGP
UPK1B
DSG2
KRT20
MUC4
−Log10 P
logFC
total = 19970 variables
B
NS
Log2 FC
p−value
p − value and log2 FC
LINC01278
LINC01420
RNU11
FAM83H−AS1
DANCR
LINC01410
LINC01001
LINC00342
LINC01106
UCA1
TTTY15
MALAT1
LINC00152
−Log10 P
logFC
total = 1271 variables


A
SHH−BMP Signature
p=0.002
Signature Score
0.8
0.6
0.4
0.2
0.0
−0.2
FGFR3+
micropapillary
urothelial
B
P53 Signature
p=0.114
Signautre Score
1.5
1.0
0.5
0.0
−0.5
FGFR3+
micropapillary
urothelial
C
FGFR3 Signature
p<0.001
Signature Score
0.6
0.4
0.2
0.0
FGFR3+
micropapillary
urothelial
D
EMT Signature
p=0.04
Signaure Score
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
FGFR3+
micropapillary
urothelial


A
p=0.95
Recurrence-free Survival
Micropap
Urothelial
Time (months)
Micropap 15 12 9 4 2 1 1
Urothelial 69 49 43 41 39 37 32
numbers at risk
B
p=0.4
Progression-free Survival
Micropap
Urothelial
Time (months)
Micropap 15 13 10 6 3 2 2
Urothelial 69 62 56 53 47 43 38
numbers at risk
C
p=0.04
Overall Survival
Micropap
Urothelial
Time (months)
Micropap 15 14 12 7 3 2 2
Urothelial 69 68 64 61 57 48 42
numbers at risk
D
p=0.49
Recurrence-free Survival
MP-LncCC
Rest
Time (months)
MP-LncCC 16 11 7 3 2 2 2
Rest 68 50 45 42 39 36 31
numbers at risk
E
p=0.008
Progression-free Survival
MP-LncCC
Rest
Time (months)
MP-LncCC 16 13 9 5 3 3 3
Rest 68 62 57 54 47 42 37
numbers at risk
F
p=0.008
Overall Survival
MP-LncCC
Rest
Time (months)
MP-LncCC 16 15 12 6 4 4 4
Rest 68 67 64 62 56 46 40
numbers at risk


Secondary Micropap.
Absent
Present
Not applicable
Microstaging
T1a
T1b
Consensus Subtype
Ba/Sq
LumNS
LumP
LumU
Stroma-rich
TCGA 2017 Subtype
Basal squamous
Luminal papillary
Luminal
Luminal infiltrated
Primary Histology
Micropapillary
Urothelial
MP-LncCC
MP-LncCC
Rest
Micropapillary Pattern
Microstaging
Consensus Subtype
TCGA 2017 Subtype
Histology
MP-LncCC
Row Z−Score
−2
−1
0
1
2
KCNF1
MESP1
CLDN3
ARRB1
GDPD3
MUC1
PROM1
SLC30A2
IGDCC3
TRPV6
LY6D
KRT5
TRIM29
EFNB1
NR4A1
ERF
MRPL4
IRF2


A
LncRNA Group
Consensus Subtype
FGFR3 Mutation
FGFR3 Fusion
TP53 Mutation
Row Z−Score
−2 −1 0 1 2
LncRNA Group
FGFR3+
MP+
Rest
Consensus Subtype
LumP
LumNS
LumU
KCNF1
MESP1
CLDN3
ARRB1
GDPD3
MUC1
PROM1
SLC30A2
IGDCC3
TRPV6
LY6D
KRT5
TRIM29
EFNB1
NR4A1
ERF
MRPL4
IRF2
B
SHH Signature
p<0.001
Signature Score
FGFR3+
MP+
Rest
C
p53 pathway
p<0.001
Signature Score
FGFR3+
MP+
Rest
D
FGFR3 Signature
p<0.001
Signature Score
FGFR3+
MP+
Rest
E
EMT Signature
p<0.001
Signature Score
FGFR3+
MP+
Rest
F
p<0.001
Overall Survival
FGFR3+
MP+
Rest
Time (months)
FGFR3+ 53 44 23 13 7 5 4
MP+ 7 3
Rest 140 111 62 37 29 25 19
numbers at risk