FGF- and FGFR-coding gene alterations in urothelial cancers

Short analysis report

Piotr Tymoszuk for the Department of Urology, Medical University of Innsbruck

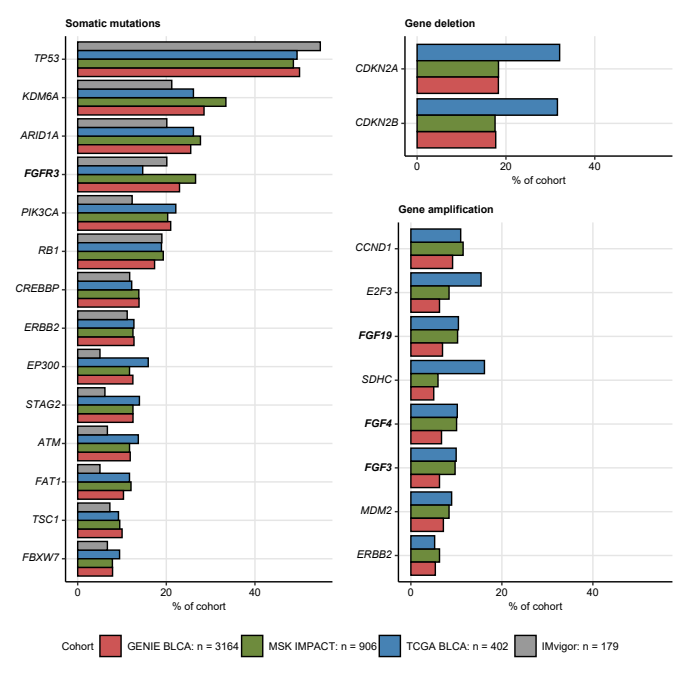
2024-05-17

# Analysis summary

* Fibroblast growth factor receptor (FGFR) signaling has emerged as a promising therapy target in urothelial cancers with approval of the pan-FGFR inhibitor erdafitinib. However, effects of genetic alterations and modulation of expression of FGF- and FGFR-coding genes in urothelial carcinoma on tumor biology, therapy response, and clinical outcome are still not fully understood.
* The primary goal of the current project was to characterize frequency, co-occurrence, and clinical implications of genetic alterations in FGF- and FGFR-coding genes in large publicly available cohorts of urothelial cancers. The secondary goal was an attempt to establish subsets of urothelial cancer specimens based on differences in their profiles of somatic mutations and gene copy number variants.
* Four collectives of urothelial cancer patients were evaluated: the GENIE BLCA cohort, an urothelial carcinoma subset of the large GENIE project (n = 3164) (1), the predominantly muscle-invasive MSK IMPACT cohort (906) (2), the muscle-invasive TCGA BLCA collective (n = 402) (3,4), and the IMvigor cohort of late stage, platinum-refractory cancers subjected to an anti-PD-L1 treatment (179) (5). Detailed information on somatic mutations and their effects on protein sequence was provided for the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts. Binarized indexes of presence/absence of somatic mutations in selected genes were available for the IMvigor cohort. Binarized data on gene amplifications and deletions were obtained for the GENIE BLCA, MSK IMPACT, and TCGA BLCA collectives. Gene expression quantification by RNA sequencing and assignment to consensus molecular classes (6) were retrieved for the TCGA BLCA and IMvigor cohorts. Overall survival information was available for the MSK IMPACT, TCGA BLCA, and IMvigor cohorts; data on disease-specific and relapse-free survival was provided for the TCGA BLCA collective. Approximate overall survival was calculated for the GENIE BLCA project participants based on age at enrollment in years and follow-up time in days. Characteristic of the investigated collectives is presented in **Table 1**.
* Somatic mutations of *FGFR3* belonged to the most common genetic features of urothelial cancers and were found in 15 to 27% of cancer samples. Somatic mutations of the remaining *FGFR1*, *FGFR2*, and *FGFR4* receptor genes were less frequent (0 to 2.5% of cancer samples). Amplifications of the *FGF3*, *FGF4*, and *FGF19* genes in urothelial cancers result from multiplication of the 11q13 chromosome region, which codes also for an essential oncogene *CCDN1*. Amplifications of *FGF3*, *FGF4*, and *FGF19* were detected in 6.3 to 10% of specimens (**Figures 1** and **2**, **Tables 2** and **3**).
* A great majority of somatic mutations of *FGFR1*, *FGFR2*, *FGFR3*, and *FGFR4* was classified as missense mutations caused by single nucleotide polymorphisms. Particularly for *FGFR3*, they affected protein residues were frequently located apart from the Ig-like domains responsible for ligand biding and besides the kinase domains vital for activation of downstream signaling (**Figures 3** - **5**, **Table 4**). A more detailed analysis revealed the Ig-like domain 1 and kinase domain as the mutation hot-spots of *FGFR1*. The kinase domain was the prime target of mutations of *FGFR2*. Hinge regions between the Ig-like domains 1 and 2, as well as residues around the membrane-spanning region between the Ig-like domain 3 and the kinase domain were identified as mutational hot-spots of *FGFR3*. In case of *FGFR4*, the Ig-like domain 1, the hinge region between the Ig-like domain 3 and kinase domain, as well as the kinase domain were prime targets of mutations of *FGFR4* (**Figures 6** - **9**). It can be hence speculated, that somatic mutations of FGF receptors may affect both the kinase activity and the conformation of the protein resulting in a higher affinity towards ligands or ligand-independent signaling.
* Co-occurrence and exclusivity of genetic features was investigated by analysis of pairwise Jacquard similarity coefficients for pairs of genetic alterations in urothelial cancer samples. Such analysis revealed little overlap between somatic mutations of *FGFR3* and other common alterations such as *TP53* and *RB1* mutations, amplification of the 11q13 chromosome region hallmarked by copy number variants of *FGF3/4/19* and *CCND1*, as well as deletion of the 9p21 chromosome region with the *MTAP*, *CDKN2A*, and *CDKN2A* genes (**Figures 10** - **Figure 13**).
* Concerning co-regulation of expression of FGF- and FGFR-coding genes, *FGFR1* and *FGFR3* mRNA levels were found to be weakly but significantly negatively associated. Increased expression of *FGFR1* in the cancer tissue correlated significantly with upregulation of genes coding for multiple FGF ligands (*FGF1*, *FGF2*, *FGF5*, *FGF7*, *FGF9*, *FGF10*, and *FGF18*). This phenomenon was absent for *FGFR3* (**Figure 14**). In a differential gene expression analysis of cancers with and without *FGFR3* mutations, we observed that presence of *FGFR3* mutations translated in significantly and strongly increased *FGFR3* mRNA levels. By contrast, cancers with *FGFR3* mutations had significantly lower expression of *FGF2*, *FGF5*, and *FGF7* (**Figure 15**, **Table 5**). This suggests two mechanisms leading to activation of pro-oncogenic FGFR signaling in urothelial carcinoma. In *FGFR3* mutant tumors, FGFR signaling may be triggered by over-expression of FGFR3 protein, improved FGF-affinity, or even in an ligand-independent manner. Alternatively, concomitant over-expression of *FGFR1* with its ligands may activated FGFR signaling in auto- or paracrine manner.
* No consistent differences in total mutation burdens (TMB), counts of protein-affecting somatic mutations, or counts of gene deletions were found for urothelial cancers with and without *FGFR3* mutations. Gene amplification numbers were in turn significantly lower in *FGFR3* mutated neoplasms as compared with *FGF3* wild-type tumors. Effect size of this difference was small (**Figure 16**).
* There were no significant differences in age and gender distribution between patients with wild-type and *FGFR3*-mutated cancers. *FGFR3* mutant cancers tended towards lower pathological stages, this effect was significant in the MSK IMPACT cohort but not in the TCGA BLCA collective (**Table 6**). In the GENIE BLCA and MSK IMPACT cohorts, patients with *FGFR3* mutated tumors demonstrated improved overall survival as compared with bearers of WT neoplasms. This phenomenon was also evident in the TCGA BLCA collective, it was only statistically significant upon exclusion of pT4 cancer patients (entire cohort: p = 0.06, without pT4 cancers: p = 0.05). There were no significant differences in disease-specific or relapse-free survival between the *FGFR3* mutation strata in the TCGA BLCA cohort, which may however result from insufficient number of events of interest (tumor-related deaths and relapses, **Figures 17** and **Figure 18**).
* Analysis of effects of amplification of the 11q13 chromosome region on expression of FGF- and FGFR-coding genes was possible solely for the TCGA BLCA collective. Presence of this genetic aberration resulting in amplification of the *FGF3*, *FGF4*, *FGF19*, and *CCND1* genes translated into lower expression of *FGFR1* and significantly higher levels of *FGF3*, *FGF4*, and *FGF19* mRNA (**Figure 19**). However, effects of those differences was weak and, in case of *FGF3* and *FGF4*, the expression levels were extremely low as compared with other FGF ligand-coding genes. Hence, biological relevance of the 11q13 amplification for FGFR signaling remains unclear.
* Basal/squamous-like cancers followed by luminal papillary tumors constituted the largest consensus molecular classes (6) of urothelial cancers in the TCGA BLCA and IMvigor cohorts (**Figure 20A**). in both cohorts, *FGFR3* mutated cancers were found to be significantly enriched in the papillary luminal class (TCGA: 33%, IMvigor: 52% of specimens). In the TCGA BLCA cohort, amplifications of *FGF3*, *FGF4*, and *FGF19* genes resulting from multiplication of the 11q13 chromosome region were found in luminal papillary (18 to 19%), stroma-rich (9.6 to 12%), and basal/squamous-like cancers (8.8% of specimens, **Figures 20** and **21**, **Table 7**).
* mRNA levels of *FGFR1*, *FGFR3*, *FGF2*, *FGF5*, *FGF7*, *FGF10*, and *FGF18* were found to vary significant between the consensus molecular classes of urothelial carcinomas in both the TCGA BLCA and IMvigor data set (**Table 8**). Expression of *FGFR1* was the highest in stroma-rich and luminal non-specified tumors and the lowest in luminal papillary cancers. In turn, maximal mRNA levels of *FGFR3* were detected in luminal papillary cancers and the lowest expression was observed in basal/squamous- and neuroendocrine-like tumors. Effect size of those differences was large (**Figure 22**). Expression of all significantly regulated *FGF* ligands was the lowest in luminal papillary cancers. *FGF2*, *FGF7*, *FGF10*, and *FGF18* transcript amounts followed the expression pattern of *FGFR1* and were detected at the highest levels in stroma-rich and non-specified luminal cancers. *FGF5* was expressed at the highest levels by basal/squamous-like, neuroendocrine-like and stroma-rich neoplasms. Effect sizes of differential regulation of these FGF-coding genes was large (**Figure 23** and **24**). Collectively, the differential gene expression analysis results support the model of activation of FGFR signaling propose above. While luminal papillary cancers may be driven by *FGFR3* mutations and over-expression resulting in FGFR signaling with little dependence on the ligands, stroma-rich and non-specified luminal cancers may exploit auto- and paracrine signaling fueled by over-expression of *FGFR1* and its ligands.
* By latent class analysis (LCA, called also latent class modeling), we could establish a classification scheme of urothelial cancers based on presence/absence of the most frequent genetic alterations detected in at least 5% of cancer specimens. This classification scheme with six ‘genetic subsets’ was developed in the GENIE BLCA training cohort and technically validated in the MSK IMPACT and TCGA BLCA cohort (**Figures 25** and **26**). The ‘mutRB1’ subset (19 to 26% of specimens) was characterized by highly frequent *TP53*, *RB1* and *ARID1A* mutations, *E2F3* amplifications, and nearly absent *FGFR3* mutations, 11q13 amplification, and 9p21 deletions. The ‘oligoMut’ subset was hallmarked by frequent mutations of *TP53* and the lowest overall mutation count. The key features of ‘hyperMut’ subset were the highest total mutation count, as well as frequent somatic mutations of *ARID1A*, *FGFR3*, *CREBBP*, *ERBB2*, *FBXW7*, and lacking copy number variants of 11q13 and 9p21. The ‘del9p21’ subset was characterized by relatively common deletions of the chromosome region 9p21, amplification of the 11q13 chromosome region, and highly frequent mutations of *TP53* and *KDM6A*. In the ‘mutFGFR3’ subset, mutations of *FGFR3* and *KDM6A*, and deletions of 9p21 were common, while mutations in *TP53*, *RB1*, and *ERBB2*, and gene amplifications were virtually absent. Finally, the ‘ampMDM2’ subset was characterized by the highest percentage of *MDM2*, highly common copy number variants of the 11p13 and 9p21 chromosome regions and essentially lacking mutations of *TP53* and *RB1* (**Figures 27** - **32**, **Table 9**).
* Numbers of mutations differed significantly between the genetic subsets with large effect size and were the highest in hyperMut cancers and the lowest in oligoMut and mutFGFR3 tumors. Gene deletion numbers were significantly higher in del9p21 and ampMDM2 than in the remaining genetic subsets in the GENIE BLCA and MSK IMPACT collectives (large effect size). Amplification counts were the highest in del9p21, ampMDM2, and mutRB1 subsets and the lowest in mutFGFR3 cancers; those differences were large and significant in all collectives (**Figures 33** and **34**).
* We could observe significant differences in expression of *FGFR1*, *FGFR3*, and *FGFR3*, as well as *FGF2*, *FGF5*, FGF7\_, and *FGF17* genes between the genetic subsets of the TCGA BLCA cohort (**Table 10**). Interestingly, mutFGFR3 cancers were found to over-express *FGFR3* and, at the same time, to suppress *FGF2*, *FGF5*, and *FGF7*. The highest expression levels of *FGFR1* were observed in mutRB1 and ampMDM2 genetic subsets; *FGFR2* was found to be expressed in oligoMut, del9p21, and ampMDM2 at the highest levels (**Figures 35** and **36**).
* There were no significant differences in age and gender of cancer patients between the genetic clusters that were shared by the GENIE BLCA, MSK IMPACT, and TCGA BLCA collectives (**Figure 37**). Significantly lower pathological staging was observed in mutFGFR3 subset in the MSK IMPACT cohort, but this difference could not be corroborated in the TCGA BLCA collective. In the TCGA BLCA data set, mutFGFR3 and ampMDM2 genetic subsets comprised predominantly cancers assigned to the luminal papillary consensus class. Basal/squamous-like consensus class cancers predominated in mutRB1, oligoMut, hyperMut, and del9p21 genetic subsets (**Figure 38**). Notably, consensus molecular classes of muscle-invasive urothelial carcinomas were established with bulk cancer transcriptome (4,6). The disparity between the molecular consensus classes and the genetic subsets proposed by us suggests that phenotypically similar cancers may exploit highly dissimilar sets of driver genetic alterations and arise from unrelated evolutionary processes.
* Significant differences in overall survival were observed between the genetic subsets in the GENIE BLCA and MSK IMPACT cohorts with the worst survival in mutRB1 subset and favorable prognosis inferred for mutFGFR3 and hyperMut cancers. No significant differences in overall survival were found in the TCGA cohort, which may, however, attributed to insufficient number of deaths in the genetic subsets (**Figure 39**). As investigated by univariable Cox regression, patients assigned to hyperMut subset were at roughly 50 - 75% significantly lower death risk, and patients in mutFGFR3 were at approximately 43 - 45% significantly lower death risk as compared with mutRB1 subset in both the GENIE BLCA and MSK IMPACT cohort. The general effect of genetic cluster assignment was found to be independent from age, sex, and pathological tumor stage (likelihood ratio test GENIE BLCA: χ²(2) = 52, p < 0.001, MSK IMPACT: χ²(3) = 56, p < 0.001). At the levels of single genetic subsets, hyperMut subset as favorable survival marker remained significant upon adjustment for age, sex, and pathological tumor stage. mutFGFR3 subset assignment was found independent from age and gender in the GENIE BLCA cohort but was confounded by age, sex, and pathological cancer stage in the MSK IMPACT collective (**Figure 40**). We could not discern significant differences in disease-specific and relapse-free survival between the genetic subsets of the TCGA BLCA cohort (not shown).
* Collectively, our analysis results suggest at least two mechanisms of FGFR signaling activation in urothelial cancers: the one exploiting FGFR3 mutations and largely independent on FGF ligands and operative predominantly in papillary cancers, and the other relying on over-expression of FGFR1 and FGF ligands predominantly by stroma-rich and non-specified luminal tumors. Furthermore, we put forward a relatively simple classification scheme of urothelial cancers based on the most common somatic mutations, gene and chromosome region deletions and amplifications established in the large GENIE BLCA collective. The key genetic features of those six genetic clusters such as subset-defining mutations and copy number variants, mutation and variant numbers could be reproduced in the MSK IMAPST and TCGA BLCA, i.e. two independent validation cohorts. Of clinical importance, the genetic subset classification bears consequences for overall survival in an age-, gender-, and stage-independent manner.
* Our analysis bears limitations. Incompleteness of clinical information in the GENIE BLCA cohort precluded an in-depth characteristic of the FGFR3 mutation strata and genetic subsets. Unavailability of exact overall survival in this collective may have interfered with results of significant analyses. Detailed mutation data as well as copy number information were not provided for the IMvigor collective, which made validation of the genetic subsets impossible in this cohort of immunotherapy patients. Finally, gene expression data were not available for the largest GENIE BLCA and MSK IMPACT cohorts, which precluded detailed analyses of differential gene expression, biological phenotype, and consensus molecular classification based on transcriptome.

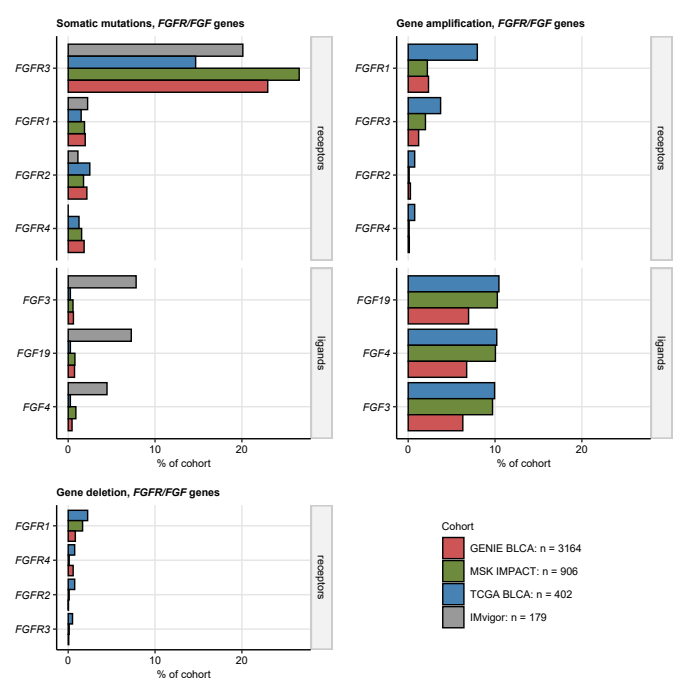
# Results

## Characteristic of investigated cohorts and frequency of genetic alterations



**Figure 1. The most frequent somatic mutations and gene copy alterations in urothelial cancer.**

*Percentages of somatic mutations, gene deletions and amplification present in at least 5% of cancers samples in all investigated cohorts (GENIE BLCA, MSK IMPACT, TCGA BLCA, and IMvigor, note: gene amplification and deletion data were not available for the IMvigor cohort) are presented in bar plots. Alterations affecting FGF- and FGFR-coding genes are highlighted with bold font. Numbers of analyzed cancer samples are indicated in the plot legend.*



**Figure 2. Frequency of somatic mutations and copy number alterations of FGF- and FGFR-coding genes.**

*Percentages of somatic mutations, deletions and amplifications of FGF- and FGFR-coding genes in the investigated cohort are presented in bar plots. Note that amplification and deletion data were not available for the IMvigor cohort. Numbers of analyzed cancer samples are indicated in the plot legend.*

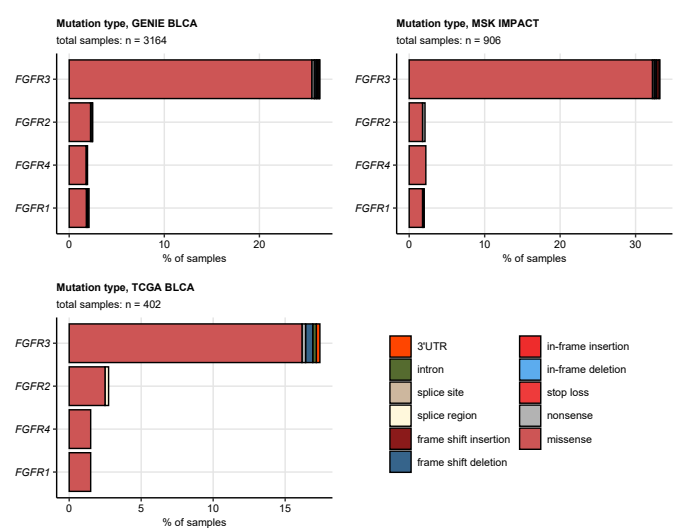
Table 1: Characteristic of the investigated cohorts. Numeric features are presented as medians with interquartile ranges and ranges. Qualitative variables are shown as percentages and observation counts of the categories.

| **Variable** | **GENIE BLCA** | **MSK IMPACT** | **TCGA BLCA** | **IMvigor** | **Significancea** | **Effect sizea** |
| --- | --- | --- | --- | --- | --- | --- |
| Age, years | 69 [IQR: 62 - 76] range: 18 - 89 complete: n = 3159 | 66 [IQR: 58 - 73] range: 28 - 92 complete: n = 906 | 69 [IQR: 60 - 76] range: 34 - 90 complete: n = 402 |  | p < 0.001 | η² = 0.013 |
| Gender | female: 27% (846) male: 73% (2313) complete: n = 3159 | female: 24% (215) male: 76% (690) complete: n = 905 | female: 26% (105) male: 74% (297) complete: n = 402 | female: 23% (41) male: 77% (138) complete: n = 179 | ns (p = 0.23) | V = 0.03 |
| Tissue | bladder: 82% (2610) non-bladder: 18% (554) complete: n = 3164 | bladder: 98% (888) non-bladder: 2% (18) complete: n = 906 | bladder: 100% (402) non-bladder: 0% (0) complete: n = 402 | bladder: 88% (158) non-bladder: 12% (21) complete: n = 179 | p < 0.001 | V = 0.22 |
| Invasiveness |  | non-muscle invasive: 6.2% (56) muscle invasive: 94% (850) complete: n = 906 | non-muscle invasive: 1.1% (4) muscle invasive: 99% (366) complete: n = 370 | non-muscle invasive: 0% (0) muscle invasive: 100% (179) complete: n = 179 |  |  |
| pT stage |  | T1: 6.2% (56) T2: 21% (190) T3: 72% (651) T4: 0.99% (9) T0: 0% (0) complete: n = 906 | T1: 0.81% (3) T2: 32% (117) T3: 52% (192) T4: 15% (57) T0: 0.27% (1) complete: n = 370 |  |  |  |
| pN stage |  |  | N0: 65% (233) N1: 12% (45) N2: 21% (74) N3: 2.2% (8) complete: n = 360 |  |  |  |
| pM stage |  |  | M0: 95% (191) M1: 5.4% (11) complete: n = 202 |  |  |  |
| Mutation burden, mutations/MB or fraction of genome | 0.16 [IQR: 0.044 - 0.33] range: 0 - 1 complete: n = 2247 | 9.5 [IQR: 5.5 - 16] range: 0 - 400 complete: n = 906 | 5.8 [IQR: 3.1 - 10] range: 0.033 - 120 complete: n = 402 | 8 [IQR: 5 - 14] range: 0 - 59 complete: n = 177 |  |  |
| Best overall response |  |  |  | SD/PD: 73% (115) CR/PR: 27% (42) complete: n = 157 |  |  |
| Mortality | 37% (1136) complete: n = 3109 | 30% (276) complete: n = 905 | 44% (177) complete: n = 402 | 64% (114) complete: n = 179 | p < 0.001 | V = 0.13 |
| aNumeric variables: Kruskal-Wallis test with η² effect size statistic. Categorical variables: χ² test with Cramer V effect size statistic. | | | | | | |

Table 2: Frequency of the most common somatic mutations and copy number variants in the investigated cohorts. Frequencies of alterations present in at least 5% of cancer samples are displayed. Statistical significance for differences between the cohorts was assessed by chi-squared test with Cramer's V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.

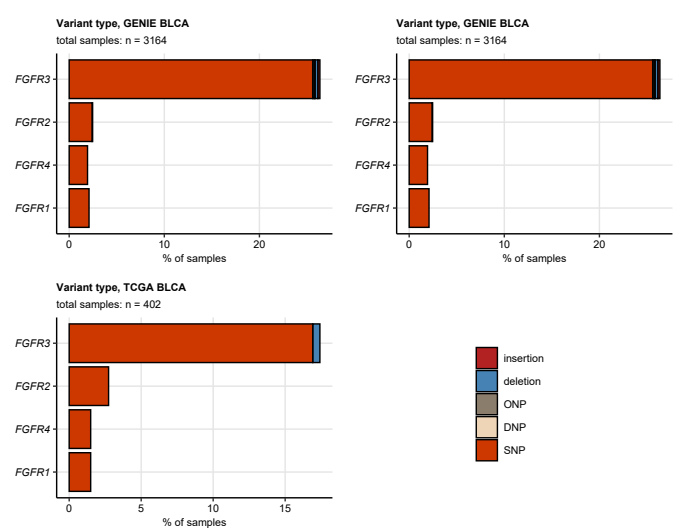
Table 3: Frequency of somatic mutations and copy number variants of FGF- and FGFR-coding genes in the investigated cohorts. Statistical significance for differences between the cohorts was assessed by chi-squared test with Cramer's V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.

## Classification of mutations in FGFR-coding genes, and their possible effects on protein structure and function



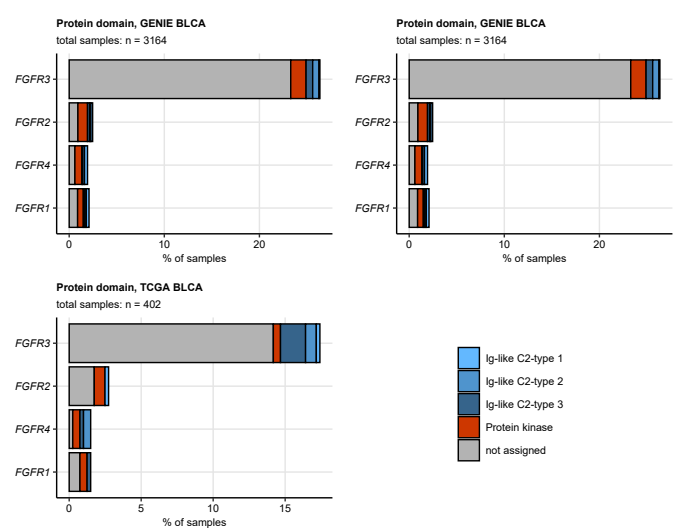
**Figure 3. Frequency of somatic mutations in FGFR1/2/3/4 genes split by mutation type.**

*Mutations in FGFR1, FGFR2, FGFR3, and FGFR4 genes in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts were classified by their effects on the coding sequence. Their frequency was expressed as percentage of cancer samples and presented in stack plots. Numbers of analyzed cancer samples are displayed in the plot captions.*



**Figure 4. Frequency of somatic mutations in FGFR1/2/3/4 genes split by variant type.**

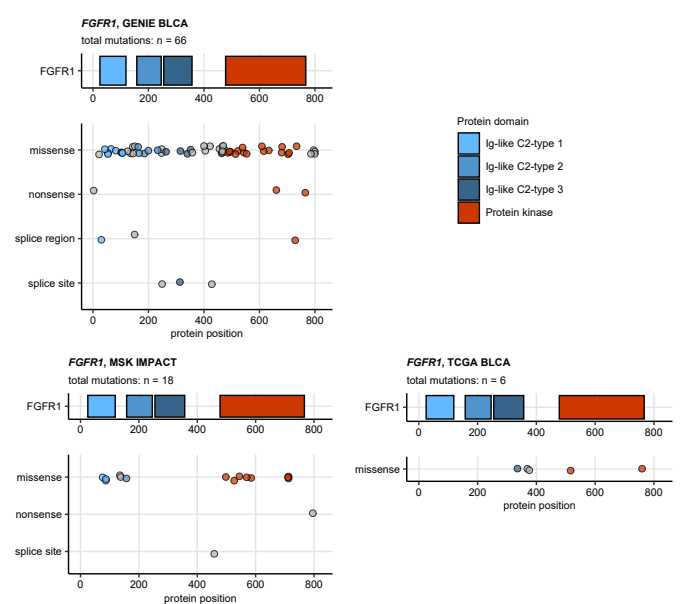
*Mutations in FGFR1, FGFR2, FGFR3, and FGFR4 genes in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts were classified by variant type. Their frequency was expressed as percentage of cancer samples and presented in stack plots. Numbers of analyzed cancer samples are displayed in the plot captions. SNP: single nucleotide polymorphism, DNP: di-nucleotide polymorphism; ONP: oligo-nucleotide polymorphism.*



**Figure 5. Frequency of somatic mutations in FGFR1/2/3/4 genes split by protein domain.**

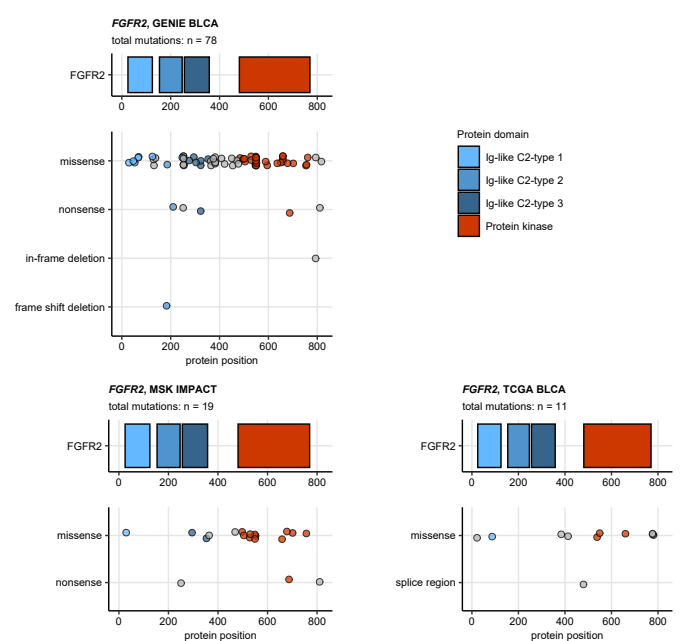
*Mutations in FGFR1, FGFR2, FGFR3, and FGFR4 genes in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts were classified by protein domain. Their frequency was expressed as percentage of cancer samples and presented in stack plots. Numbers of analyzed cancer samples are displayed in the plot captions.*

Table 4: Frequency of mutations in the FGFR1/2/3/4 genes in the GENIE BLCA and TCGA BLCA cohorts classified by mutation type, variant type, and protein domain. The table is available as a supplementary Excel file.



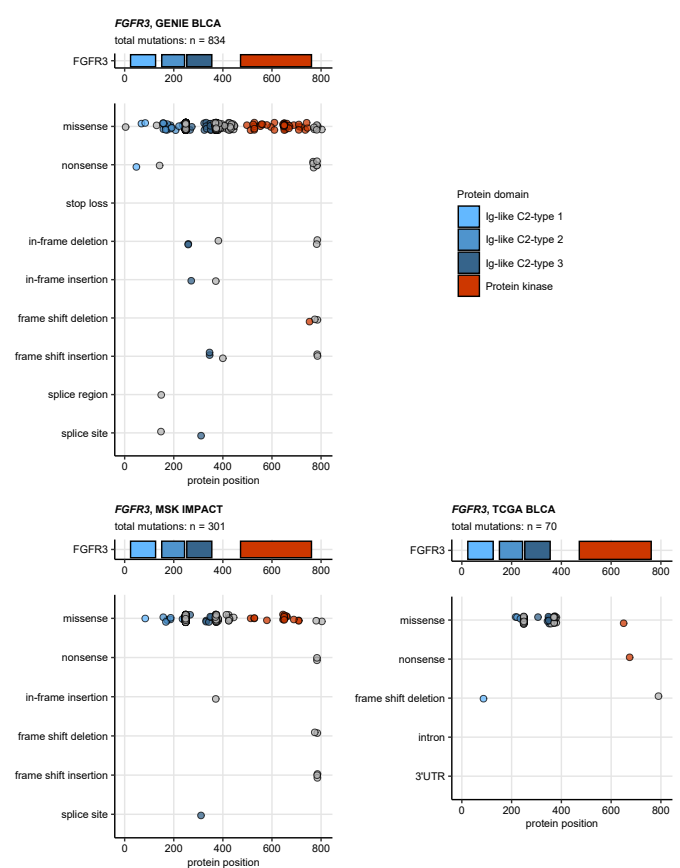
**Figure 6. Residues of FGFR1 protein affected by somatic mutations.**

*Somatic mutations in the FGFR1 gene in the GENIE BLCA, MSK IMPACT, and TCGA BLCA were classified by their effects on nucleotide sequence and mapped to amino acids of the FGFR1 protein. Mutation localization and type are visualized in scatter plots below the FGFR1 protein domain schemes. Each point represents a single somatic mutation. Point color codes for protein domain. Total mutation numbers are displayed in the plot captions.*



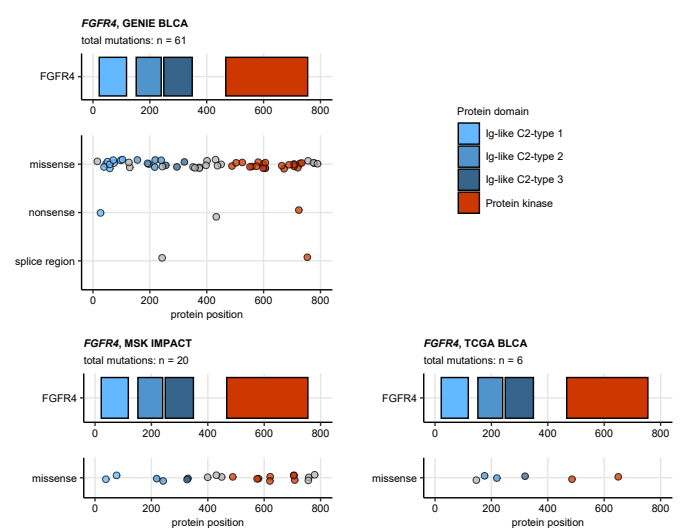
**Figure 7. Residues of FGFR2 protein affected by somatic mutations.**

*Somatic mutations in the FGFR2 gene in the GENIE BLCA, MSK IMPACT, and TCGA BLCA were classified by their effects on nucleotide sequence and mapped to amino acids of the FGFR2 protein. Mutation localization and type are visualized in scatter plots below the FGFR2 protein domain schemes. Each point represents a single somatic mutation. Point color codes for protein domain. Total mutation numbers are displayed in the plot captions.*



**Figure 8. Residues of FGFR3 protein affected by somatic mutations.**

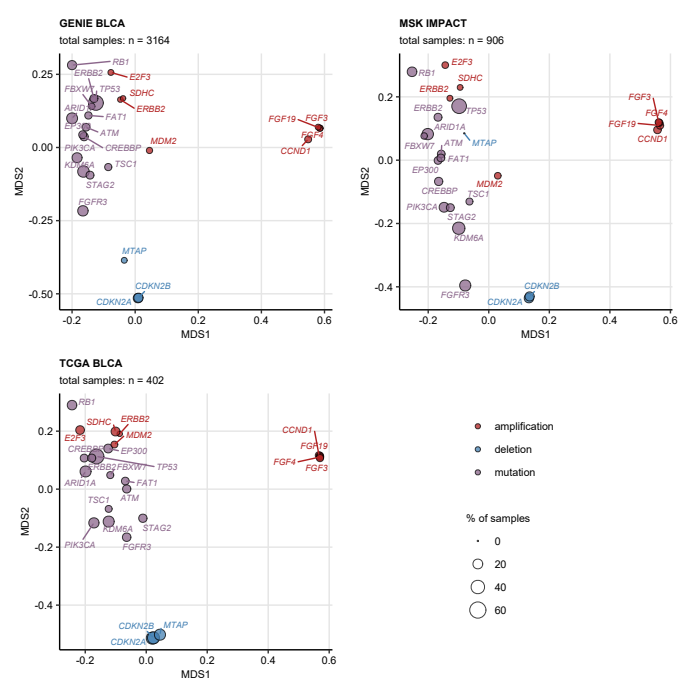
*Somatic mutations in the FGFR3 gene in the GENIE BLCA, MSK IMPACT, and TCGA BLCA were classified by their effects on nucleotide sequence and mapped to amino acids of the FGFR3 protein. Mutation localization and type are visualized in scatter plots below the FGFR3 protein domain schemes. Each point represents a single somatic mutation. Point color codes for protein domain. Total mutation numbers are displayed in the plot captions.*



**Figure 9. Residues of FGFR4 protein affected by somatic mutations.**

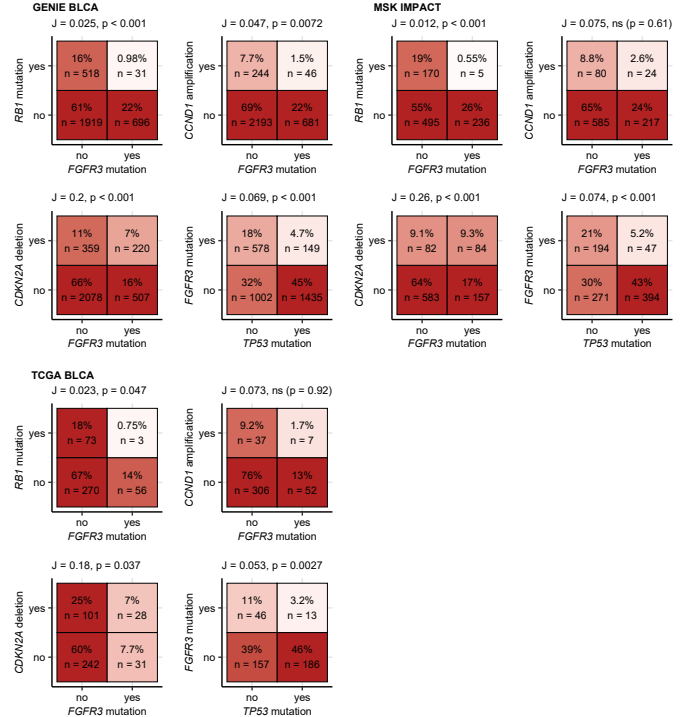
*Somatic mutations in the FGFR4 gene in the GENIE BLCA, MSK IMPACT, and TCGA BLCA were classified by their effects on nucleotide sequence and mapped to amino acids of the FGFR4 protein. Mutation localization and type are visualized in scatter plots below the FGFR4 protein domain schemes. Each point represents a single somatic mutation. Point color codes for protein domain. Total mutation numbers are displayed in the plot captions.*

## Co-occurrence of genetic alterations in FGF- and FGFR-coding genes with the most frequent genetic features of urothelial cancers



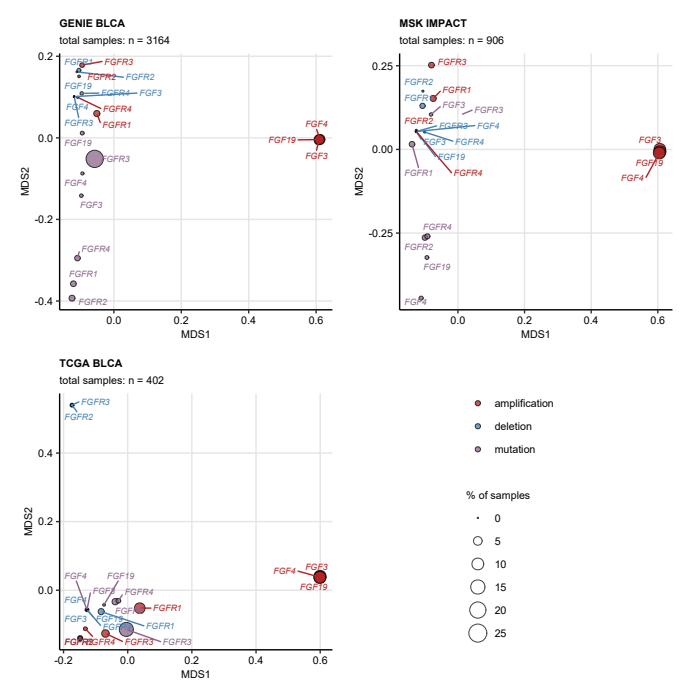
**Figure 10. Co-occurrence of the most common genetic alterations in urothelial cancer.**

*Co-occurrence of the most common genetic alterations in urothelial cancers (at least 5% of cancer samples) was investigated in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts by two-dimensional MDS (multi-dimensional scaling) of pairwise Jaccard’s distances.* *MDS results are visualized in scatter plots with single genetic alterations depicted as points. Point color codes for alteration type (somatic mutation, deletion, amplification), point size represents frequency of the alteration in the data set. Note that genetic features displayed in the plots close to each other are expected to co-occur in a high fraction of cancer samples. Total numbers of samples are indicated in the plot captions.*



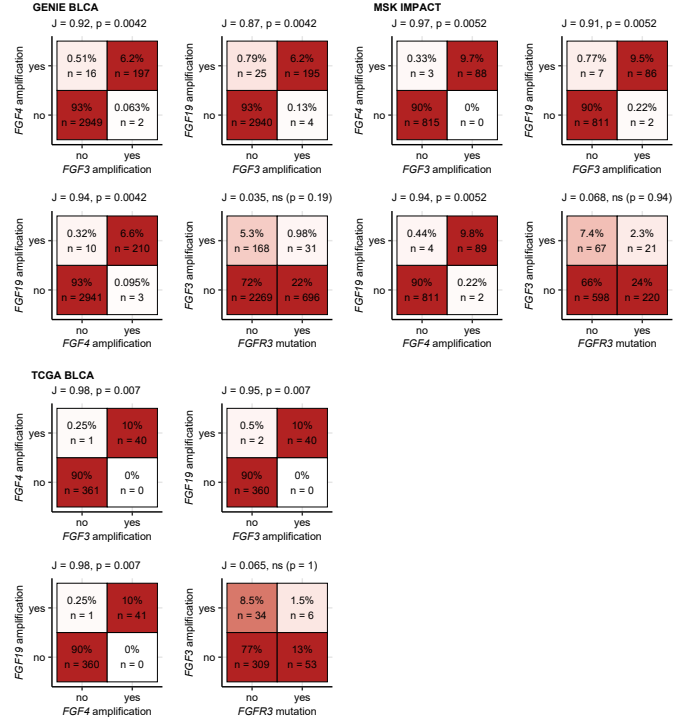
**Figure 11. Co-occurrence of selected genetic alterations in urothelial cancer.**

*Co-occurrence of the most common genetic alterations in urothelial cancers (at least 5% of cancer samples) was investigated in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts by pairwise Jaccard’s similarity coefficients J. Statistical significance of co-occurrence, i.e. , was assessed by false discovery rate-adjusted bootstrap tests.* *Frequencies of samples with and without genetic alterations are visualized as heat maps of contingency tables for selected pairs of genetic features. Sample percentages and counts within the total sample sets are indicated in the heat map tiles. J and p values are presented in the plot captions.*



**Figure 12. Co-occurrence of genetic alterations of FGF- and FGFR-coding genes in urothelial cancer.**

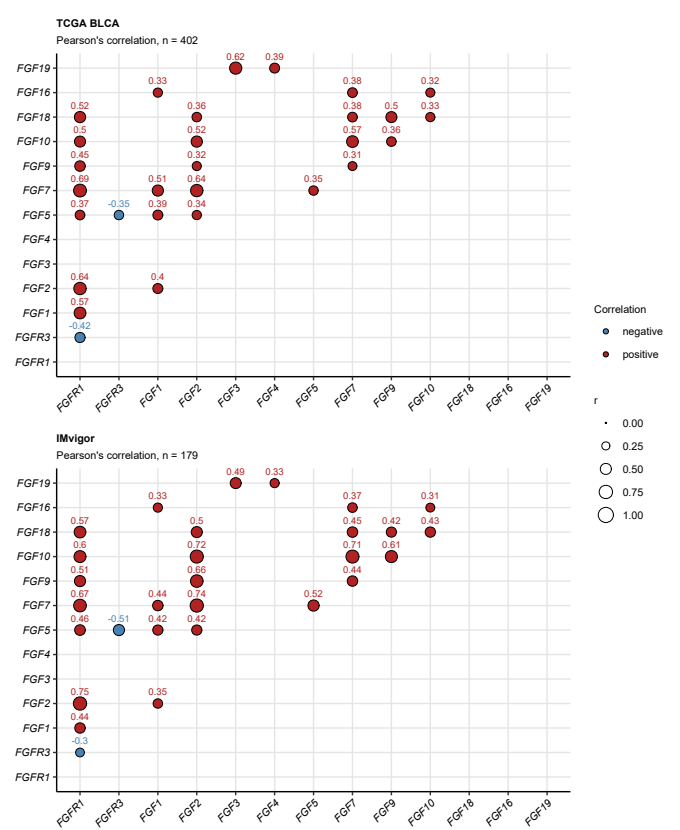
*Co-occurrence of somatic mutations, deletions, and amplifications of FGF- and FGFR-coding genes was investigated in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts by two-dimensional MDS (multi-dimensional scaling) of pairwise Jaccard’s distances.* *MDS results are visualized in scatter plots with single genetic alterations depicted as points. Point color codes for alteration type (somatic mutation, deletion, amplification), point size represents frequency of the alteration in the data set. Note that genetic features displayed in the plots close to each other are expected to co-occur in a high fraction of cancer samples. Total numbers of samples are indicated in the plot captions.*



**Figure 13. Co-occurrence of genetic alterations selected FGF- and FGFR-coding genes in urothelial cancer.**

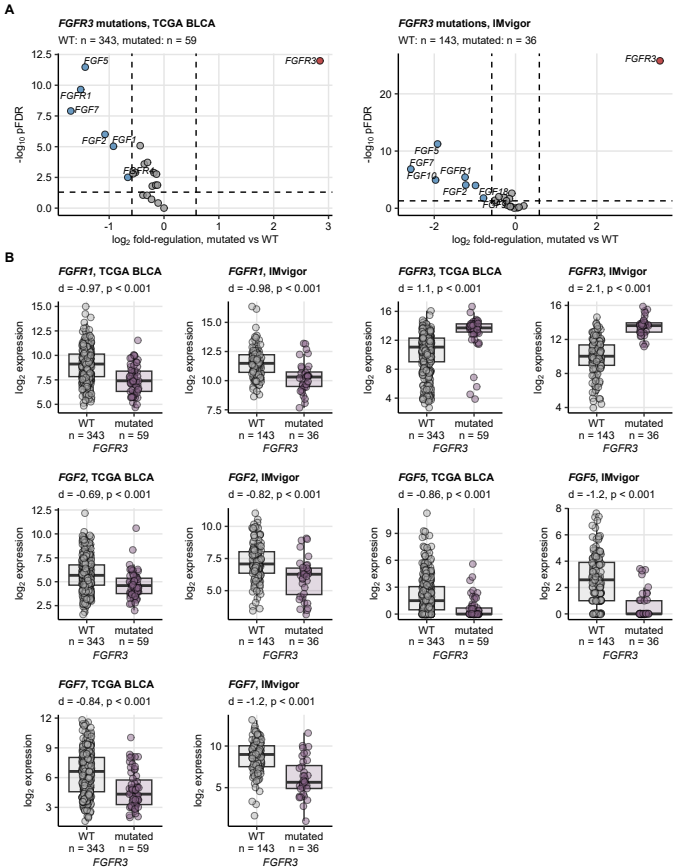
*Co-occurrence of somatic mutations, deletions, and amplifications in FGF- and FGFR-coding genes was investigated in the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts by pairwise Jaccard’s similarity coefficients J. Statistical significance of co-occurrence, i.e. , was assessed by false discovery rate-adjusted bootstrap tests.* *Frequencies of samples with and without genetic alterations are visualized as heat maps of contingency tables for selected pairs of genetic features. Sample percentages and counts within the total sample sets are indicated in the heat map tiles. J and p values are presented in the plot captions.*

## Clinical and biological significance of genetic alterations of FGF- and FGFR-coding genes



**Figure 14. Correlation of mRNA levels of FGF- and FGFR-coding genes.**

*Pairwise correlation of log2-transformed mRNA levels of FGF- and FGFR-coding genes was investigated in the TCGA BLCA and IMvigor cohorts by false discovery rate-adjusted Pearson’s test. Correlation coefficients r for significant gene pairs are depicted in bubble plots. Color of the points codes for correlation sign. Point size represents absolute values of r. Numbers of analyzed cancer samples are displayed in the plot captions.*



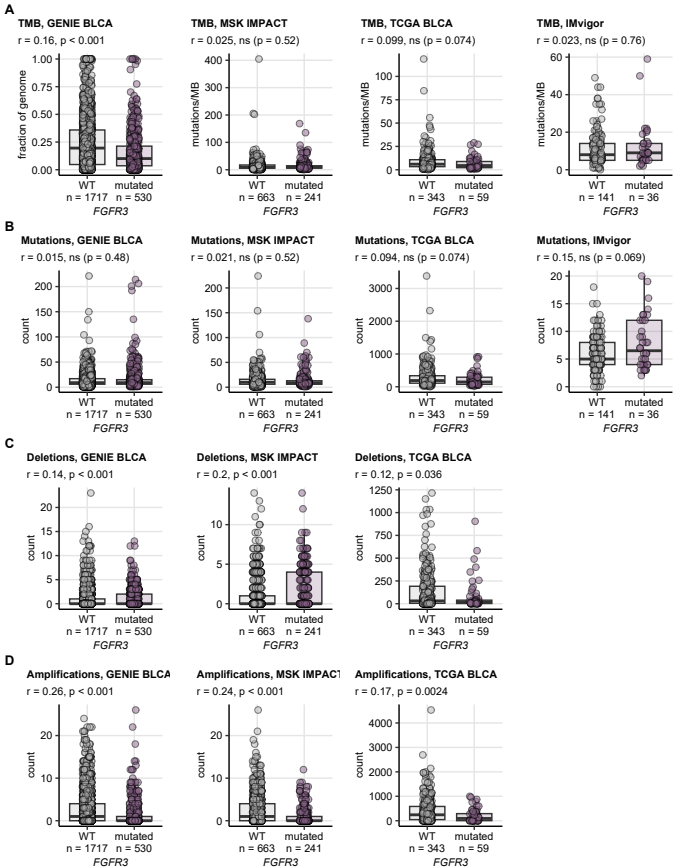
**Figure 15. Expression of FGF- and FGFR-coding genes in urothelial cancers stratified by presence of FGFR3 mutations.**

*Differences in expression of FGF- and FGFR-coding genes between cancers with and without somatic mutations in the FGFR3 gene were assessed by two-tailed T test with Cohen’s d effect size metric in the TCGA BLCA and IMvigor. P values were corrected with the false discovery rate (FDR) method. Differentially regulated genes were identified by pFDR < 0.05 and at least 1.5 fold-regulation of gene expression in mutated samples as compared with WT specimens.*

*(A) pFDR values and log2 fold-regulation estimates of gene expression in mutated vs WT samples are presented in Volcano plots. Each point represents a single gene. Point color codes for significance and regulation sign, differentially regulated genes are labeled with their symbols. The significance and fold-regulation cutoffs of differential gene expression are depicted as dashed lines. Numbers of analyzed cancer samples are indicated in the plot caption.*

*(B) log2-transformed mRNA levels for genes found to be differentially regulated in both cohorts are presented in box plots. Median expression values with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single observations are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

Table 5: Differential expression of FGF- and FGFR-coding genes in urothelial cancers with and without FGFR3 mutations. log2-transformed expression levels are presented as medians with interquartile ranges and ranges. Statistical significance was determined by two-tailed T test with Cohen's d effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.

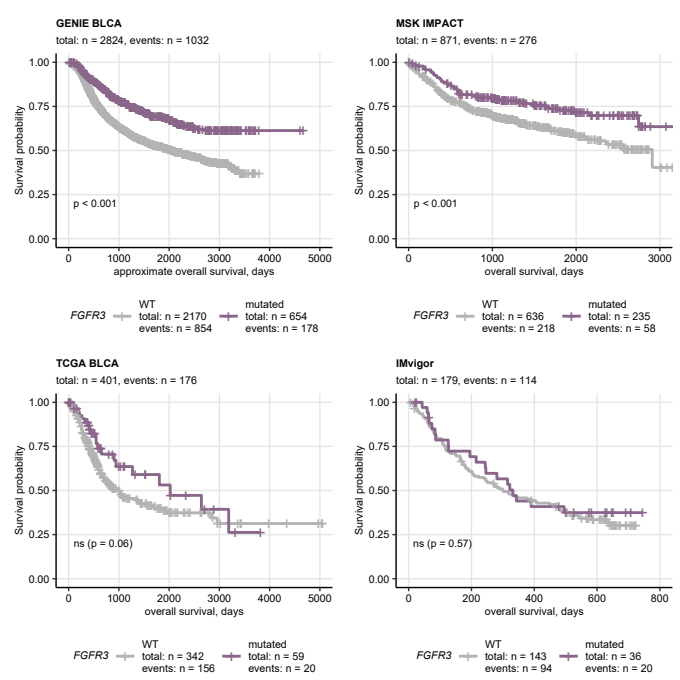


**Figure 16. Total mutation burden and counts of copy number alterations in FGFR3 WT and FGFR3-mutated cancers.**

*Total mutation burdens (TMB) were provided for the cohorts by the study authors. Counts of mutations, deletions and amplifications were computed for each cancer sample based on the published information. Note, that information on deletions and amplification was not available for the IMvigor cohort. TMB (A), and counts of mutations (B), deletions and amplifications (C, D) were compared between cancers with and without FGFR3 mutations by Mann-Whitney test with r effect size statistic. P values were adjusted for multiple testing with the false discovery rate method. Median values with interquartile ranges are shown as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

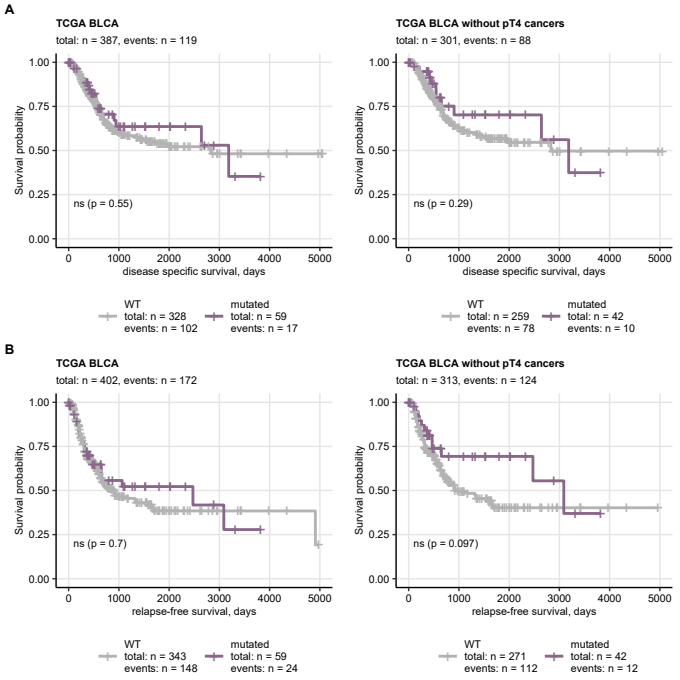
Table 6: Demographic, clinical, and pathological characteristic of cancers with WT and mutated FGFR3. Numeric features are presented as medians with interquartile ranges and ranges. Qualitative variables are shown as percentages and observation counts of the categories.

| **Cohort** | **Variable** | **FGFR3 wild-type** | **FGFR3 mutated** | **Significancea** | **Effect sizea** |
| --- | --- | --- | --- | --- | --- |
| GENIE BLCA | Age | 69 [IQR: 62 - 76] range: 18 - 89 complete: n = 2432 | 69 [IQR: 63 - 75] range: 31 - 89 complete: n = 727 | ns (p = 0.79) | η² = -0.00029 |
| Gender | female: 26% (629) male: 74% (1803) complete: n = 2432 | female: 30% (217) male: 70% (510) complete: n = 727 | ns (p = 0.056) | V = 0.038 |
| Tissue | bladder: 84% (2050) non-bladder: 16% (387) complete: n = 2437 | bladder: 77% (560) non-bladder: 23% (167) complete: n = 727 | p < 0.001 | V = 0.078 |
| MSK IMPACT | Age | 67 [IQR: 59 - 73] range: 28 - 92 complete: n = 665 | 65 [IQR: 56 - 72] range: 30 - 88 complete: n = 241 | p = 0.035 | η² = 0.0042 |
| Gender | female: 21% (141) male: 79% (523) complete: n = 664 | female: 31% (74) male: 69% (167) complete: n = 241 | p = 0.0068 | V = 0.098 |
| Tissue | bladder: 98% (653) non-bladder: 1.8% (12) complete: n = 665 | bladder: 98% (235) non-bladder: 2.5% (6) complete: n = 241 | ns (p = 0.7) | V = 0.022 |
| Invasiveness | non-muscle invasive: 1.7% (11) muscle invasive: 98% (654) complete: n = 665 | non-muscle invasive: 19% (45) muscle invasive: 81% (196) complete: n = 241 | p < 0.001 | V = 0.31 |
| pT stage | T1: 1.7% (11) T2: 15% (102) T3: 82% (545) T4: 1.1% (7) complete: n = 665 | T1: 19% (45) T2: 37% (88) T3: 44% (106) T4: 0.83% (2) complete: n = 241 | p < 0.001 | V = 0.42 |
| TCGA BLCA | Age | 69 [IQR: 61 - 77] range: 34 - 90 complete: n = 343 | 67 [IQR: 59 - 72] range: 42 - 85 complete: n = 59 | ns (p = 0.081) | η² = 0.008 |
| Gender | female: 28% (97) male: 72% (246) complete: n = 343 | female: 14% (8) male: 86% (51) complete: n = 59 | ns (p = 0.08) | V = 0.12 |
| Tissue | non-bladder: 0% (0) bladder: 100% (343) complete: n = 343 | non-bladder: 0% (0) bladder: 100% (59) complete: n = 59 |  |  |
| Invasiveness | non-muscle invasive: 0.94% (3) muscle invasive: 99% (317) complete: n = 320 | non-muscle invasive: 2% (1) muscle invasive: 98% (49) complete: n = 50 | ns (p = 1) | V = 0.035 |
| pT stage | T0: 0.31% (1) T1: 0.62% (2) T2: 29% (92) T3: 55% (176) T4: 15% (49) complete: n = 320 | T0: 0% (0) T1: 2% (1) T2: 50% (25) T3: 32% (16) T4: 16% (8) complete: n = 50 | ns (p = 0.08) | V = 0.18 |
| pN stage | N0: 63% (193) N1: 14% (42) N2: 22% (67) N3: 1.9% (6) complete: n = 308 | N0: 77% (40) N1: 5.8% (3) N2: 13% (7) N3: 3.8% (2) complete: n = 52 | ns (p = 0.18) | V = 0.13 |
| pM stage | M0: 96% (151) M1: 4.4% (7) complete: n = 158 | M0: 91% (40) M1: 9.1% (4) complete: n = 44 | ns (p = 0.49) | V = 0.085 |
| IMvigor | Gender | female: 21% (30) male: 79% (113) complete: n = 143 | female: 31% (11) male: 69% (25) complete: n = 36 | ns (p = 0.63) | V = 0.091 |
| Tissue | bladder: 89% (127) non-bladder: 11% (16) complete: n = 143 | bladder: 86% (31) non-bladder: 14% (5) complete: n = 36 | ns (p = 0.87) | V = 0.034 |
| Invasiveness | non-muscle invasive: 0% (0) muscle invasive: 100% (143) complete: n = 143 | non-muscle invasive: 0% (0) muscle invasive: 100% (36) complete: n = 36 |  |  |
| aNumeric variables: Kruskal-Wallis test with η² effect size statistic. Categorical variables: χ² test with Cramer V effect size statistic. | | | | | |



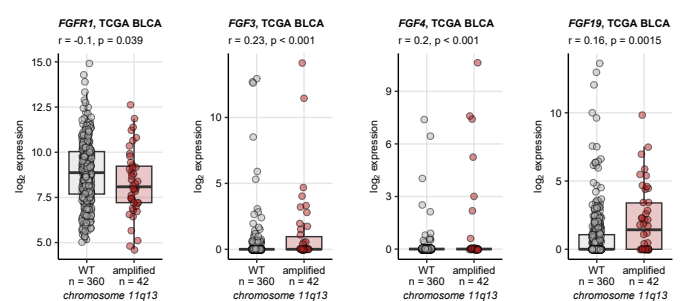
**Figure 17. Overall survival of patients with FGFR3 WT and FGFR3-mutated cancers.**

*Overall survival information was provided for the MSK IMPACT, TCGA BLCA, and IMvigor study participants. Approximate overall survival was calculated for the GENIE BLCA cohort based on age at enrollment in years and age at the last follow-up or death in days. Overall survival was subsequently compared between patients with cancers with and without FGFR3 mutations. Statistical significance was determined by false discovery rate-adjusted Peto-Peto test. Fractions of surviving patients are presented in Kaplan-Meier plots. P values are indicated in the plots, numbers of observations and deaths are displayed in the plot captions. Numbers of observations and deaths in the FGFR3 mutation strata are presented in the plot legends.*



**Figure 18. Disease-specific and relapse-free survival of patients with FGFR3 WT and FGFR3 mutated cancers.**

*Disease-specific and relapse-free survival information was provided for the TCGA BLCA cohort. Differences in survival between patients with cancers with and without FGFR3 mutations were investigated by false discovery rate-adjusted Peto-Peto test. Fractions of surviving patients are presented in Kaplan-Meier plots. P values are indicated in the plots, numbers of observations and deaths are displayed in the plot captions. Numbers of observations and deaths in the FGFR3 mutation strata are presented in the plot legends.*

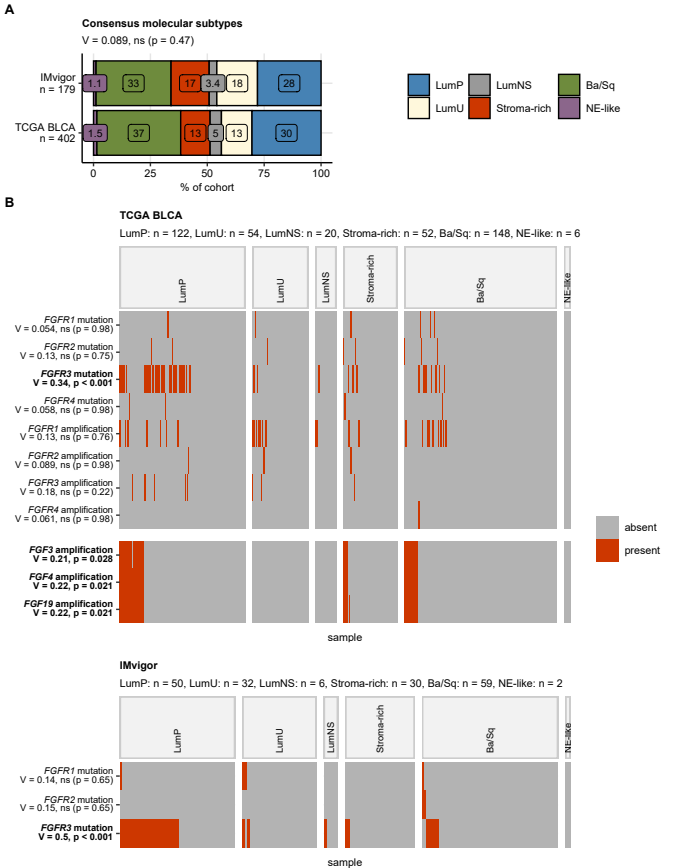


**Figure 19. Expression of FGF- and FGFR-coding genes in cancers with and without amplification of the 11q13 chromosome region.**

*Amplification of the 11q13 chromosome region with FGF3, FGF4, FGF19 and CCND genes was identified as frequent genetic alteration in urothelial carcinoma. Differences in expression of FGF- and FGFR-coding genes between cancers with and without 11q13 amplification were assessed in the TCGA BLCA cohort by Mann-Whitney test with r effect size metric. P values were corrected for multiple testing with the false discovery rate method. Median log2-transformed mRNA levels with interquartile ranges are represented by boxes with whiskers spanning over 150% of the interquartile ranges for the significantly regulated genes. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

## Consensus molecular subtypes of urothelial cancers

Table 7: Frequency of somatic mutations and copy number variants of FGF- and FGFR-coding genes in the consensus molecular classes of urothelial cancers. The frequencies are presented as percentages of the consensus class. Statistical significance of differences between the consensus classes was determined by chi-squared test with Cramer's V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.



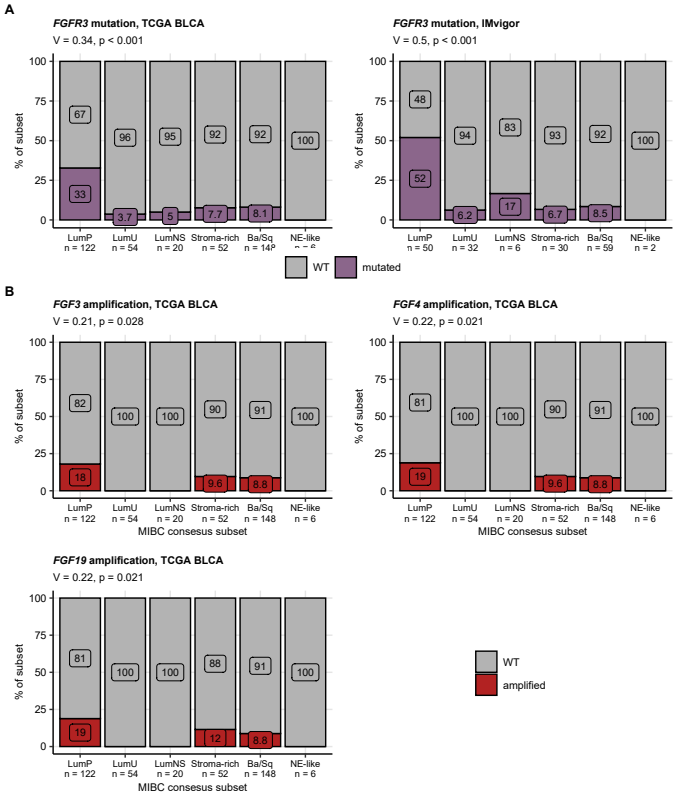
**Figure 20. Alterations of FGF- and FGFR-coding genes in consensus molecular classes of urothelial cancers.**

*Cancer samples from the GENIE BLCA and TCGA BLCA data sets were assigned to the MIBC consensus molecular subtypes based on ComBat-adjusted whole-genome mRNA levels by the nearest centroid algorithm (R package consensusMIBC).* *Differences in frequency of somatic mutations and copy number variants of FGF- and FGFR-coding genes between the consensus molecular classes were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Note that copy number variant information was not available for the IMvigor cohort.*

*(A) Percentages of cancer samples in the consensus molecular classes were visualized in a stack plot. Statistical significance of difference in class distribution between the cohorts was assessed by test with Cramer’s V effect size statistic. The effect size and p value are shown in the plot caption. Total numbers of cancer samples are indicated in the Y axis.*

*(B) Presence/absence of selected genetic alterations in the consensus molecular classes visualized in heat maps. Alteration names with effect sizes and p values of differences between the consensus classes are indicated in the Y axis. Significant effects are highlighted with bold font. Numbers of cancer samples in the consensus classes are displayed in the plot captions.*

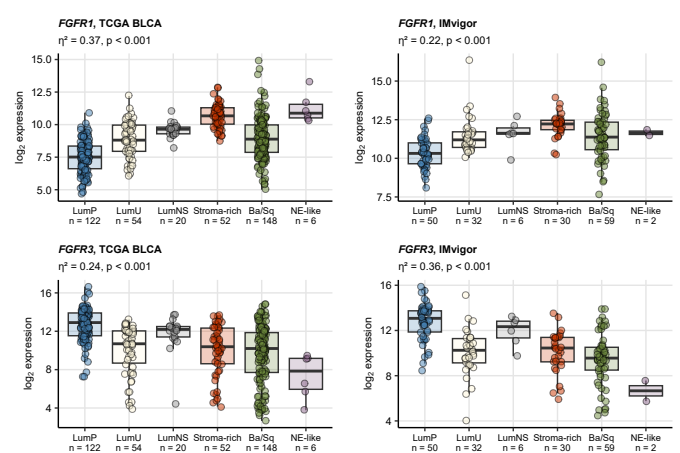
*LumP: luminal papillary; LumU: luminal genetically unstable; LumNS: luminal non-specified; Stroma-rich: luminal stroma-rich; Ba/Sq: basal/squamous-like; NE-like: neuroendocrine-like.*



**Figure 21. Frequency of somatic mutations of FGFFR3 and amplification of FGF3/4/19 in consensus molecular classes of urothelial cancer.**

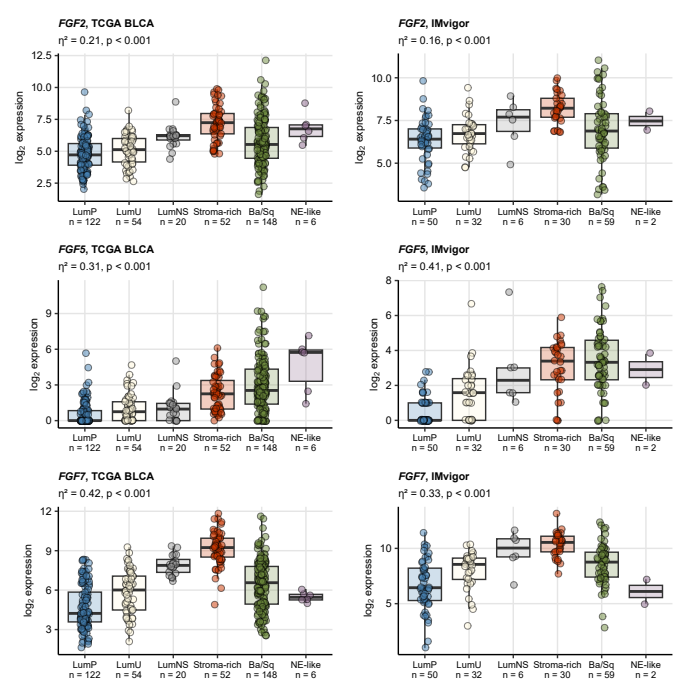
*Assignment of cancer samples to consensus molecular classes and testing for differences in genetic alterations between the consensus classes were presented in Figure 20. Percentages of FGFR3 mutations, and amplifications of the FGF3, FGF4 and FGF19 genes in the consensus molecular classes are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of observations in the consensus classes are indicated in the X axes.* *LumP: luminal papillary; LumU: luminal genetically unstable; LumNS: luminal non-specified; Stroma-rich: luminal stroma-rich; Ba/Sq: basal/squamous-like; NE-like: neuroendocrine-like.*

Table 8: Expression of FGF- and FGFR-coding genes in the consensus molecular classes of urothelial cancers. log2-transformed expression levels are presented as medians with interquartile ranges and ranges. Statistical significance of differences between the molecular classes was determined by one-way ANOVA with eta-square effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.



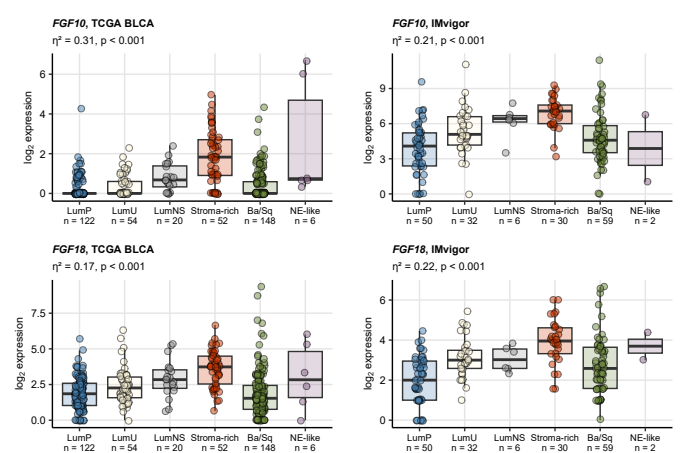
**Figure 22. Differential expression of FGFR-coding genes in the consensus molecular classes of urothelial cancers.**

*Assignment of cancer samples from the TCGA BLCA and IMvigor to the consensus molecular classes was presented in Figure 20. Differences in log2-transformed mRNA levels of FGF- and FGFR-coding genes between the consensus molecular classes were assessed by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the false discovery rate method.* *Expression of FGFR1 and FGFR3 was found to differ significantly between the consensus classes in both cohorts. Median log2-transformed mRNA levels with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the consensus classes are indicated in the X axes.* *LumP: luminal papillary; LumU: luminal genetically unstable; LumNS: luminal non-specified; Stroma-rich: luminal stroma-rich; Ba/Sq: basal/squamous-like; NE-like: neuroendocrine-like.*



**Figure 23. Differential expression of FGF-coding genes in the consensus molecular classes of urothelial cancers.**

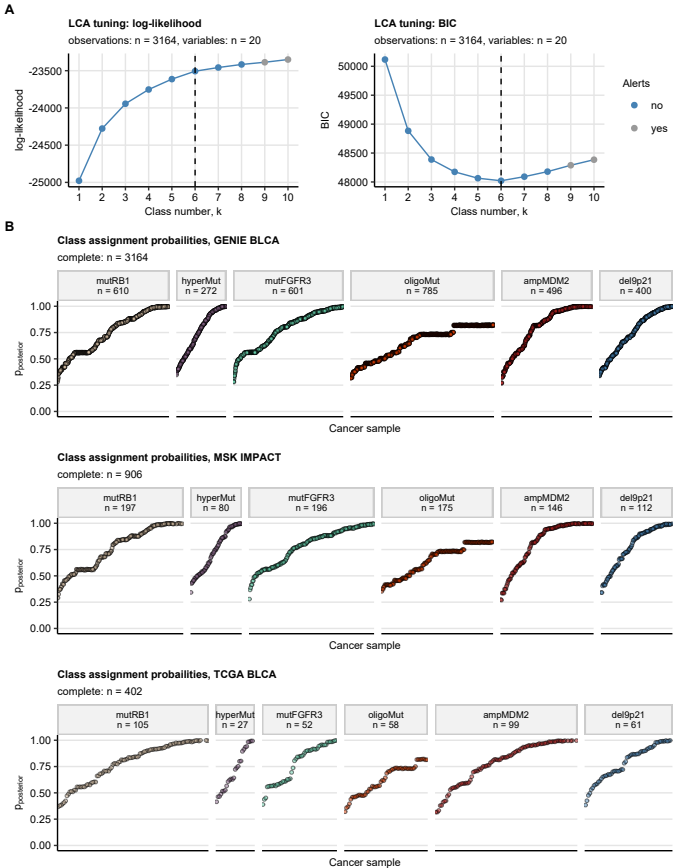
*Assignment of cancer samples from the TCGA BLCA and IMvigor to the consensus molecular classes was presented in Figure 20. Differences in log2-transformed mRNA levels of FGF- and FGFR-coding genes between the consensus molecular classes were assessed by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the false discovery rate method.* *Expression of FGF2, FGF5, and FGF7 was found to differ significantly between the consensus classes in both cohorts. Median log2-transformed mRNA levels with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the consensus classes are indicated in the X axes.* *LumP: luminal papillary; LumU: luminal genetically unstable; LumNS: luminal non-specified; Stroma-rich: luminal stroma-rich; Ba/Sq: basal/squamous-like; NE-like: neuroendocrine-like.*



**Figure 24. Differential expression of FGF-coding genes in the consensus molecular classes of urothelial cancers.**

*Assignment of cancer samples from the TCGA BLCA and IMvigor to the consensus molecular classes was presented in Figure 20. Differences in log2-transformed mRNA levels of FGF- and FGFR-coding genes between the consensus molecular classes were assessed by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the false discovery rate method.* *Expression of FGF10 and FGF18 was found to differ significantly between the consensus classes in both cohorts. Median log2-transformed mRNA levels with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the consensus classes are indicated in the X axes.* *LumP: luminal papillary; LumU: luminal genetically unstable; LumNS: luminal non-specified; Stroma-rich: luminal stroma-rich; Ba/Sq: basal/squamous-like; NE-like: neuroendocrine-like.*

## Genetic subsets of urothelial carcinoma

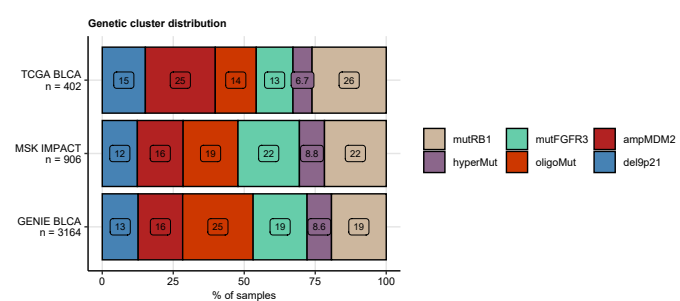


**Figure 25. Development of genetic subsets of urothelial carcinoma in the GENIE BLCA training cohort. Prediction of the genetic subset assignment for the MSK and TCGA BLCA cohort samples.**

*The genetic subsets of urothelial cancers were developed in the GENIE BLCA training cohort by latent class modeling with binary indexes (presence/absence) of genetic alterations present in at least 5% of samples as modeling factors. Presence of FGF3, FGF4, FGF19, or CCND1 amplification was subsumed under 11q13 amplification prior to the modeling. Presence of CDKN2A, CDKN2B, or MTAP deletions was subsumed under 9p21 deletion prior to the modeling.*

*(A) Selection of the genetic subset number in the GENIE BLCA training cohort based on log-likelihood and Bayesian Information Criterion (BIC). Note the minimal BIC for the 6-class model, which is subsequently used for definition of the genetic subsets.*

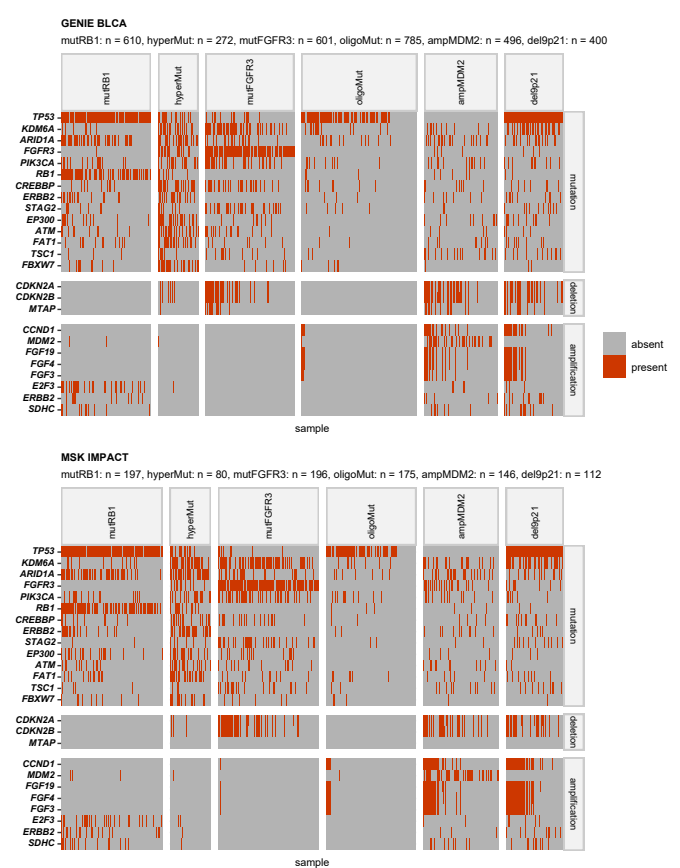
*(B) The 6-class model was trained in the GENIE BLCA cohort and used for assignment of cancer samples from the MSK IMPACT and TCGA BLCA test collectives to the genetic subsets. Class assignment probabilities (posterior p) of single cancer samples are visualized in scatter plots. Note that high posterior p indicate high confidence at class assignment and stability of the modeling solution. Total observation numbers are displayed in the plot captions. Numbers of samples in the genetic subsets are indicated in the plot facets.*



**Figure 26. Distribution of sizes of the genetic subsets in the training GENIE BLCA cohort, and the MSK and TCGA BLCA test collectives.**

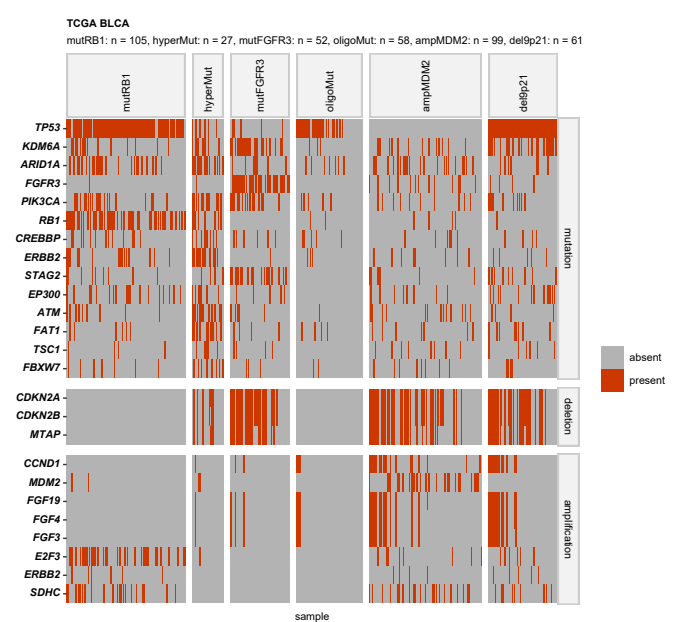
*Genetic subsets were developed in the GENIE BLCA cohort and predicted for cancers samples in the MSK IMPACT and TCGA BLCA collectives as presented in Figure 25. Sizes of the genetic subsets expressed as percentages of cancer samples in the cohort are presented in a stack plot.*

Table 9: Frequency of the most frequent somatic mutations and copy number alterations in the genetic subsets of urothelial cancers. The frequencies are presented as percentages of samples with genetic alterations in the genetic subsets. Differences between the genetic subsets were assessed by chi-squared test with Cramer's V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.



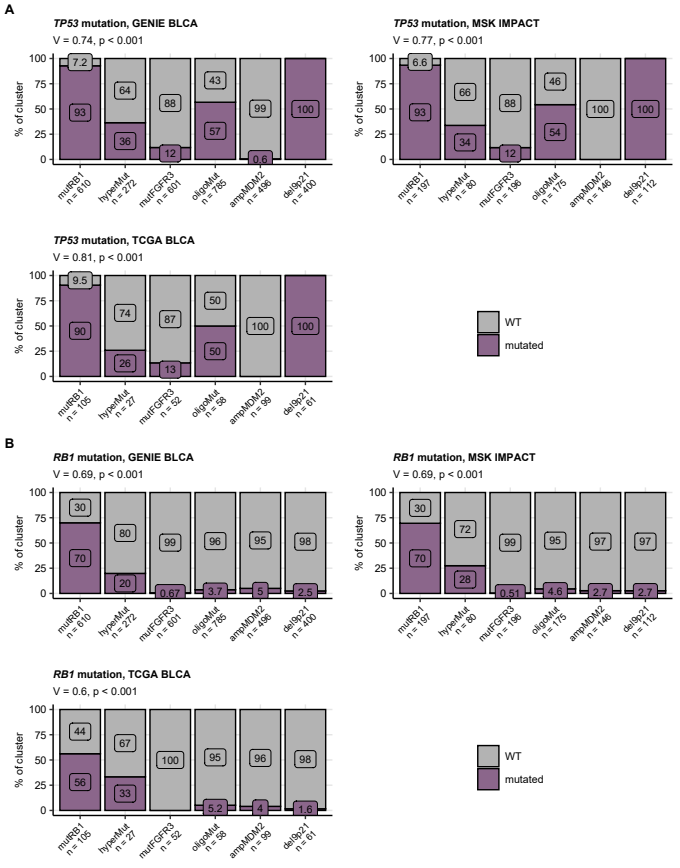
**Figure 27. Distribution of the most frequent somatic mutations and copy number alterations in the genetic subsets of urothelial cancers in the GENIE BLCA and MSK cohorts.**

*Differences in frequency of the most common somatic mutations and copy number variants ( 5% of samples) between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Presence of the genetic features in the genetic subsets was presented in a heat map. Significant alterations are highlighted in the Y axis with bold font. Numbers of analyzed cancer samples in the subsets are displayed in the plot captions.*



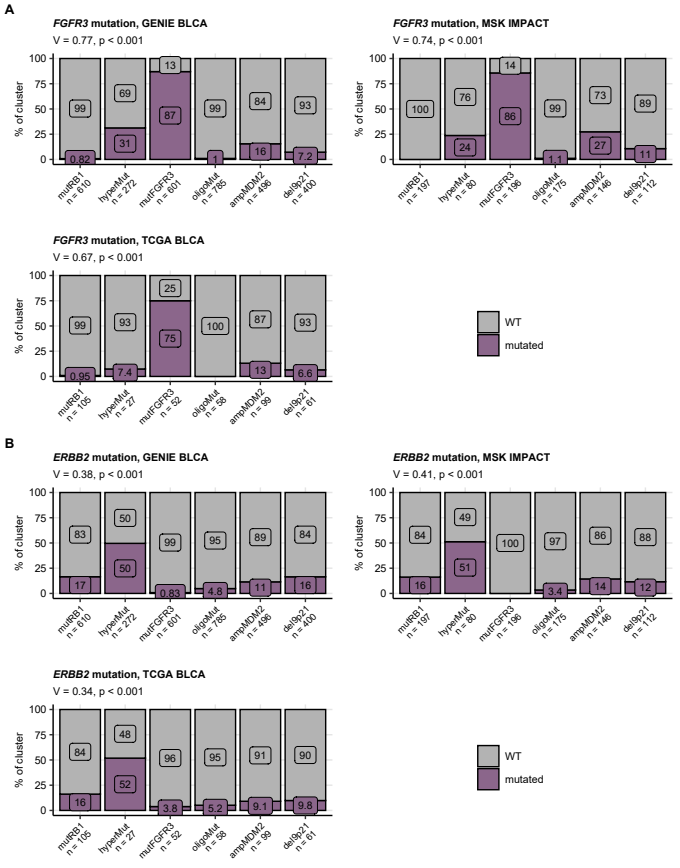
**Figure 28. Distribution of the most frequent somatic mutations and copy number alterations in the genetic subsets of urothelial cancers in the TCGA BLCA cohort.**

*Differences in frequency of the most common somatic mutations and copy number variants ( 5% of samples) between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Presence of the genetic features in the genetic subsets was presented in a heat map. Significant alterations are highlighted in the Y axis with bold font. Numbers of analyzed cancer samples in the subsets are displayed in the plot captions.*



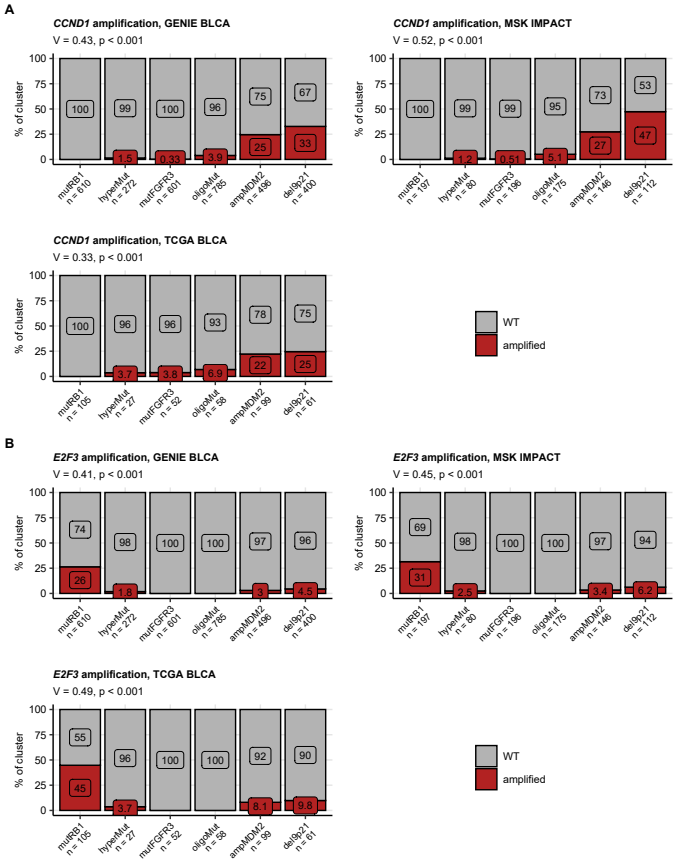
**Figure 29. Frequency of somatic mutations of TP53 and RB1 in the genetic subsets of urothelial cancers.**

*Differences in frequency of TP53 and RB1 mutations between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of the mutations in the genetic subsets are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the subsets are indicated in the X axes.*



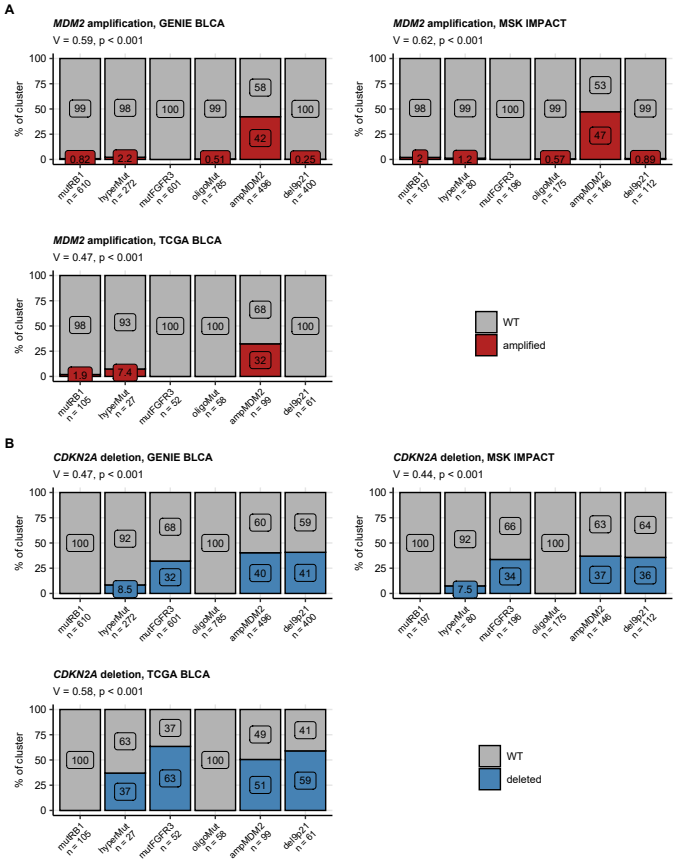
**Figure 30. Frequency of somatic mutations of FGFR3 and ERBB2 in the genetic subsets of urothelial cancers.**

*Differences in frequency of FGFR3 and ERBB2 mutations between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of the mutations in the genetic subsets are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the subsets are indicated in the X axes.*



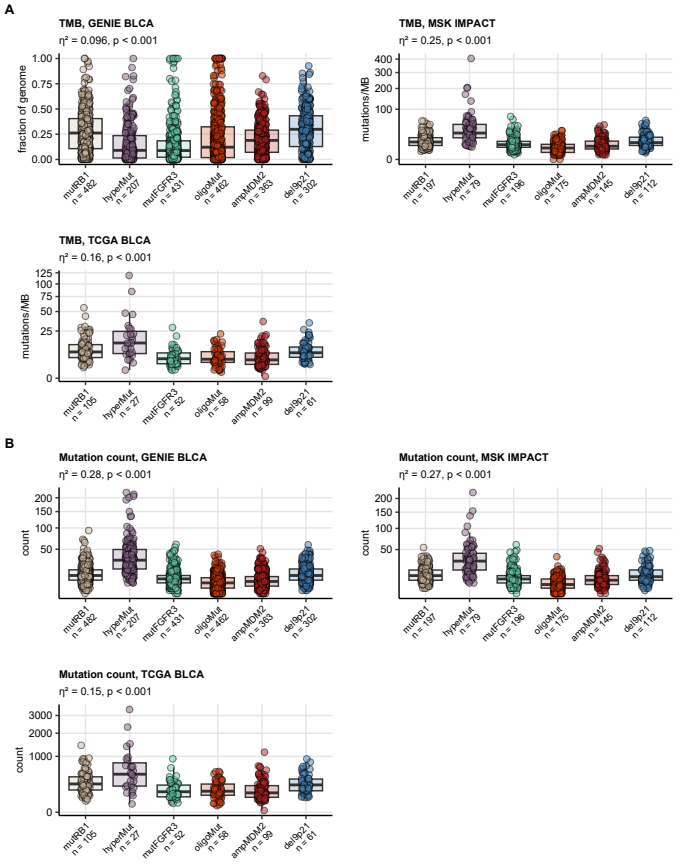
**Figure 31. Frequency of amplifications of CCND1 and E3F3 in the genetic subsets of urothelial cancers.**

*Differences in frequency of CCND1 (hallmark alteration for 11q13 chromosome region amplification) and E2F3 amplifications between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of the copy number variants in the genetic subsets are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the subsets are indicated in the X axes.*



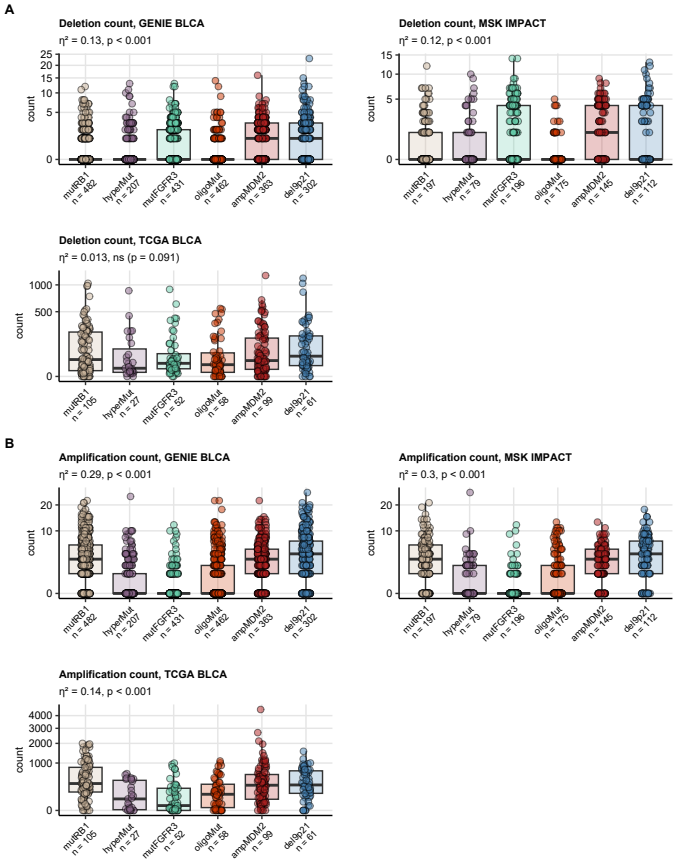
**Figure 32. Frequency of amplifications of MDM2 and deletions of CDKN2A in the genetic subsets of urothelial cancers.**

*Differences in frequency of MDM2 amplifications and CDKN2A deletions (characteristic for 9p21 chromosome region deletion) between the genetic subsets of urothelial cancers were assessed by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of the copy number variants in the genetic subsets are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the subsets are indicated in the X axes.*



**Figure 33. Total mutation burden and counts of mutations in protein-coding genes in the genetic subsets of urothelial cancers.**

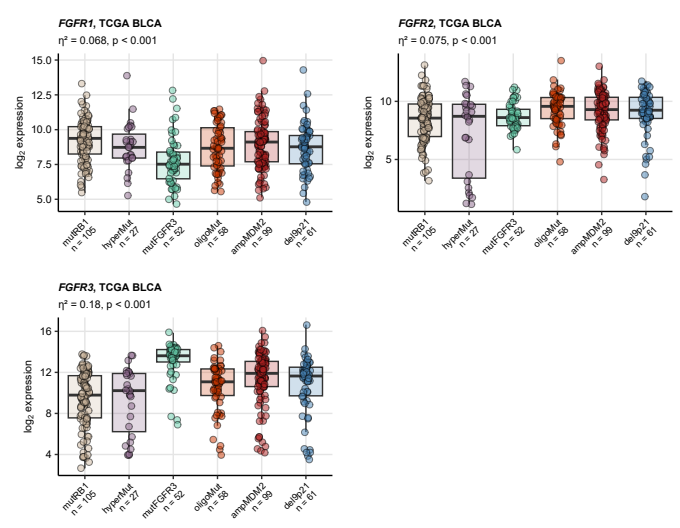
*Total mutation burdens (TMB) were provided for the GENIE BLCA, MSK IMPACT, and TCGA BLCA cohorts by the study authors. Counts of mutations were computed for each cancer sample based on the published information. TMB and counts of mutations were compared between the genetic subsets of urothelial cancers by Kruskal-Wallis test with effect size statistic. P values were adjusted for multiple testing with the false discovery rate method. Median values with interquartile ranges are shown as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*



**Figure 34. Counts of gene deletions and amplifications in the genetic subsets of urothelial cancers.**

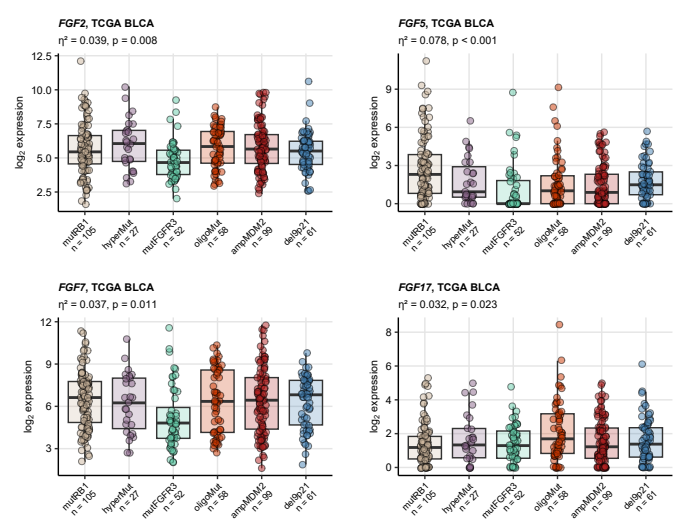
*Counts of deletions and amplifications were computed for cancer samples in the GENIE BLCA, MSK IMPACT, and TCGA BLCA based on the published information, and were compared between the genetic subsets of urothelial cancers by Kruskal-Wallis test with effect size statistic. P values were adjusted for multiple testing with the false discovery rate method. Median values with interquartile ranges are shown as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

Table 10: Expression of FGF- and FGFR-coding genes in the genetic subsets of urothelial cancers of the TCGA BLCA cohort. log2-transformed expression levels are presented as medians with interquartile ranges and ranges. Statistical significance of differences between the genetic subsets was determined by one-way ANOVA with eta-square effect size statistic. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.



**Figure 35. Differential expression of FGFR-coding genes in the genetic subsets of urothelial cancers.**

*Differences in log2-transformed mRNA levels of FGF- and FGFR-coding genes were compared between the genetic subsets of urothelial cancers in the TCGA BLCA cohort by one-way ANOVA with effect size statistic. Differences in expression of FGFR1, FGFR2, and FGFR3 were found significant. Median log2-transformed expression values with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

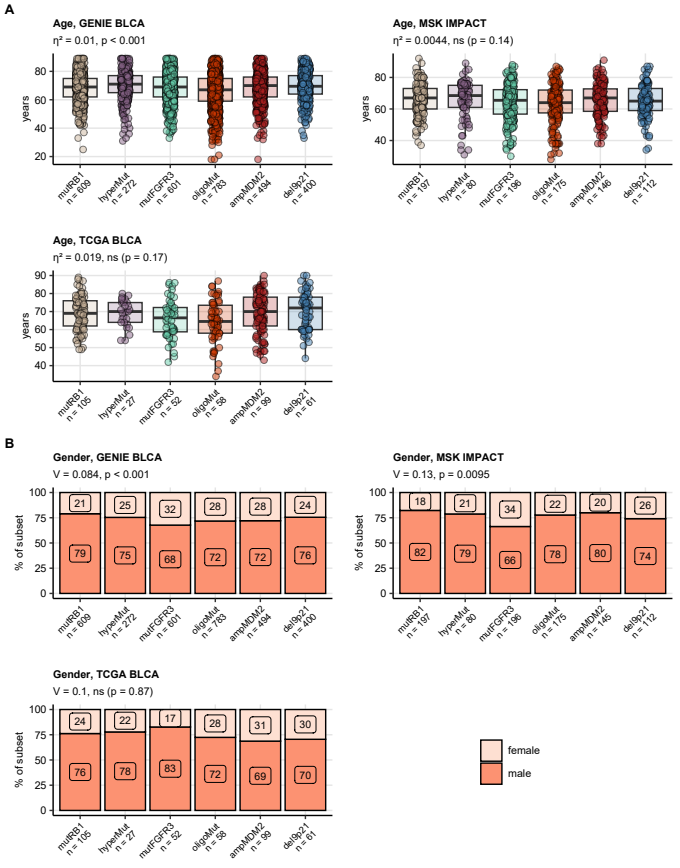


**Figure 36. Differential expression of FGF-coding genes in the genetic subsets of urothelial cancers.**

*Differences in log2-transformed mRNA levels of FGF- and FGFR-coding genes were compared between the genetic subsets of urothelial cancers in the TCGA BLCA cohort by one-way ANOVA with effect size statistic. Differences in expression of FGF2, FGF5, FGF7, and FGF17 were found significant. Median log2-transformed expression values with interquartile ranges are depicted as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of analyzed cancer samples are indicated in the X axes.*

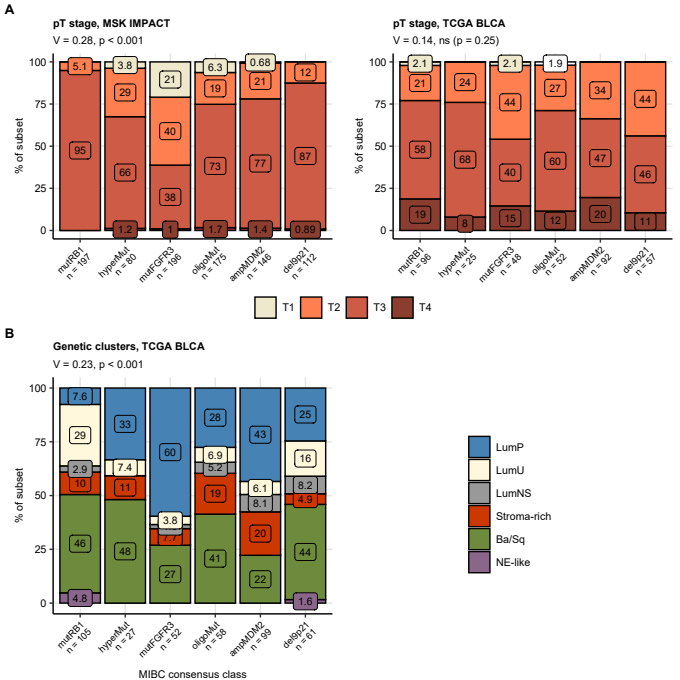
Table 11: Demographic, clinical, and pathological characteristic of the genetic subsets of urothelial cancers. Numeric features are presented as medians with interquartile ranges and ranges. Qualitative variables are shown as percentages and observation counts of the categories.

| **Cohort** | **Variable** | **mutRB1** | **hyperMut** | **mutFGFR3** | **oligoMut** | **ampMDM2** | **del9p21** | **Significancea** | **Effect sizea** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GENIE BLCA | Age | 69 [IQR: 62 - 75] range: 25 - 89 complete: n = 609 | 71 [IQR: 65 - 77] range: 31 - 89 complete: n = 272 | 69 [IQR: 62 - 76] range: 33 - 89 complete: n = 601 | 67 [IQR: 59 - 75] range: 18 - 89 complete: n = 783 | 70 [IQR: 62 - 76] range: 18 - 89 complete: n = 494 | 70 [IQR: 64 - 77] range: 33 - 89 complete: n = 400 | p < 0.001 | η² = 0.01 |
| Gender | female: 21% (128) male: 79% (481) complete: n = 609 | female: 25% (67) male: 75% (205) complete: n = 272 | female: 32% (194) male: 68% (407) complete: n = 601 | female: 28% (221) male: 72% (562) complete: n = 783 | female: 28% (138) male: 72% (356) complete: n = 494 | female: 24% (98) male: 76% (302) complete: n = 400 | p < 0.001 | V = 0.084 |
| Tissue | bladder: 93% (565) non-bladder: 7.4% (45) complete: n = 610 | bladder: 80% (218) non-bladder: 20% (54) complete: n = 272 | bladder: 78% (470) non-bladder: 22% (131) complete: n = 601 | bladder: 78% (614) non-bladder: 22% (171) complete: n = 785 | bladder: 81% (403) non-bladder: 19% (93) complete: n = 496 | bladder: 85% (340) non-bladder: 15% (60) complete: n = 400 | p < 0.001 | V = 0.14 |
| MSK IMPACT | Age | 67 [IQR: 60 - 73] range: 37 - 92 complete: n = 197 | 68 [IQR: 61 - 75] range: 31 - 89 complete: n = 80 | 66 [IQR: 57 - 72] range: 30 - 88 complete: n = 196 | 64 [IQR: 58 - 72] range: 28 - 87 complete: n = 175 | 67 [IQR: 58 - 73] range: 38 - 91 complete: n = 146 | 65 [IQR: 59 - 73] range: 34 - 87 complete: n = 112 | ns (p = 0.14) | η² = 0.0044 |
| Gender | female: 18% (35) male: 82% (162) complete: n = 197 | female: 21% (17) male: 79% (63) complete: n = 80 | female: 34% (66) male: 66% (130) complete: n = 196 | female: 22% (39) male: 78% (136) complete: n = 175 | female: 20% (29) male: 80% (116) complete: n = 145 | female: 26% (29) male: 74% (83) complete: n = 112 | p = 0.0095 | V = 0.13 |
| Tissue | bladder: 99% (196) non-bladder: 0.51% (1) complete: n = 197 | bladder: 99% (79) non-bladder: 1.2% (1) complete: n = 80 | bladder: 96% (189) non-bladder: 3.6% (7) complete: n = 196 | bladder: 98% (172) non-bladder: 1.7% (3) complete: n = 175 | bladder: 97% (141) non-bladder: 3.4% (5) complete: n = 146 | bladder: 99% (111) non-bladder: 0.89% (1) complete: n = 112 | ns (p = 0.2) | V = 0.09 |
| Invasiveness | non-muscle invasive: 0% (0) muscle invasive: 100% (197) complete: n = 197 | non-muscle invasive: 3.8% (3) muscle invasive: 96% (77) complete: n = 80 | non-muscle invasive: 21% (41) muscle invasive: 79% (155) complete: n = 196 | non-muscle invasive: 6.3% (11) muscle invasive: 94% (164) complete: n = 175 | non-muscle invasive: 0.68% (1) muscle invasive: 99% (145) complete: n = 146 | non-muscle invasive: 0% (0) muscle invasive: 100% (112) complete: n = 112 | p < 0.001 | V = 0.34 |
| pT stage | T1: 0% (0) T2: 5.1% (10) T3: 95% (187) T4: 0% (0) complete: n = 197 | T1: 3.8% (3) T2: 29% (23) T3: 66% (53) T4: 1.2% (1) complete: n = 80 | T1: 21% (41) T2: 40% (79) T3: 38% (74) T4: 1% (2) complete: n = 196 | T1: 6.3% (11) T2: 19% (33) T3: 73% (128) T4: 1.7% (3) complete: n = 175 | T1: 0.68% (1) T2: 21% (31) T3: 77% (112) T4: 1.4% (2) complete: n = 146 | T1: 0% (0) T2: 12% (14) T3: 87% (97) T4: 0.89% (1) complete: n = 112 | p < 0.001 | V = 0.28 |
| TCGA BLCA | Age | 69 [IQR: 62 - 76] range: 49 - 89 complete: n = 105 | 70 [IQR: 64 - 75] range: 54 - 80 complete: n = 27 | 66 [IQR: 59 - 72] range: 42 - 86 complete: n = 52 | 64 [IQR: 58 - 74] range: 34 - 87 complete: n = 58 | 70 [IQR: 62 - 78] range: 43 - 90 complete: n = 99 | 72 [IQR: 60 - 78] range: 44 - 90 complete: n = 61 | ns (p = 0.17) | η² = 0.019 |
| Gender | female: 24% (25) male: 76% (80) complete: n = 105 | female: 22% (6) male: 78% (21) complete: n = 27 | female: 17% (9) male: 83% (43) complete: n = 52 | female: 28% (16) male: 72% (42) complete: n = 58 | female: 31% (31) male: 69% (68) complete: n = 99 | female: 30% (18) male: 70% (43) complete: n = 61 | ns (p = 0.87) | V = 0.1 |
| Invasiveness | non-muscle invasive: 2.1% (2) muscle invasive: 98% (94) complete: n = 96 | non-muscle invasive: 0% (0) muscle invasive: 100% (25) complete: n = 25 | non-muscle invasive: 2.1% (1) muscle invasive: 98% (47) complete: n = 48 | non-muscle invasive: 1.9% (1) muscle invasive: 98% (51) complete: n = 52 | non-muscle invasive: 0% (0) muscle invasive: 100% (92) complete: n = 92 | non-muscle invasive: 0% (0) muscle invasive: 100% (57) complete: n = 57 | ns (p = 0.87) | V = 0.099 |
| pT stage | T0: 0% (0) T1: 2.1% (2) T2: 21% (20) T3: 58% (56) T4: 19% (18) complete: n = 96 | T0: 0% (0) T1: 0% (0) T2: 24% (6) T3: 68% (17) T4: 8% (2) complete: n = 25 | T0: 0% (0) T1: 2.1% (1) T2: 44% (21) T3: 40% (19) T4: 15% (7) complete: n = 48 | T0: 1.9% (1) T1: 0% (0) T2: 27% (14) T3: 60% (31) T4: 12% (6) complete: n = 52 | T0: 0% (0) T1: 0% (0) T2: 34% (31) T3: 47% (43) T4: 20% (18) complete: n = 92 | T0: 0% (0) T1: 0% (0) T2: 44% (25) T3: 46% (26) T4: 11% (6) complete: n = 57 | ns (p = 0.25) | V = 0.14 |
| pN stage | N0: 58% (56) N1: 12% (12) N2: 27% (26) N3: 2.1% (2) complete: n = 96 | N0: 80% (20) N1: 8% (2) N2: 12% (3) N3: 0% (0) complete: n = 25 | N0: 72% (33) N1: 8.7% (4) N2: 17% (8) N3: 2.2% (1) complete: n = 46 | N0: 63% (33) N1: 12% (6) N2: 21% (11) N3: 3.8% (2) complete: n = 52 | N0: 68% (63) N1: 15% (14) N2: 15% (14) N3: 1.1% (1) complete: n = 92 | N0: 57% (28) N1: 14% (7) N2: 24% (12) N3: 4.1% (2) complete: n = 49 | ns (p = 0.87) | V = 0.1 |
| pM stage | M0: 93% (42) M1: 6.7% (3) complete: n = 45 | M0: 93% (13) M1: 7.1% (1) complete: n = 14 | M0: 94% (33) M1: 5.7% (2) complete: n = 35 | M0: 96% (24) M1: 4% (1) complete: n = 25 | M0: 94% (49) M1: 5.8% (3) complete: n = 52 | M0: 97% (30) M1: 3.2% (1) complete: n = 31 | ns (p = 0.99) | V = 0.056 |
| aNumeric variables: Kruskal-Wallis test with η² effect size statistic. Categorical variables: χ² test with Cramer V effect size statistic. | | | | | | | | | |



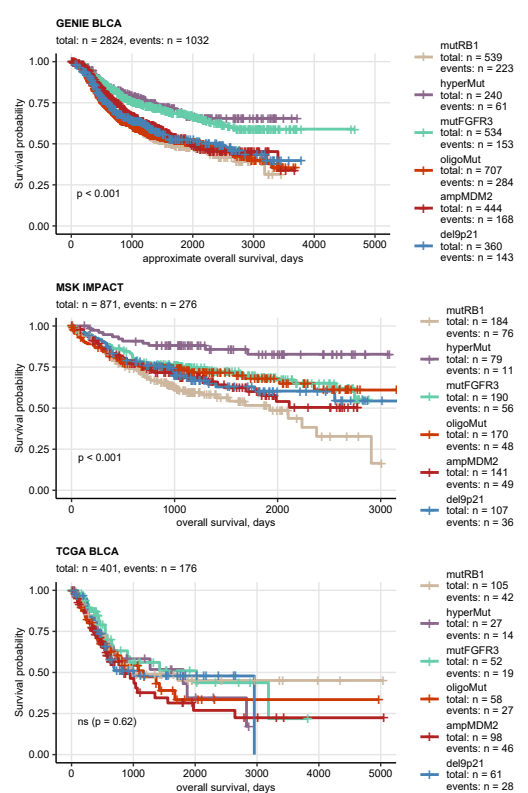
**Figure 37. Age and gender in the genetic subsets.**

*Differences in age (A) and gender (B) between the genetic subsets of urothelial cancers in the GENIE BLCA and TCGA BLCA were investigated by Kruskal-Wallis test and test, respectively. and Cramer’s V served as effect size metrics. P values were corrected for multiple testing with the false discovery rate method. Median age with the interquartile range is presented in box plots with whiskers spanning over 150% of the interquartile range and single cancer patients visualized as points. Percentages of women and men in the genetic subsets are shown in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of observations in the genetic subsets are indicated in the X axes.*



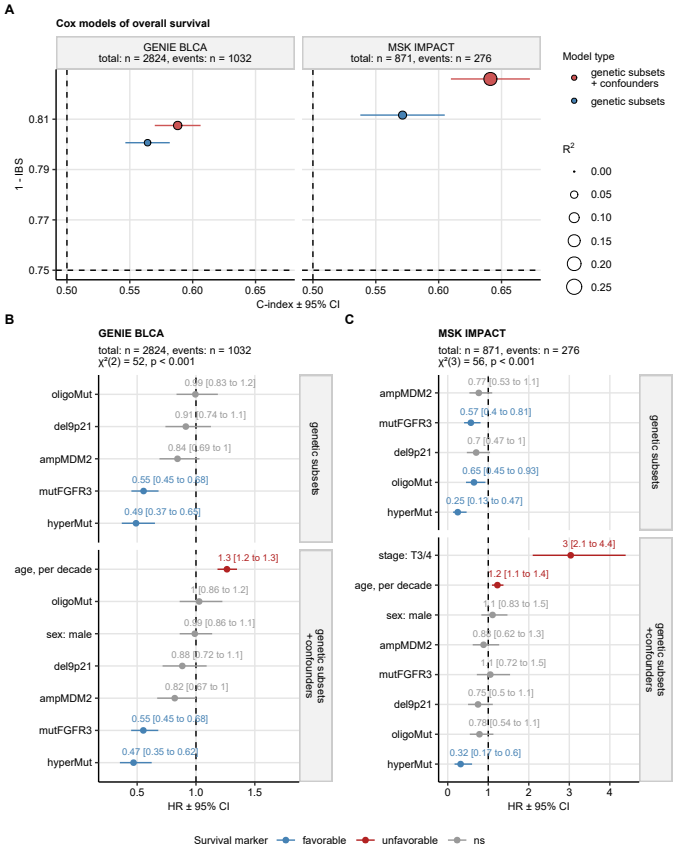
**Figure 38. Pathological tumor stage and consensus molecular class distribution in the genetic subsets of urothelial cancers.**

*Differences in distribution of pathological tumor stages (A) and consensus molecular classes (B) between the genetic subsets were investigated by test with Cramer’s V effect size statistic. P values were adjusted for multiple testing with the false discovery rate method. Percentages of pathological stages and consensus classes within the genetic subsets are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of complete observations are indicated in the X axes.*



**Figure 39. Overall survival in the genetic subsets of urothelial cancers.**

*Overall survival information was provided for the GENIE BLCA, MSK IMPACT, and TCGA BLCA study participants. Approximate overall survival was calculated for the GENIE BLCA cohort based on age at enrollment in years and age at the last follow-up or death in days. Overall survival was subsequently compared between the genetic subsets by false discovery rate-corrected Peto-Peto test. Fractions of surviving patients are presented in Kaplan-Meier plots. P values for globals differences in survival between the subsets are indicated in the plots, numbers of observations and deaths are displayed in the plot captions. Numbers of observations and deaths in the FGFR3 mutation strata are presented in the plot legends.*



**Figure 40. Cox modeling of overall survival in the genetic subsets of urothelial carcinoma.**

*Overall survival (MSK IMPACT) and approximate overall survival (GENIE BLCA) was modeled as a function of genetic subset assignment alone or as a function of genetic subset assignment and clinical confounders (GENIE BLCA: age and gender, MSK IMPACT: age, gender, and pathological tumor stage) with Cox proportional hazard regression. Significance of the genetic subset term was investigated by likelihood ratio test (LRT) between the model with genetic subset assignment only and the model with genetic subset assignment and clinical confounders as explanatory factors.*

*(A) Numeric statistics of performance of Cox models with genetic subset assignment and Cox models with genetic subset assignment and clinical confounders as explanatory factors. Concordance between the observed and predicted survival was assessed by Harrell’s concordance index (high values: good concordance), model calibration was investigated by integrated Brier score (IBS; low values: good calibration), and fraction of survival variance explained by the model was gauged by R^2 statistic. The statistic values are visualized in a bubble plot with point color coding for model type and point size representing R^2. Numbers of complete observations and deaths are indicated in the plot facets.*

*(B, C) Cox model coefficient estimates expressed as hazard ratios (HR). HR with 95% confidence intervals are visualized with Forest plots faceted by model type. Numbers of complete observations and deaths, and results of LRT are presented in the plot captions.*

# Methods

## Software

Data handling and analysis was accomplished with R version 4.2.3. For general transformation tasks for matrix, tabular, and text data, base R, *tidyverse* package bundle (7), *rlang* (8), [*trafo*](https://github.com/PiotrTymoszuk/trafo), and *stringi* (9) were utilized. General data visualization tasks were performed with *ggplot2* (10), *cowplot* (11), and *flextable* (12). R packages specific for particular analysis and visualization steps are referenced below.

## Data sources

Four collectives of urothelial carcinoma patients were evaluated in the current report: the urothelial cancer subset of the GENIE cohort referred to as GENIE BLCA (1), the MSK IMPACT collective (2), the bladder cancer TCGA collective referred to as TCGA BLCA (3,4), and the IMvigor clinical study cohort (5).

The GENIE BLCA cohort (1) includes n = 3164 cases of bladder and urinary tract cancers selected by oncotree classification ‘BLAD’, ‘BLCA’, ‘BLSC’, ‘UA’, ‘UAD’, ‘UCA’, ‘UCU’, ‘UPA’, ‘URCA’, ‘USCC’, or ‘UTUC’. The available data sets for the GENIE BLCA cohort included basic clinical information (age at sequencing in years, sex, follow-up in days, mortality), total mutation burden (TMB) expressed as fraction of genome, detailed information on location, variant type, nucleotide sequence and protein residues affected by somatic mutations, as well as 0/1 coded absence/presence of gene deletions and amplifications. Approximate overall survival was computed as the time interval between the last follow-up in days and age at sequencing in years. In case of multiple samples per participant, the earliest sample, i.e. the specimen with the lowest ID, was included in the current analysis. The GENIE BLCA data sets were fetched from the [cBioportal API](https://www.cbioportal.org/) with tools provided by the *cbioportalR* package (13) and formatted with in-house developed R scripts.

The MSK IMPACT collective consisted of n = 906 patients with predominantly muscle-invasive urothelial cancers (2).  
In case of multiple samples available for a patient, the sample with the lowest ID was included in the analysis. The clinical information included entries on participants age, gender, overall survival, cancer pathological stage, and cancer treatment. The mutation data set included detailed information on mutation nucleotide sequence, variant type, and affected protein residues. The copy number variant data were provided in 0/1 coded presence of gene deletions and amplifications. The data sets for the MSK impact study were downloaded from [cBioportal](https://www.cbioportal.org/) and implemented in R with in-house developed scripts.

Data sets for the TCGA BLCA cohort (3,4) was fetched from the [cBioportal](https://www.cbioportal.org/) repository with in-house developed scripts for n = 402 predominantly muscle-invasive urothelial cancer patients. The clinical information consisted of demographic variables (e.g. age, sex, BMI, smoking history), pathological features (e.g. pathological staging, histology, grade, extracapsular extension), survival information (overall, tumor-specific and relapse-free survival), and TMB expressed as mutation count per MB. The somatic mutation data set included detailed information on location, variant type, nucleotide sequence and affected protein residues. The copy number variant data set consisted of 0/1 coded presence of gene amplifications and deletions. The expression data set for the TCGA BLCA collective consisted of -transformed RSEM-normalized gene-integrated transcript counts obtained by bulk cancer RNA sequencing.

The IMvigor data set included a subset of 179 bladder and ureter muscle non-invasive and muscle-invasive cancers of the IMvigor 210 trial subjected to RNAseq and DNA sequencing (5,14). The study participants were treated with an anti-PD-L1 drug (atezolizumab) following platinum chemotherapy. The clinical data set included basic demographic, clinical and pathological features, best overall response (complete/partial response [CR/PR], stable/progressive disease [SD/PD]), overall survival, and TMB quantified as mutation count per MB. The mutation data set consisted of 0/1 coded basic information on presence of any somatic mutations in a particular gene. The expression data set consisted of -transformed gene-integrated transcript counts obtained by bulk cancer RNA sequencing. The IMvigor data sets were obtained from the R package *IMvigor210CoreBiologies*.

Assignment of the TCGA BLCA and IMvigor cancer samples to consensus molecular classes was done with the nearest centroid algorithm fed with ComBat-normalized whole-genome log\_2 mRNA expression levels (R package *consensusMIBC*) (6).

JSON files with domain structures of FGFR1 (ID: P11362), FGFR2 (ID: P21802), FGFR3 (ID: P22607), and FGFR4 (ID: P22455) proteins were downloaded from [UniProt](https://www.uniprot.org/) and processed with the *jsonlite* package (15).

## Descriptive statistics, statistical inference, and effect size

If not indicated otherwise, descriptive statistics for quantitative variables included median, interquartile range, and range. Descriptive statistic for qualitative variables were percentage and count within categories. Differences of means or medians of qualitative variables were investigated by Mann-Whitney, tow-tailed T test, Kruskal-Wallis test, and one-way ANOVA, depending on the number of analysis groups and normality of distribution of the variable. Correlation was assessed by Pearson’s correlation test. Effect size of differences in variable levels, distribution, and correlation was assessed with the following statistics (16):

* r for differences between two groups, non-parametric tests: small for r < 0.2, medium for 0.2 r < 0.4, large for r 0.4
* Cohen’s d for differences between two groups, parametric tests: small for d < 0.5, medium for 0.5 d < 0.8, large for d 0.8
* for differences between three or more groups: small for < 0.06, moderate for 0.06 < 0.14, large for 0.14
* Cramer’s V for comparison of category frequencies of a numeric variable: small for r < 0.2, medium for 0.2 r < 0.4, large for r 0
* Pearson’s for correlation, parametric tests: small for < 0.3, medium for 0.3 < 0.5, large for r 0.5

Hypothesis testing results were adjusted for multiple comparisons separately for each analysis step and investigated cohort with the false discovery rate method (FDR) (17). Differences with p < 0.05 were considered statistically significant. Statistical hypothesis testing was accomplished with the development packages [*ExDA*](https://github.com/PiotrTymoszuk/ExDA) and (18,19) internally employing statistical tools of base R, *stats*, *rstatix*, and *rcompanion* (18,19).

## Survival analysis

Differences in overall, disease-specific, and relapse-free survival between patients stratified by genetic features such as *FGFR3* mutations of genetic subsets were investigated with Peto-Peto test. Fractions of surviving patients were visualized with Kaplan-Meier plots. Survival analyses and visualizations were done with packages *survival* and *survminer* (20,21).

## Co-occurrence of genetic features

Co-occurrence analysis was performed in two settings: (1) for the most common genetic features (mutations, amplifications, deletions) present in at least 5% of cancer samples in all cohorts and (2) for all available mutations, amplifications, and deletions of FGF- and FGFR-coding genes. Co-occurrence of genetic features was investigated by Jaccard’s similarity coefficient J and Jaccard’s distance computed with the formula. Significance of co-occurrence, i.e. , was assessed by bootstrap test with B = 10000 iterations. Mutual co-occurrence (e.g. between *FGFR3* mutations and *RB1* mutations) was depicted as a heat map of contingency table.

For visualization of co-occurrence of multiple genetic features in two-dimensional Euclidean space, matrices of pairwise Jaccard’s distances were subjected to multi-dimensional scaling (MDS). Subsequently, the genetic features of interest were displayed in scatter plots with point size corresponding to the general frequency of the genetic alteration.

Co-occurrence analysis was done with [*clustTools*](https://github.com/PiotrTymoszuk/clustTools) and *jaccard* R packages (22).

## Genetic subsets of urothelial cancers

Definition and evaluation of genetic subsets of urothelial cancers was done for the GENIE BLCA training cohort, and the MSK IMPACT and TCGA BLCA test collectives, for which both mutation and copy number variant data were provided.

Genetic subsets of urothelial cancers were developed for the GENIE BLCA training cohort with latent class modeling (23,24). The explanatory variables were binary indexes (presence/absence) of genetic alterations present in at least 5% of samples in all data sets. Presence of FGF3, FGF4, FGF19, or CCND1 amplification was subsumed under 11q13 amplification prior to the modeling. Presence of CDKN2A, CDKN2B, or MTAP deletions was subsumed under 9p21 deletion prior to the modeling. The latent class models were fit with R function tune\_lca() (package [*polcaExtra*](https://github.com/PiotrTymoszuk/polcaExtra)) with the maximal iteration number set to 3000 and numbers of replicates, i.e. random starts of the modeling procedure, set to 5. Selection of the optimal subset number was motivated by the minimum of Bayesian Information Criterion (BIC). With this selection criterion, a six-subset model whose classes were named after characteristic gene and chromosome alterations was chosen for definition of the genetic subsets.

Subsequently, assignment of cancer samples from the MSK IMPACT and TCGA BLCA test collectives to the genetic subsets was predicted based on conditional subset assignment probabilities using predict() method from [*polcaExtra*](https://github.com/PiotrTymoszuk/polcaExtra) package. Quality of the subset assignment in the training and test collective was assessed by a visual inspection of the class assignment probabilities (posterior p) and by comparison of frequency of the subset-defining genetic features between the subsets ( test with Cramer’s V effect size statistic).

## Surviival modeling

For each of the GENIE BLCA and MSK IMPACT cohorts, two types of Cox proportional hazard models of overall survival were constructed: (1) a model with genetic subset assignment as a sole explanatory factor, and (2) a model with genetic subset assignment and clinically relevant confounders (GENIE BLCA: age and gender, MSK IMPACT: age, gender, and pathological tumor stage). The Cox models were constructed, checked for the proportional hazard assumption and evaluated with tools provided by R packages *survival* (21,25) and [*coxExtensions*](https://github.com/PiotrTymoszuk/coxExtensions). Overall Cox model performance was evaluated by Harrell’s C index (high values: good concordance between the predicted and observed survival), integrated Brier score (IBS; low values: good model calibration), and R^2 as a measure of survival variance explained by the model. Significance of the genetic subset term and its independence from clinical confounders was determined by likelihood ratio test between the model with subset assignment and the model with cluster assignment and clinical confounders as explanatory factors (function anova()). Estimates of model coefficients were retrieved with summary() method.

# Data and code availability

Data from publicly available sources were analyzed. The R analysis pipeline is available as a [GitHub repository](https://github.com/PiotrTymoszuk/FGFR-BLCA)

# References

1. Sweeney SM, Cerami E, Baras A, Pugh TJ, Schultz N, Stricker T, Lindsay J, Del Vecchio Fitz C, Kumari P, Micheel C, et al. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer discovery* (2017) 7:818–831. doi: [10.1158/2159-8290.CD-17-0151](https://doi.org/10.1158/2159-8290.CD-17-0151)

2. Clinton TN, Chen Z, Wise H, Lenis AT, Chavan S, Donoghue MTA, Almassi N, Chu CE, Dason S, Rao P, et al. Genomic heterogeneity as a barrier to precision oncology in urothelial cancer. *Cell Reports* (2022) 41: doi: [10.1016/j.celrep.2022.111859](https://doi.org/10.1016/j.celrep.2022.111859)

3. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* (2018) 173:400–416.e11. doi: [10.1016/J.CELL.2018.02.052](https://doi.org/10.1016/J.CELL.2018.02.052)

4. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* (2017) 171:540–556.e25. doi: [10.1016/J.CELL.2017.09.007](https://doi.org/10.1016/J.CELL.2017.09.007)

5. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, Loriot Y, Necchi A, Hoffman-Censits J, Perez-Gracia JL, et al. Atezolizumab as first-line therapy in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet (London, England)* (2017) 389:67. doi: [10.1016/S0140-6736(16)32455-2](https://doi.org/10.1016/S0140-6736(16)32455-2)

6. Kamoun A, Reyniès A de, Allory Y, Sjödahl G, Robertson AG, Seiler R, Hoadley KA, Groeneveld CS, Al-Ahmadie H, Choi W, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *European urology* (2020) 77:420. doi: [10.1016/J.EURURO.2019.09.006](https://doi.org/10.1016/J.EURURO.2019.09.006)

7. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. Welcome to the Tidyverse. *Journal of Open Source Software* (2019) 4:1686. doi: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686)

8. Henry L, Wickham Hadley. rlang: Functions for Base Types and Core R and ’Tidyverse’ Features. (2022) <https://cran.r-project.org/web/packages/rlang/index.html>

9. Gagolewski M, Tartanus B. Package ’stringi’. (2021) <https://cran.r-project.org/web/packages/stringi/index.html http://cran.ism.ac.jp/web/packages/stringi/stringi.pdf>

10. Wickham Hadley. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. New York: Springer-Verlag (2016). [https://ggplot2.tidyverse.org](https://ggplot2.tidyverse.org/)

11. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st ed. Sebastopol: O’Reilly Media (2019).

12. Gohel D. flextable: Functions for Tabular Reporting. (2022) <https://cran.r-project.org/web/packages/flextable/index.html>

13. Whiting K. cbioportalR: R package to wrap cBioPortal’s API to pull data from public or private cBioPortal databases. (2024) <https://github.com/karissawhiting/cbioportalR>

14. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, Koeppen H, Astarita JL, Cubas R, et al. TGF attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature 2018 554:7693* (2018) 554:544–548. doi: [10.1038/nature25501](https://doi.org/10.1038/nature25501)

15. Ooms J. A Simple and Robust JSON Parser and Generator for R [R package jsonlite version 1.8.8]. (2023) <https://cran.r-project.org/package=jsonlite>

16. Cohen J. Statistical Power Analysis for the Behavioral Sciences. *Statistical Power Analysis for the Behavioral Sciences* (2013) doi: [10.4324/9780203771587](https://doi.org/10.4324/9780203771587)

17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* (1995) 57:289–300. doi: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)

18. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation. (2022) <https://cran.r-project.org/package=rcompanion>

19. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2021) <https://cran.r-project.org/package=rstatix>

20. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using ’ggplot2’. (2016) <https://cran.r-project.org/package=survminer>

21. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. 1st ed. New York: Springer Verlag (2000).

22. Chung NC, Miasojedow B, Startek M, Gambin A. jaccard: Test Similarity Between Binary Data using Jaccard/Tanimoto Coefficients. (2018) <https://cran.r-project.org/web/packages/jaccard/index.html>

23. Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology* (2020) 46:287–311. doi: [10.1177/0095798420930932/ASSET/IMAGES/LARGE/10.1177\_0095798420930932-FIG1.JPEG](https://doi.org/10.1177/0095798420930932/ASSET/IMAGES/LARGE/10.1177_0095798420930932-FIG1.JPEG)

24. Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software* (2011) 42:1–29. doi: [10.18637/jss.v042.i10](https://doi.org/10.18637/jss.v042.i10)

25. Grambsch PM, Therneau TM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* (1994) 81:515. doi: [10.2307/2337123](https://doi.org/10.2307/2337123)