The consensus molecular classification of muscle-invasive bladder cancer

Authors

Aurélie Kamoun[1]§, Aurélien de Reyniès[1]*, Yves Allory[2]*, Gottfrid Sjödahl[3]*, A. Gordon Robertson[4]*, Roland Seiler[5], Katherine A. Hoadley[6], Hikmat Al-Ahmadie[7], Woonyoung Choi[8], Clarice S. Groeneveld[9], Mauro A. A. Castro[9], Jacqueline Fontugne[2], Pontus Eriksson[10], Qianxing Mo[11], Jordan Kardos[6], Alexandre Zlotta[12], Arndt Hartmann[13], Colin P. Dinney[14], Joaquim Bellmunt[15], Thomas Powles[16], Núria Malats[17], Keith S. Chan[18], William Y. Kim[19], David J. McConkey[20], Peter C. Black[21], Lars Dyrskjøt[22], Mattias Höglund[10], Seth P. Lerner[23], Francisco X. Real[24], François Radvanyi[25], The Bladder Cancer Molecular Taxonomy Group[+]

§ corresponding author, *these authors contributed equally to this work
+ A full list of all consortium members appears at the end of the article

1. Cartes d'Identité des Tumeurs Program, French League Against Cancer, Paris, France

2. Department of Pathology, Institut Curie Hospital Group, Paris, France

3. Division of Urological Research, Department of Translational Medicine, Lund University, Skåne University Hospital Malmö, Sweden

4. Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, Canada

5. Department of Urology, Bern University Hospital, Switzerland

6. Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

7. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

8. Johns Hopkins Greenberg Bladder Cancer Institute and Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA

9. Bioinformatics and Systems Biology Laboratory, Federal University of Paraná, Polytechnic Center, Curitiba, Brazil

10. Division of Oncology and Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden

11. Department of Medicine, Baylor College of Medicine, Houston, TX, USA

12. Department of Surgery, Division of Urology, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, ON, Canada

13. Institute of Pathology, University Erlangen-Nürnberg, Krankenhausstr 8-10, Erlangen, Germany

14. Department of Urology and Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

15. Bladder Cancer Center, Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, MA, USA

16. Barts Cancer Institute ECMC, Barts Health and the Royal Free NHS Trust, Queen Mary University of London, London, UK

17. Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

18. Molecular & Cellular Biology/Scott Department of Urology, Baylor College of Medicine, One Baylor Plaza, Houston, TX, USA

19. Department of Genetics, Department of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

20. Johns Hopkins Greenberg Bladder Cancer Institute and Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA

21. Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada

22. Department of Molecular Medicine, Aarhus University Hospital, Aarhus 8200, Denmark

23. Scott Department of Urology, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA

24. Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

25. Molecular Oncology, CNRS UMR 144, Institut Curie, Paris, France

Abstract

Muscle-Invasive Bladder Cancer (MIBC) is a molecularly diverse disease with heterogeneous clinical outcomes. Several molecular classifications have been proposed, yielding diverse sets of subtypes, which hampers the clinical implications of such knowledge. Here, we report the results of a large international effort to reach a consensus on MIBC molecular subtypes. Using 1750 MIBC transcriptomes and a network-based analysis of six independent MIBC classification systems, we identified a consensus set of six molecular classes: Luminal Papillary (24%), Luminal Non-Specified (8%), Luminal Unstable (15%), Stroma-rich (15%), Basal/Squamous (35%), and Neuroendocrine-like (3%). These consensus classes differ regarding underlying oncogenic mechanisms, infiltration by immune and stromal cells, and histological and clinical characteristics. This consensus system offers a robust framework that will enable testing and validating predictive biomarkers in future clinical trials.

Bladder cancer is one of the most frequently diagnosed cancers in North America and Europe (4th in men and 9th in women). Most bladder cancers are urothelial carcinoma, which are classified for operational reasons as either non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC). MIBC is usually diagnosed *de novo*, but may arise from the 10 to 20% of NMIBC cases that eventually progress. MIBC is the most aggressive disease state and is associated with a five-year survival rate of 60% for patients with localized disease, and less than 10% for patients with distant metastases.

At the molecular level, MIBC is a heterogeneous disease that is characterized by genomic instability and a high mutation rate. Many chromosomal rearrangements and more than 50 oncogenes and tumour suppressor genes have been identified as recurrently altered[1,2]. Transcriptomic profiling facilitates stratifying bladder cancer into molecular subtypes in order to more precisely classify a patient's cancer according to prognosis and therapeutic options. Various teams have been working on the molecular stratification of bladder cancers, and several expression-based classification schemes have been proposed, either considering the full spectrum of bladder cancers[3–6] or focusing separately on MIBC[2,7–13] or on NMIBC[14]. These classifications have considerably advanced our understanding of bladder cancer biology; for example, the association between molecular subtypes and urothelial differentiation, and similarities between subtypes in bladder cancer and other cancers. In addition, specific genomic alterations were found to be enriched in particular molecular subtypes, including mutations targeting genes involved in cell cycle regulation, chromatin remodelling and receptor tyrosine kinase signaling. Importantly, several reports have highlighted the clinical importance of MIBC molecular stratification, suggesting that responses to chemotherapy and immunotherapy may be enriched in specific MIBC subtypes[9,15–17].

The published MIBC classifications share many characteristics, including subtype-specific molecular features; however, the classifications are diverse, containing between two to seven molecular subtypes, and having both shared and unique subtype names. This diversity has hampered transferring subtypes into clinical practice, and highlights that identifying a single set of consensus molecular subtypes would facilitate work to achieve such a transfer.

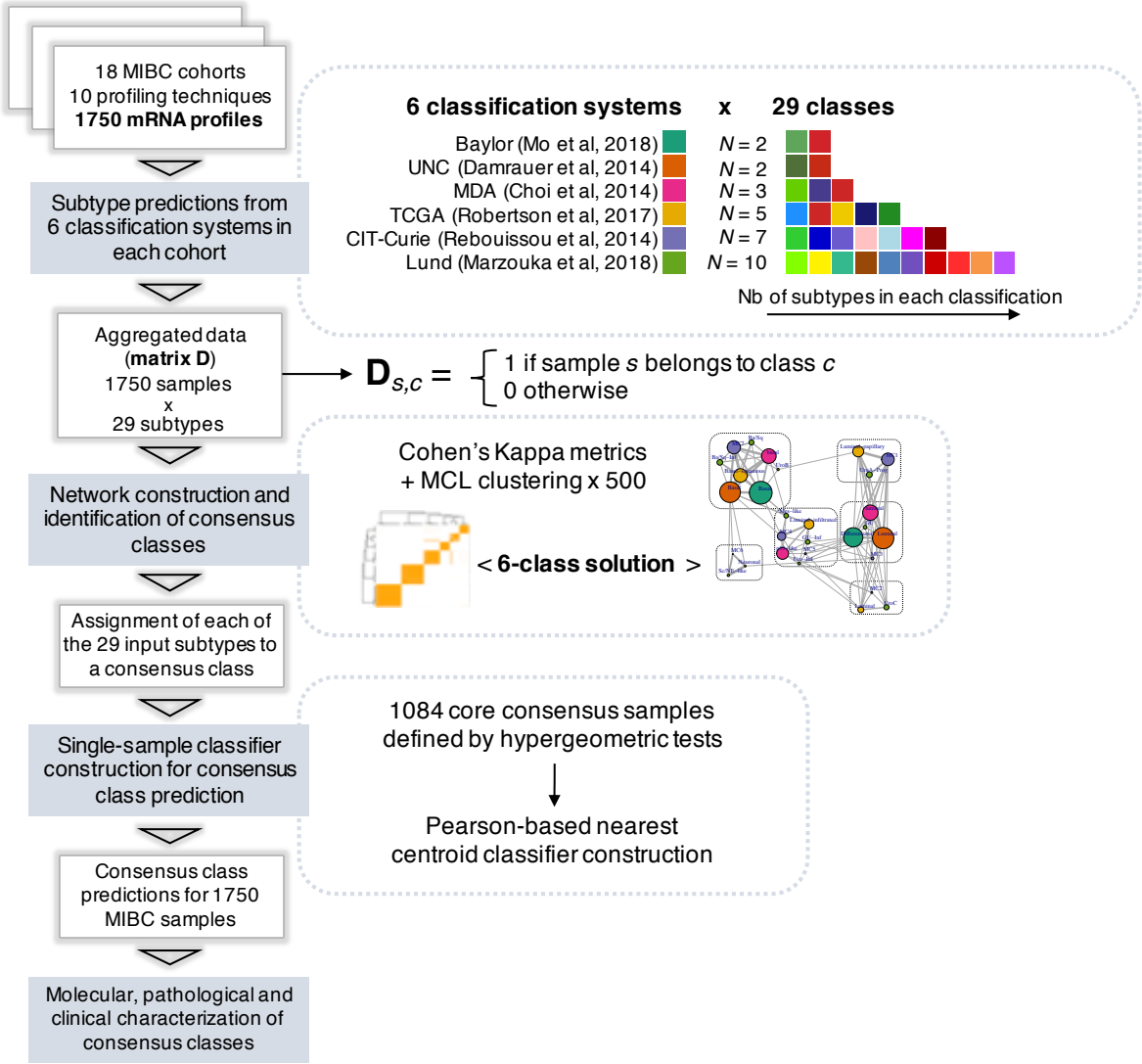
Here, we report the results of an international collaborative effort to reconcile molecular MIBC classifications, involving pathologists, urologists, oncologists, biologists, and bioinformaticians. By analysing six previously published classification schemes and combining public transcriptome data for 1750 tumours, we established a six-class, consensus molecular classification for MIBC. We characterized the consensus classes using additional molecular, pathological and clinical data. To support the use of this consensus molecular classification, we offer a freely available transcriptomic classifier that assigns consensus class labels to single tumour samples (https://github.com/cit-bioinfo/consensusMIBC).

Results

Published molecular classifications of MIBC converge on six classes.

We used six published MIBC molecular classifications to define a unified consensus subtyping system. We refer to these input classifications as Baylor (Tumour differentiation)[13], UNC[7], CIT-Curie[8], MDA[9], Lund[10], and TCGA[2]. Following the approach outlined in Extended data figure 1, we selected 18 MIBC mRNA datasets (n=1750, Supplementary Table 1), and assigned each sample to a subtype in each of the six classification systems.

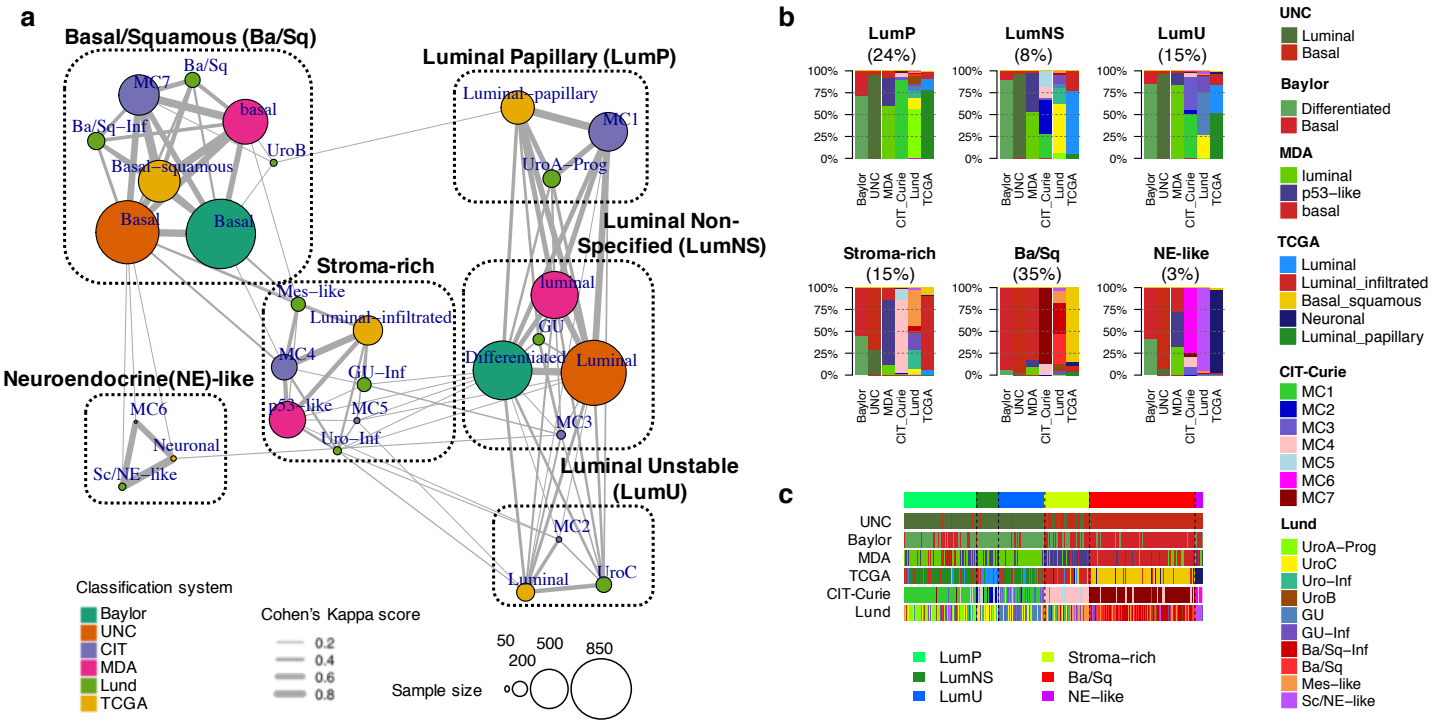
Extended data figure 1 : Analytical workflow



We built a weighted network of these input subtypes, using Cohen's Kappa metric to quantify similarities between subtypes from different classification systems, and applied a Markov cluster clustering algorithm (MCL) to identify robust network substructures corresponding to potential consensus classes (Methods, Supplementary Figure 1). We identified a 6-cluster solution, defining six biologically relevant consensus molecular classes, which we labeled as: Luminal Papillary (LumP), Luminal Non-Specified (LumNS), Luminal Unstable (LumU), Stroma-rich, Basal/Squamous

(Ba/Sq), and Neuroendocrine-like (NE-like) (Figure 1a). Considerations motivating our choices for these consensus names are detailed in the Supplementary Note.

Figure 1 : The six consensus classes and their relation to input molecular subtypes.



The six molecular classes had variable sample sizes, with Ba/Sq and LumP being the most prevalent (35% and 24% of all samples, respectively). The remaining 41% of samples were split into LumU (15%), Stroma-rich (15%), LumN (8%), and NE-like (3%) tumours (Figure 1b). The consensus classification was strongly associated with each of the initial classification systems (Chi-square P<10-165), as illustrated in Figure 1 and Supplementary Figure 2.

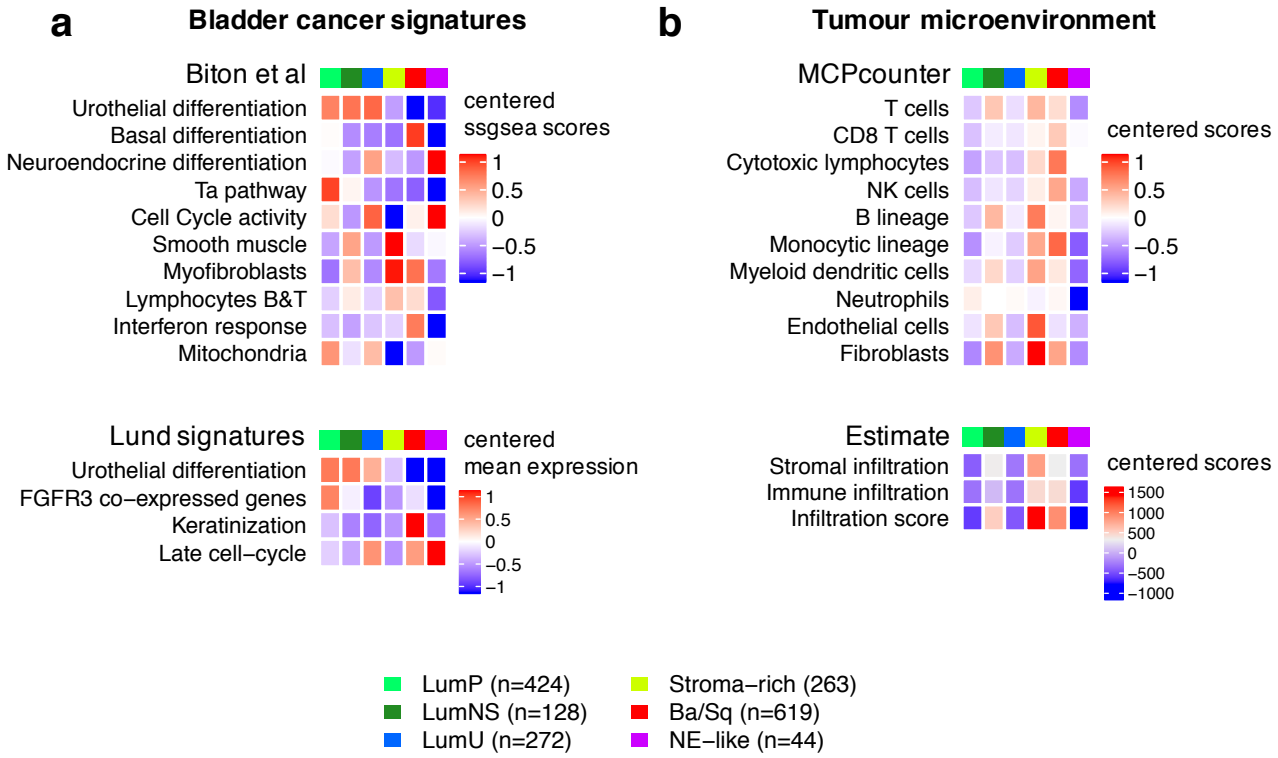
We compared the consensus classes to the 15 TCGA pan-cancer integrative clusters[18] that contained MIBC tumours (Supplementary Figure 2b). We observed enrichments between the Ba/Sq consensus class and the squamous cell carcinoma C27:Pan-SCC pan-cancer cluster (P=1.10x10^{-11}), and between the Stroma-rich class

and the stroma-driven C20:Mixed(Stromal/Immune) pan-cancer cluster (P=<2.2x10^{-16}).

Transcriptomic characterization of the six consensus molecular classes

We used mRNA data from all 1750 samples to characterize consensus classes with published molecular gene signatures for bladder cancer pathways and for tumour microenvironment infiltration (Figure 2, Supplementary Table 2).

Figure 2 : Characterization of tumour and stroma signals using published mRNA signatures



Differentiation-associated mRNA signatures were strongly associated with the consensus classes. Tumours from the three luminal classes overexpressed urothelial differentiation signatures (P<10^{-16}), including the PPARG/GATA3/FOXA1-related Lund signature[19]. In contrast, Ba/Sq and NE-like tumours respectively overexpressed gene

signatures associated with basal ($P<10^{-16}$) and neuroendocrine differentiation ($P=4.2x10^{-16}$).

In addition to their urothelial differentiation status, the three luminal classes exhibited distinct molecular signatures. LumP tumours were characterized by high expression of a non-invasive Ta pathway signature[20] ($P<10^{-16}$) and were strongly associated with FGFR3 transcriptional activity as measured by an FGFR3 co-expressed genes signature[5] ($P<10^{-16}$). LumNS tumours displayed elevated stromal infiltration signatures, mainly fibroblastic, as compared to the other luminal tumours ($P<10^{-16}$). LumU tumours had a high cell cycle activity, and notably overexpressed a late cell cycle signature ($P<10^{-16}$).

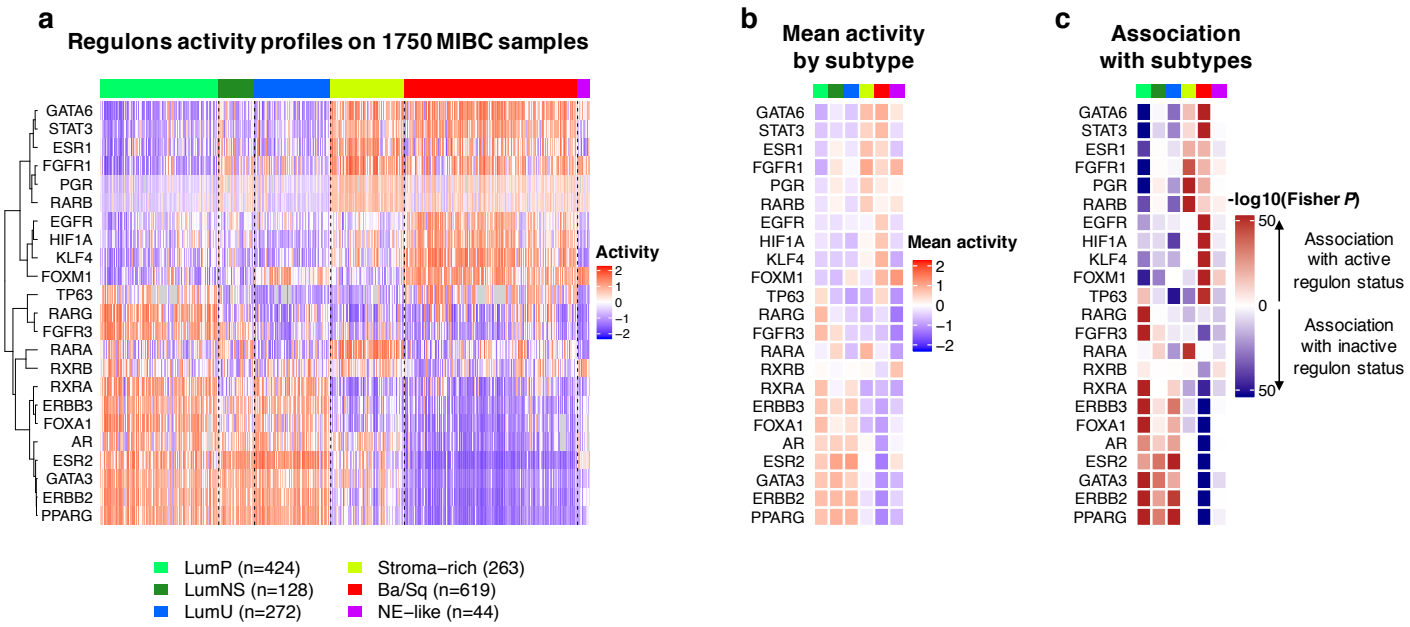
Stroma-rich tumours displayed intermediate and heterogeneous levels of urothelial differentiation. They were mainly characterized by stromal infiltration as summarized by ESTIMATE[21] stromal scores, with a specific overexpression of smooth muscle and endothelial cell signatures ($P<10^{-16}$). Fibroblasts and myofibroblasts signatures were also overexpressed within the Stroma-rich tumours ($P<10^{-16}$).

Immune infiltration was mainly found within Ba/Sq and Stroma-rich tumours, but the two classes were characterized by distinct immune cell populations, as measured by MCPcounter signatures[22]. Ba/Sq tumours were enriched in cytotoxic lymphocytes and NK cells ($P<10^{-16}$), whereas Stroma-rich tumours overexpressed T cell and B cell markers ($P<10^{-16}$). LumNS tumours were the only luminal tumours associated with moderate immune infiltration signals (mainly B and T lymphocytes). We detected no transcriptomic markers of immune infiltration in NE-like tumours.

Analysis of regulatory units (i.e. regulons) for 23 regulator genes previously reported as associated with bladder cancer[2,23] were consistent with the mRNA signatures assessed (Extended data figure 2). *PPARG* and *GATA3* regulons were

activated within the luminal tumours, which overexpressed strong urothelial differentiation signals. The *FGFR3* regulon was specifically activated within LumP tumours, and Ba/Sq tumours showed the strongest association with the *STAT3* regulon activation, consistent with their expressing a keratinization gene signature. Additionally, the regulon analysis showed an elevated *HIF1A* activity specifically in Ba/Sq tumours, suggesting that this class is associated with a highly hypoxic microenvironment. *EGFR* activity was also specifically associated with Ba/Sq tumours, as previously reported[8].

Extended data figure 2 : **Regulons activity within consensus classes.**

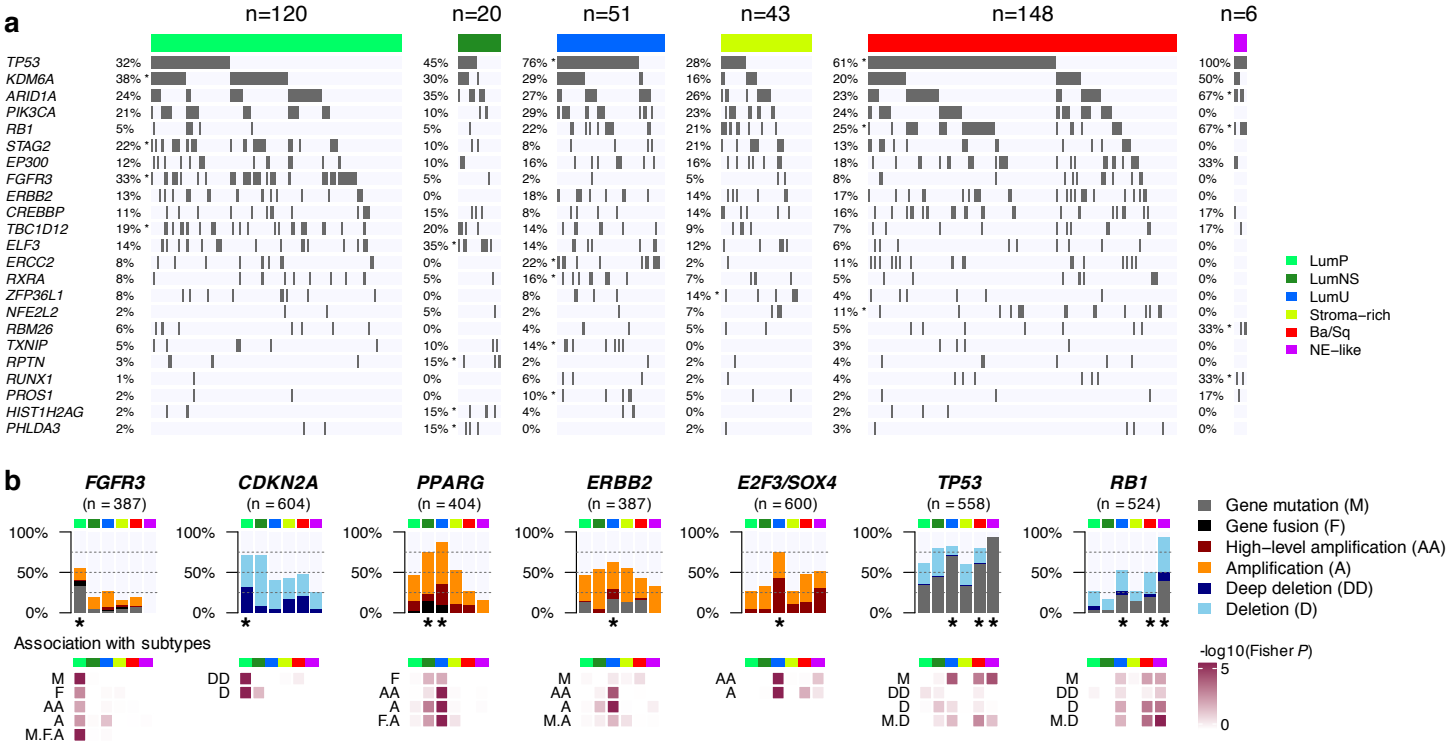


Genomic alterations associated with the consensus molecular classes

We used TCGA exome data to identify class-specific mutations (Figure 3a, Supplementary Table 3) and ran GISTIC2[24] on 600 available copy number profiles grouped by consensus class to identify class-specific copy number variations (CNV) (Supplementary Table 4). In addition, we combined all CNV, gene fusion, and gene mutation data from the 18 cohorts to generate comprehensive profiles of genomic

alterations by consensus class, for seven key bladder cancer key genes: *FGFR3*, *CDKN2A*, *PPARG*, *ERBB2*, *E2F3*, *TP53* and *RB1* (Figure 3b).

Figure 3 : Genomic alterations associated with consensus classes

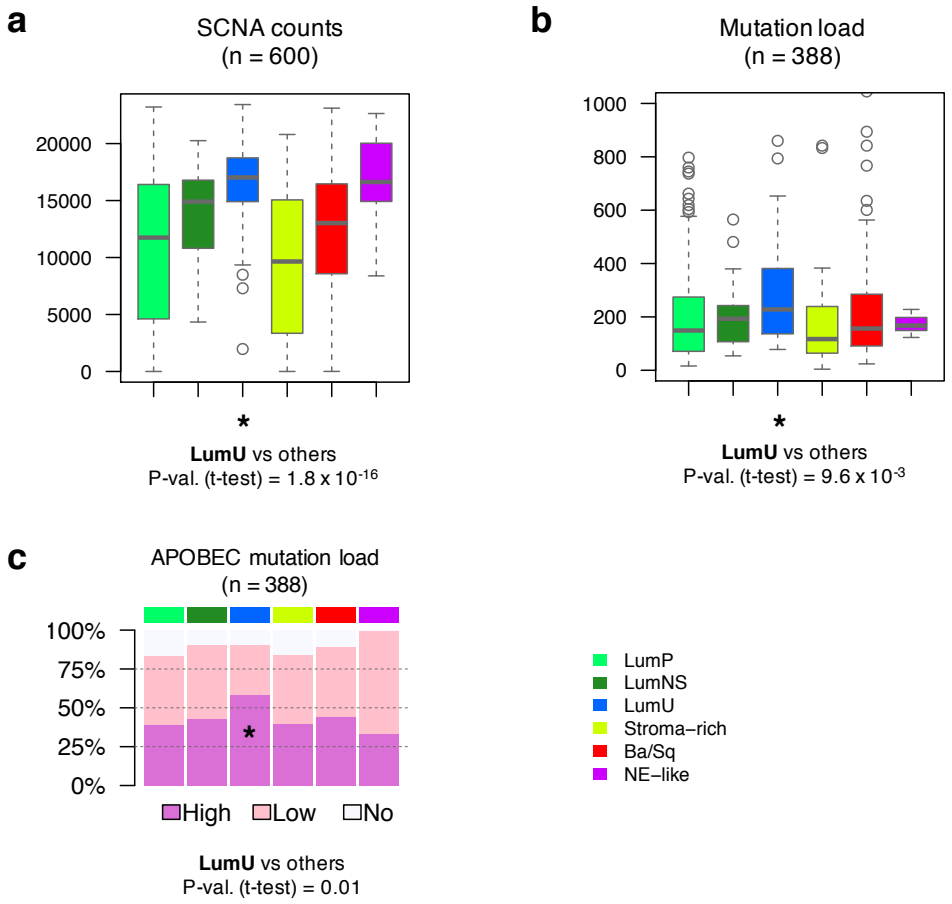


LumP tumours were enriched in *FGFR3* (P=1.4x10^{-11}), *KDM6A* (P=0.002) and *STAG2* mutations (P=0.01). Aggregating data for 643 LumP tumours, the proportion of *FGFR3*-mutated tumours reached 40% (P=1.6x10^{-23}). Assembling mutations, fusions, and copy number amplifications, *FGFR3* alterations were enriched in LumP tumours (P=1.9x10^{-11}). *CDKN2A* MLPA (Multiplex Ligation-dependent Probe Amplification) and CNV data for 604 tumours revealed 33% of *CDKN2A* homozygous/deep deletions in LumP tumours, corresponding to a strong enrichment as compared to other tumours (P=3.8x10^{-8}). These deletions were consistent with the enrichment of LumP tumours within the TCGA pan-cancer iCluster C7:Mixed(Chr9 del) (P=1.6x10^{-10}), which is characterized by Chr 9 deletions (Supplementary Figure 2b).

The LumNS class was mainly characterized by an enrichment of mutations in *ELF3* (35%, *P*=0.004), which is an early regulator of normal urothelium, and is activated by PPARγ[25]. *PPARG* was significantly altered as well, with 76% of LumNS tumours harbouring either amplifications or fusions involving this gene ($P=5.7\times10^{-3}$).

LumU tumours also harboured frequent *PPARG* alterations (89%, $P=1.9\times10^{-11}$), and high-level amplifications of a 6p22.3 region that contains *E2F3* and *SOX4* (76%, $P=3.0\times10^{-12}$). Genomic amplifications of *ERBB2* were overrepresented in LumU tumours ($P=4.3\times10^{-8}$), but no significant association was found between *ERBB2* mutations and any of the consensus classes. In contrast with the other luminal tumours, LumU tumours were associated with *TP53* mutations (76%, $P=3.4\times10^{-5}$), and with mutations in the core nucleotide-excision repair gene *ERCC2* (22%, *P*=0.006). More generally, LumU was the most genomically altered class (Extended data figure 3), displaying the highest number of copy number alterations ($P=1.8\times10^{-16}$), the highest somatic mutation load (*P*=0.009), and including more APOBEC-induced mutations than other consensus classes (*P*=0.01). These features of genomic instability and the association with *ERBB2* amplifications were consistent with the enrichment of LumU tumours within the TCGA pan-cancer subtypes C2:BRCA(HER2 amp) (characterized by frequent *ERBB2* amplifications, $P=4.0\times10^{-5}$) and C13:Mixed(Chr8 del) (enriched in highly aneuploid tumours, $P=3.8\times10^{-9}$) (Supplementary Figure 2b).

Extended data figure 3: Distributions of SCNA, and total somatic and APOBEC mutation loads across consensus classes



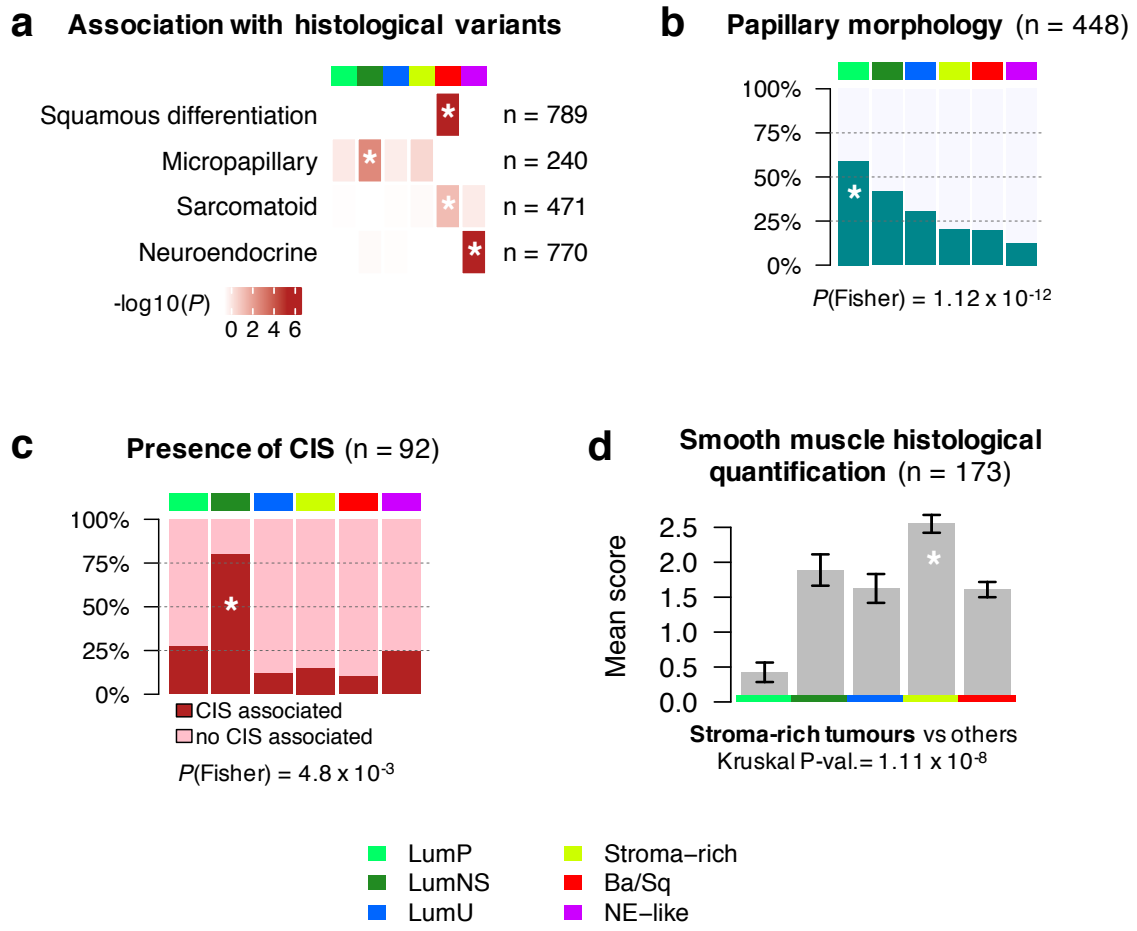
For Ba/Sq tumours, as shown previously[26], the most frequently mutated genes were *TP53* (P=5.8x10^{-4}), *NFE2L2* (P=0.002) and *RB1* (P=0.002). Aggregated mutation data revealed that 58% (134/232, P=0.009) and 20% (43/224, P=0.007) of Ba/Sq tumours contained mutations in *TP53* and *RB1*, respectively. These mutations co-occurred in 14% (32/224) of cases. Ba/Sq tumours were also strongly associated with genomic deletions of a 3p14.2 region, which occurred in 49% of cases (P = 1.5x10^{-13}).

Finally, combining all available data on *TP53* and *RB1* genomic alterations, we observed a strong enrichment of *TP53* and *RB1* inactivation in NE-like tumours. *TP53* was ubiquitously mutated in these tumours (94%, P=9.7x10^{-5}), and co-occurred with *RB1* inactivation by either mutations or deletions (94%, P=2.2x10^{-6}).

Histological patterns associated with the consensus molecular classes

To characterize the consensus molecular classes from a histological perspective, we assembled sample annotations for urothelial histological variants and specific morphologic patterns (Figure 4). As expected, Ba/Sq tumours included 79% of histologically reviewed tumours with squamous differentiation (126/159, P=3.6x10^{-32}, Supplementary Figure 3). The Ba/Sq class did however extend beyond this histological subtype, with only 42% (126/303) of Ba/Sq tumours associated with squamous differentiation. Similarly, NE-like tumours were strongly associated with neuroendocrine variant histology, with 72% of histologically reviewed NE-like tumours showing neuroendocrine differentiation (13/18, P=9.7x10^{-22}). LumP tumours were enriched with papillary morphology as compared to other consensus classes (P=1.2x10^{-12}). This pattern was observed in 59% (82/139) of histologically reviewed LumP tumours, although frequently found in other luminal tumours (42% in LumNS and 31% in LumU tumours). LumNS tumours were enriched in micropapillary variant histology (36%, 9/25, P=0.001) and with the presence of carcinoma *in situ* (CIS) lesions (80%, 4/5, P=0.005).

Figure 4 : Histopathological associations with consensus classes



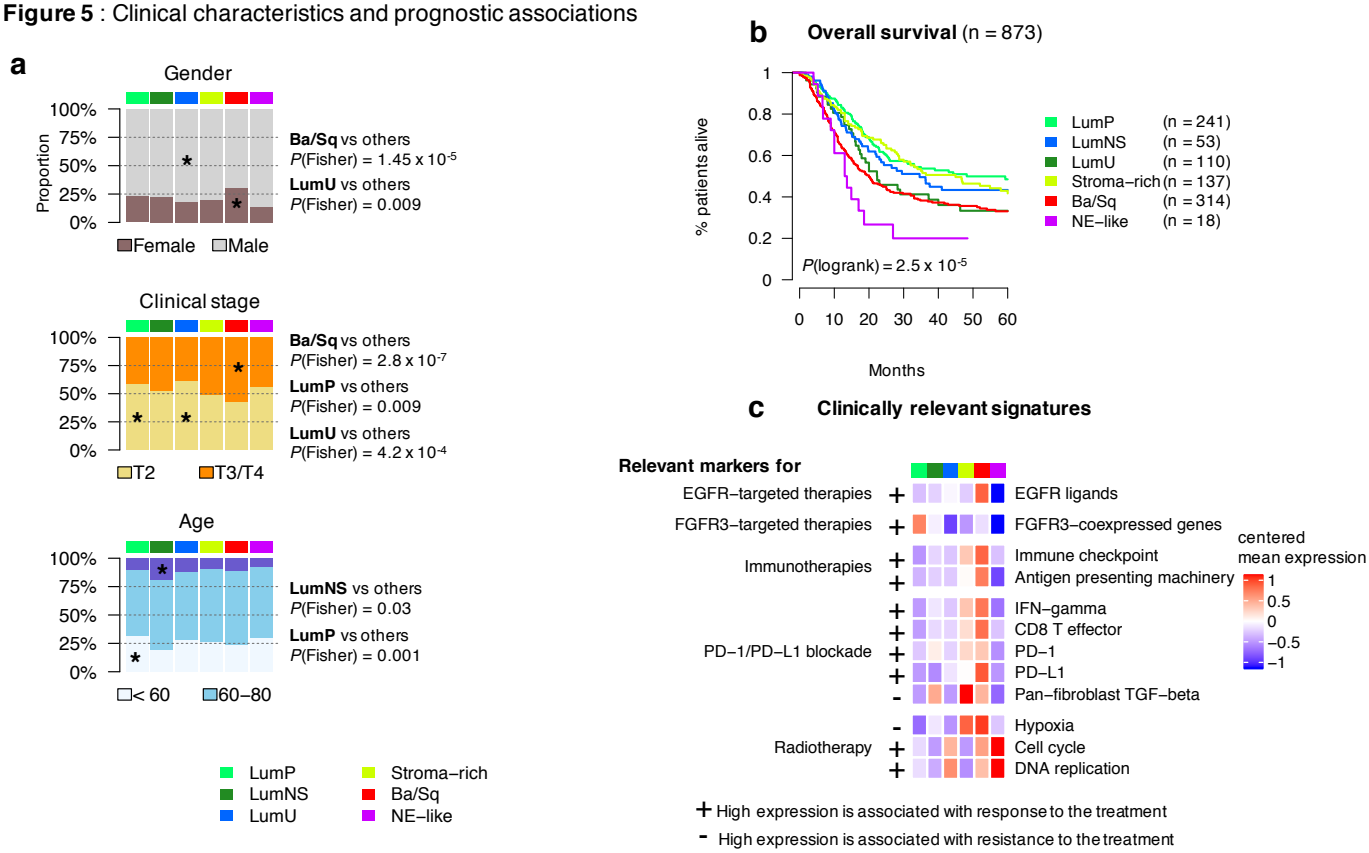
A pathological review of stromal infiltration in TCGA tumour sample slide images confirmed that Stroma-rich tumours were associated with a higher proportion of smooth muscle cells (Kruskal P=1.1x10^{-8}), consistent with the strong smooth muscle-related mRNA expression characterizing these tumours.

The consensus molecular classes are associated with distinct clinical characteristics, survival outcomes, and therapeutic opportunities.

We confirmed previously reported associations with gender, stage, and age (Figure 5a), such as the overrepresentation of Ba/Sq tumours in females and in higher clinical stages (P=1.4x10^{-5} and P=2.8x10^{-7} respectively). The LumP and LumU

consensus classes were enriched in T2 vs T3-4 tumours (P=0.009 and P=4.2x10^{-4}) as compared to other classes. Patients less than 60 years old were overrepresented among LumP tumours (P=0.001), whereas the LumNS consensus class was enriched with older patients (> 80 years old; P=0.03).

Figure 5 : Clinical characteristics and prognostic associations



Overall survival was strongly associated with the consensus classes (Figure 5b, P=2.5x10^{-5}). Patients with LumP tumours had the best prognosis when compared to all consensus classes (HR=0.65, P=2.1x10^{-4}, Supplementary Table 5a). The two other luminal classes were associated with poorer prognoses ($HR_{LumNS/LumP}$=1.51, P=4.7x10^{-2}; and $HR_{LumU/LumP}$=1.32, P=0.12), although the differences were modest or not significant in this setting. Despite the variable differentiation states among samples from the Stroma-rich class, patients with these tumours showed a similar overall survival to that associated with LumP tumours ($HR_{Stroma\text{-}rich/LumP}$=1.18, CI_{95} = [0.85,

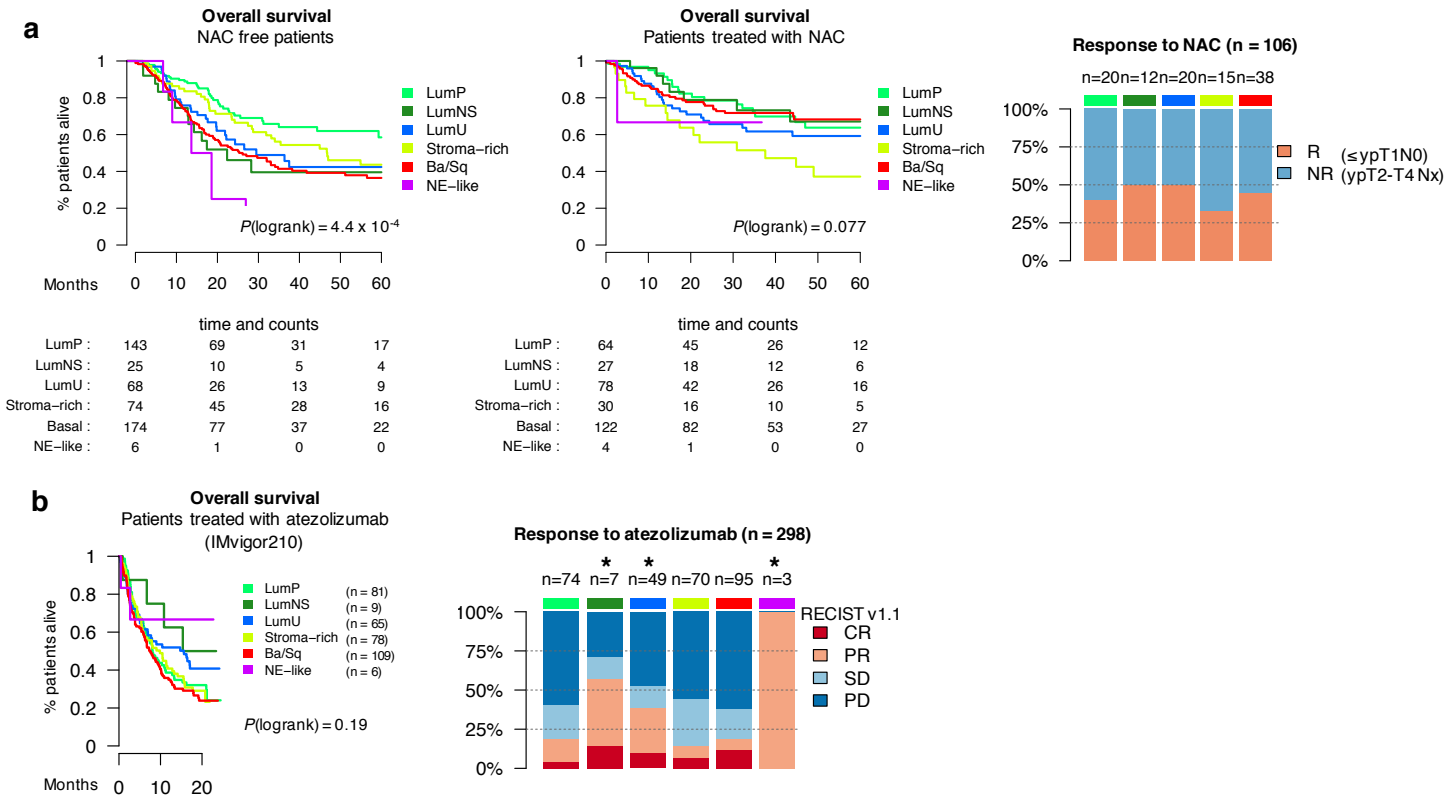
1.63]). Ba/Sq tumours were associated with a poor prognosis ($HR_{BaSq/LumP}$=1.8, P=5.7x10^{-6}), consistent with previous studies. Finally, NE-like tumours were associated with the worst prognosis ($HR_{NE\text{-}like/LumP}$=2.4, P=3.3x10^{-3}). Ba/Sq and NE-like consensus classes remained significantly associated with worse overall survival in a multivariate Cox model that combines consensus classes (with the LumP class as reference), TNM, and patient age (respectively P=0.002 and P=0.05, Supplementary Table 5b).

We characterized the consensus classes using several clinically relevant mRNA signatures (Figure 5c, Supplementary Table 6). FGFR3 activity signature was strongly and specifically expressed in LumP tumours, suggesting prospects for FGFR3-targeted therapies within this class. Ba/Sq tumours expressed high levels of the EGFR ligands, which may be associated with a sensitivity to EGFR-targeted therapies, as suggested by previously reported *in vitro* and *in vivo* experiments[8]. Ba/Sq tumours also strongly expressed immune checkpoint markers and antigen-presenting machinery genes, suggesting possibilities for immunotherapies within this class. Studies integrating mRNA signatures with response data to anti-PD1/PD-L1 therapies[17,27] have reported associations of anti-PD1/PD-L1 response with high levels of CD8 T cells, high interferon gamma signals, and low activity of the TGF-beta pathway; however, no consensus class had an expression profile suggesting either response or resistance to anti-PD1/PD-L1 therapies. In contrast, NE-like and LumU tumours both had a profile associated with response to radiotherapy[28,29], showing elevated cell cycle activity and low hypoxia signals.

Finally, we performed a consensus class-based retrospective analysis of clinical outcome from patients receiving neoadjuvant chemotherapy[9,16] (NAC) and patients treated with the anti-PD-L1 atezolizumab[17] (IMvigor210) (Extended data figure 4). Analysis of overall survival and response showed that consensus classes were

associated with distinct responses to the treatments. The results suggested an improved overall survival in the NAC setting for LumNS, LumU and Ba/Sq tumours, and an enrichment in atezolizumab responders in LumNS (*P*=0.05), LumU (*P*=0.003) and NE-like (*P*=0.01) tumours.

Extended data figure 4 : Response to cisplatin-based Neoadjuvant Chemotherapy (NAC) and PD-L1 blockade (atezolizumab)

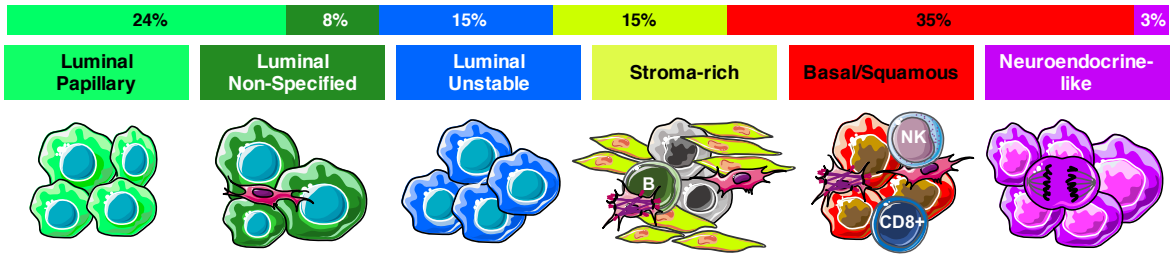


Discussion

The diversity of published MIBC classifications has delayed transferring subtypes into clinical trials or clinical practice. Here, we offer two resources to support work towards such a transfer. First, we analysed the relationships among six different published classification systems, based on 1750 MIBC transcriptomic profiles. We identified six consensus MIBC molecular classes: Basal/Squamous (Ba/Sq) (35%), Luminal Papillary (LumP) (24%), Luminal Unstable (LumU) (15%), Stroma-rich (15%), Luminal Non-Specified (LumNS) (8%), and Neuroendocrine-like (NE-like) (3%). Each

consensus class has distinct differentiation patterns, oncogenic mechanisms, tumour microenvironments, and histological and clinical associations (Figure 6). At this point, NE-like and Ba/Sq classes are the most stably classified, while the three Luminal classes appear to be less clearly defined. Second, we make available an R-based, single-sample classifier that will identify which consensus class a tumour sample's transcriptome corresponds to.

Figure 6 : Summary of main characteristics of the six consensus classes



	Luminal Papillary	Luminal Non-Specified	Luminal Unstable	Stroma-rich	Basal/Squamous	Neuroendocrine-like
Differentiation	Urothelial / Luminal				Basal	Neuroendocrine
Oncogenic mechanisms	FGFR3 ++ CDKN2A -	PPAR-γ ++	PPAR-γ ++ E2F3 +, ERBB2 + Genomic instability		EGFR +	TP53 --, RB1 --, Cell cycle +
Mutations	*FGFR3* (40%), *KDM6A* (38%), *STAG2* (22%)	*ELF3* (35%)	*TP53* (76%), *ERCC2* (22%) TMB +, APOBEC +		*TP53* (61%), *RB1* (25%)	*TP53* (94%) *RB1* (39%)
Stromal infiltrate		Fibroblasts		Smooth muscle Fibroblasts Myofibroblasts	Fibroblasts Myofibroblasts	
Immune infiltrate				B cells	CD8 T cells NK cells	
Histology	Papillary morphology	Micropapillary variants			Squamous differentiation	Neuroendocrine differentiation
Clinical	T2 stage +	Older patients + (80+)			Women + T3/T4 stage +	
Median overall survival (years)	4	1.8	2.9	3.8	1.2	1

This consensus classification fully concurs with MIBC differentiation-based stratification, revealing tumour classes that are primarily characterized by urothelial differentiation (Luminal classes), basal/squamous differentiation (Ba/Sq) and neuroendocrine differentiation (NE-like). Additional features including genomic alterations, and pathological or clinical characteristics are strongly associated with one or several classes (Figure 6).

LumP tumours are mainly characterized by strong transcriptional activation of FGFR3, involving a genetic mechanism in more than 50% of LumP samples (mutation, fusion, amplification). Papillary morphology is more frequent for these tumours (59%), although found in more than 30% of other luminal tumours. LumP tumours strongly express transcriptomic markers of the Ta pathway, and are consistently associated with the best prognosis among all MIBC tumours. These data suggest that these tumours result from progression of papillary Ta/T1 NMIBC.

The LumNS class included a relatively small number of tumours with characterizing data, precluding using a more descriptive name. Nevertheless, our results point to interesting associations, such as an enrichment in *ELF3* mutations (35%, n=7/20, P=0.004) and an association with micropapillary morphology (36%, n=9/25, P=0.001). A weak association with CIS (80%, n=4/5, P=0.005) is also observed for these tumours. The LumNS tumours are the only luminal tumours expressing stromal and immune signals. Their associated prognosis is the worst of the three luminal classes (P=0.05).

LumU tumours display typical features of genomic instability, such as a higher tumour mutation burden that includes more APOBEC-induced mutations, and more copy number alterations. The “Unstable” descriptor for this class refers to the Genomic Unstable tumours from the Lund classification, which are all included within this class. These tumours are particularly enriched in *TP53* (76%) and *ERCC2* (22%) mutations. LumU tumours are associated with high late cell cycle activity, and *ERBB2* activation through mutations or amplification (63%).

Stroma-rich tumours are mainly characterized by high expression of non-tumour cell markers. Smooth muscle cells dominate the infiltration signals associated with these tumours, but endothelial cells and B lymphocytes are also overrepresented. As

assessed by a urothelial differentiation signature[19] and by differentiation-based classification systems, this class contains both luminal and non-luminal tumours (Supplementary Figure 4). However, patients with luminal and non-luminal Stroma-rich tumours have very similar survival, suggesting that although this subgroup is heterogeneous in regards to tumour cell phenotype, stroma could be the main parameter that, given current treatments, drives its clinical features.

Ba/Sq tumours have high expression of basal differentiation markers, and are strongly associated with squamous differentiation. 42% of Ba/Sq tumours have squamous histological variants, and 79% of such variants are observed in Ba/Sq tumours. Although the Ba/Sq tumours are characterized by *KRT14*, *KRT5*/*6* and lack of *GATA3, FOXA1*, and *PPARG*, downregulation of *PPARG* and *GATA3* is not observed in normal basal cells[31]. In this regard the Ba/Sq class is more similar to squamous urothelial metaplasia, consistent which enrichment in the squamous-cell carcinoma-associated C27 pan-cancer iCluster. Ba/Sq tumours express strong fibroblast and myofibroblast infiltration signals, as well as immune infiltration signals from cytotoxic T cells and NK cells. *EGFR* and *STAT3* activation are specific to this class.

The NE-like class includes virtually all (13 of 16, 81%) of tumours with histological neuroendocrine differentiation, and 72% of NE-like tumours have small-cell neuroendocrine variant histology. These tumours show high cell cycle activity, and all have both *TP53* and *RB1* genes inactivated by mutations or deletions. They have the worst prognosis of all MIBC tumours.

We generated the MIBC consensus classification following a procedure similar to that used by Guinney *et al*[32] to identify consensus subtypes in colorectal cancer. Given the diverse nature of the six input classification systems that we used to build

the consensus classes (distinct classification methods, strongly varying numbers of classes), we anticipate that the resulting consensus classification captures most of the molecular heterogeneity described, and that it is currently the best consensus solution for MIBC molecular classification.

Except for the Lund sub-stratification, which used IHC, the original subtype classifications analysed in this study were based on transcriptome data, and mainly considered coding transcripts. Considering other types of DNA, RNA or protein data may refine and subdivide the consensus classes further, helping to decipher the diverse biology and heterogeneity of molecular processes within MIBC.

Some bladder tumours show histological and molecular intra-tumour heterogeneity[33,34]. Our consensus subtyping system addresses inter-tumour heterogeneity and focuses on defining the main molecular subtypes in MIBC. Our transcriptomic classifier will classify tumours according to the dominant class within the tumour sample analysed. However, we recognize that tumour samples may contain multiple subtypes, and we address how such mixtures are likely to interfere with our single-sample classifier by having the classifier report not simply a class label, but also correlation values with all consensus classes. Further studies are required to assess the importance of intra-tumor heterogeneity in prognosis and response to treatment.

The consensus classification suggests possible therapeutic implications. Both the high rate of *FGFR3* mutations and translocations in LumP tumours, and the FGFR3 activation signature associated with these tumours, suggest that tumours that have an FGFR3 activation signature may respond to FGFR inhibitors, irrespective of the *FGFR3*'s mutation or translocation status. Novel fibroblast growth factor receptor inhibitors have been reported to clinically benefit MIBC patients that harbour mutations

or translocations (about 20% of MIBC patients) and/or overexpression (about 40% of MIBC patients) of the tyrosine kinase receptor *FGFR3*[35–38].

Targeting the tumour microenvironment can be an effective option for cancer treatment. Immunotherapy targeting PD1 or PD-L1 immune checkpoints is now included in the standard of care in the US and most of Europe, for patients with locally advanced or metastatic urothelial cancer who relapse after cisplatin-based chemotherapy or are considered cisplatin ineligible, with a 20% objective response rate. A phase 3 clinical trial has demonstrated the efficacy of targeting tumour vasculature in MIBC using an anti-VEGFR2 inhibitor[39]. The different stromal components within consensus classes, identified by transcriptomic signatures, as well as our analysis of the IMvigor210 data, suggest that our consensus classification should be considered for further clinical studies involving immunotherapy or anti-angiogenic therapy.

Similarities between MIBC consensus classes and other cancer molecular subtypes may also be considered for future basket trials. We showed that such similarities are observed, for instance, between Ba/Sq MIBC tumours, Head and Neck Squamous Cell, Lung Squamous Carcinoma, and Cervical Squamous Cell carcinomas, which were placed together in the C27 PanCanAtlas TCGA cluster. LumU tumours and other *ERBB2*-amplified tumours in breast and stomach cancers were also grouped together in the C2 TCGA PanCanAtlas cluster. More generally, Damrauer *et al* have shown that bladder cancer and breast cancer luminal tumours share molecular similarities[7]. Indeed, in both cancers the luminal subtypes rely on GATA3 and FOXA1, two transcription factors that are necessary for luminal differentiation, and on a nuclear receptor: the estrogen receptor in breast cancer, and PPARG[20] in MIBC. Intriguingly, in both cancers there is evidence that the nuclear receptor is involved in differentiation,

while also having protumorigenic effects. Such comparisons across tumour types may help transfer treatment information from tumours bearing similar characteristics into bladder cancer.

We emphasize that we report biological rather than clinical classes, that can be tested for applications in treatment stratification. We offer the classification and the classifier as resources to apply on a single-patient basis in the work required to refine how such classes can best be used clinically. Notably, we propose the consensus classification as a framework for future studies and clinical trials that are intended to identify better predictive markers. Future sub-stratifications may allow defining a system that is more predictive of response to treatments; in such work, the clinical/strategical issue will be to decide the subtype granularity or resolution[30] that is appropriate for a specific problem.

Online Methods

Subtyping of MIBC samples according to published MIBC molecular classifications.

The six classification systems were mainly built on transcriptomic data, as follows: Mo *et al*[13] (Baylor/Tumour differentiation) developed a 18-gene tumour differentiation signature that molecularly define urothelial differentiation, and used this signature to stratify MIBC patients into two groups, namely basal and differentiated.; Damrauer *et al*[7] (UNC) performed consensus clustering on four aggregated datasets totalling 236 tumours and identified two major clusters, termed luminal-like and basal-like, based on similarities with breast cancer subtypes; Rebouissou *et al*[8] (CIT-Curie) performed a hierarchical clustering in seven independent datasets including 370 tumours and identified seven meta-clusters (MC) by measuring similarities between

the clusters obtained in each dataset; Choi *et al*[9] (MDA) identified three subtypes through hierarchical clustering of a 73 tumours dataset, that were named basal, luminal and p53-like relatively to the transcriptomic markers and signatures expressed within each cluster; Marzouka *et al*[10] (Lund) generated gene expression data from 307 tumours and subdivided the cohort into six groups by hierarchical clustering, followed by further sub-stratification into ten levels using immunohistochemistry (IHC); Robertson *et al*[2] (TCGA) performed a consensus hierarchical clustering of RNA-seq profiles from 412 tumours compiled by TCGA and identified five expression subtypes. Transcriptomic classifiers for Baylor, UNC, MDA, CIT-Curie, Lund, and TCGA classification systems were provided and/or validated by the respective teams. All classifiers were merged into an R package (https://github.com/cit-bioinfo/BLCAsubtyping).

We used these classifiers on 18 MIBC mRNA datasets (*N* = 1750 samples) profiled on ten different gene expression platforms (Supplementary Table 1), and assigned each sample to a subtype in each of the six classification systems. 16 datasets were retrieved from public repositories, and two unpublished datasets were shared by L.D. The normalisation method applied on each dataset is detailed in Supplementary Table 1. The six classifiers were applied on each dataset independently.

Network construction and identification of consensus classes

Classification results from the six classifiers were merged for all 18 datasets, and transformed into a binary matrix D of 1750 samples (rows) x 29 classes (columns), where D(s, c) is set to 1 if sample s belongs to class c and 0 otherwise, each row associated with a given sample contains exactly six 1's, reflecting the six class labels

predicted by the six classification systems. For each pair of classes, Cohen's Kappa scores were computed, evaluating the agreement between the two corresponding binary columns of the matrix, i.e. between the 1750 pairs of belong/don't belong class assignments. We could then build a weighted network, with 29 nodes encoding the input molecular subtype, and weighted edges encoding Cohen's Kappa scores. If two subtypes were related by a Cohen's Kappa score < 0, no edge was built between them, for negative values mean a complete absence of agreement between the two subtypes assignments. To quantify the statistical significance of the remaining edges, we performed hypergeometric tests for overrepresentation of samples classified to one subtype in another. The resulting p-values were adjusted for multiple hypothesis testing using the Benjamini–Hochberg (BH) method, and only edges corresponding to $P < 0.001$ were kept to build the network represented in Figure 1a. Consensus classes were then identified by partitioning this network into clusters using bootstrap iterations and MCL[40] (Markov cluster algorithm) as described in Guinney *et al*[32].

Clustering results were evaluated for MCL inflation factor I ranging from 3 to 15, with 0.3 increments, and using 500 resampling iterations. Each iteration consisted of randomly selecting 80% of samples before constructing the network and running MCL for a given inflation factor. For each inflation factor, we calculated a consensus clustering matrix, defined by the frequency that each pair of input subtypes is partitioned into the same cluster over all iterations. To evaluate clustering robustness, we computed a mean weighted silhouette width (MWSW) for each clustering result as previously described[32]. The weighted silhouette width extends the silhouette width by giving more weight to subtypes that are more stable within their assigned clusters. Here, we computed a stability score for each input subtype, defined as the average

frequency over all iterations that its within-cluster associations with other subtypes is the same as predicted by MCL on the network generated with all samples. We then used these stability scores as subtypes' weights to compute weighted silhouette width, and considered the mean over all subtypes as a measure of clustering robustness (Supplementary Figure 1a). Clustering generated four- to six-cluster solutions, all of them yielding a mean silhouette width > 0.95 for at least one inflation factor value (Supplementary Figure 1b). The K=4 solution was very robust (MWSW = 0.99) but poorly informative, revealing one cluster of basal subtypes, one cluster of luminal subtypes, one cluster of infiltrated classes, and one cluster of neuroendocrine-associated subtypes. The K=5 solution isolated an additional cluster containing only two subtypes (CIT MC7 and TCGA Luminal subtypes), which was not enough to clearly define a consensus class. The K=6 solution generated robust clusters (MWSW = 0.95), all containing a minimum of 3 subtypes). Heatmaps of consensus matrices for the three solutions illustrate the robustness of the clusters (Supplementary Figure 1b).

Identification of a core set of consensus samples

For each MIBC sample we performed a hypergeometric test for overrepresentation of the sample's assigned input subtypes in the set of input subtypes associated with each consensus class. A sample was assigned to a consensus class if the corresponding overrepresentation test was significant ($P < 0.001$). Using this approach, a core set of 1084 samples were identified to be highly representative of one of the 6 consensus classes and were labelled as consensus samples. We used these consensus samples to build and validate a single-sample mRNA classifier for the consensus classes, then used this classifier to assign consensus labels to all 1750 MIBC samples.

Single-sample transcriptomic classifier construction

We performed feature selection using a training core set of consensus samples from Sjödahl2017 (n=129) and TCGA (n=274) mRNA datasets, both of these sample sets including at least three consensus samples for each consensus class. In each dataset we performed LIMMA moderated t-tests (limma_3.39.1 R package) for each consensus class relative to the others and computed the AUC associated with each gene for the prediction of each class. We summarized the results for each gene common to both datasets (n=17381), using Stouffer's method to aggregate p-values, and computing a mean fold-change for each class comparison. For each class, we selected the genes with Stouffer $P<0.05$ and AUC>0.6 in at least one of the two datasets, and ordered them according to their mean fold-change. We used these ordered gene lists to generate several lists of varying sizes, by selecting the N top upregulated genes and the N top down-regulated genes in each consensus class, with N varying from 10 to 125. A Pearson nearest-centroid classifier was built on the 129 Sjödahl2017 core samples for each of these gene lists, and its mean balanced accuracy was tested on the independent 681 consensus samples that had not been used for feature selection. The gene list that optimized mean balanced accuracy (97.23%, Supplementary Figure 5) comprised 857 unique genes, and was used to build the final classifier. Six centroids corresponding to the six consensus classes (i.e. the mean mRNA profile of the 857 genes over each consensus class) were computed on the 129 consensus samples from Sjödahl2017 dataset. To classify the 1750 samples into one of the consensus classes, a Pearson correlation was computed between each sample and each centroid. Each sample was then assigned the consensus class whose centroid was the most correlated with the sample profile. If the maximal

correlation for a given sample was less than 0.2, no consensus class label was assigned. This Pearson-based approach does not require to add a pre-processing step to the usual batch normalization of gene expression data, as long as the data are log-transformed, and can therefore be used in a single-sample setting. As shown in Supplementary Figure 5c, the classifier accuracy was similar when using Affymetrix, Illumina, or RNA-seq data. The classifier is publicly available as an R package at https://github.com/cit-bioinfo/consensusMIBC.

Comparison with TCGA pan-cancer classifications.

The consensus bladder cancer classification schemed was compared to the TCGA's PanCancerAtlas pan-cancer subtypes[18]. We visualized the overlap of classification schemes by calculating the percentage within each MIBC consensus class across the TCGA PanCancer Atlas iCluster classification. We then normalized each row (consensus class) by setting the sum of squares equal to 1. We clustered these data using 1-pearson correlation and used a heatmap for visualisation. To evaluate the significance of the enrichment of consensus classes with certain pan-cancer classifications, we calculated the Chi-Square or Fisher's Exact test p-value from a 2x2 contingency table for the given two classifications of interest. To account for multiple testing, we calculated the Bonferroni p-value threshold for 441 pairwise comparisons to be $P<0.00011$.

Extraction of bladder cancer gene signatures from Biton *et al*

In their study, Biton *et al* identified and characterized several major bladder cancer signals by an independent component analysis of bladder cancer transcriptome data[20]. We used ten of these independent components to extract gene sets associated

with the Ta pathway (CIT-13), basal differentiation (CIT-6), cell cycle (CIT-7), urothelial differentiation (CIT-9), smooth muscle (CIT-3), lymphocytes B&T (CIT-8), myofibroblasts (CIT-12), interferon response (CIT-5), neuroendocrine differentiation (CIT-18), and mitochondria (CIT-4). We retrieved the sample contribution vectors associated with each of these components and correlated these values to each gene of the CIT mRNA dataset. Genes that had a Pearson correlation greater than 0.6 (or less than -0.6, depending on the direction of the component association with the biological signal) were selected as representative gene sets for the biological signals associated to the component. The resulting gene sets are given in Supplementary Table 2. For each mRNA dataset included in the study, the R package GSVA[41] (1.30.0) was used to compute single-sample GSEA (Gene Set Enrichment Analysis) scores for the 10 gene sets obtained. The scores were scaled and centered by gene in order to aggregate all datasets. Mean scores were then computed for each consensus class.

Computation of regulon activity scores for 23 regulators

A transcriptional regulatory network for 23 regulators reported as associated with bladder cancer was reconstructed from the TCGA (n=404) MIBC RNA-seq data[2] using the RTN R package (2.6.0). This regulatory network reconstruction was provided as an RTN TNI-class object, and used to calculate regulon activity scores for 18 cohorts, individually. In each sample in each cohort, for each regulon we used RTN's tni.gsea2 function to calculate two-tailed GSEA tests[23]. This generated regulon activity profiles (RAPs) for each cohort; such a profile shows regulon activities of samples, relative to other samples in the same cohort. Regulons were also assigned discrete status as 'activated', 'neutral' and 'inactivated' in each sample based on their activity.

Statistical analyses

We measured association between consensus classes and categorical variables by Fisher's exact or Chi-square tests. We evaluated differences of continuous variables distributions between consensus classes by Kruskal-Wallis tests, ANOVA or LIMMA moderated t-tests (limma_3.39.1 R package).

We built multivariate Cox models integrating consensus classes and clinical risk factors, stratified on cohort of patients (separate baseline hazard functions were fit for each strata). We used Wald tests to assess survival differences associated with different levels of a given factor included in the Cox models. For each factor level, we computed Hazard Ratios (HR) and 95% Confidence Intervals (CI). We constructed Kaplan-Meier curves to visualize overall survival stratified by consensus class and used log-rank tests to compare the survival of corresponding patient groups.

All statistical and bioinformatics analyses were performed with R software environment (version 3.5.1).

References

1. Knowles, M. A. & Hurst, C. D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat. Rev. Cancer* **15**, 25–41 (2015).
2. Robertson, A. G. *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171**, 540-556.e25 (2017).
3. Blaveri, E. *et al.* Bladder cancer outcome and subtype classification by gene expression. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **11**, 4044–4055 (2005).
4. Lindgren, D. *et al.* Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. *Cancer Res.* **70**, 3463–3472 (2010).
5. Sjödahl, G. *et al.* A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 3377–3386 (2012).
6. Volkmer, J.-P. *et al.* Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2078–2083 (2012).
7. Damrauer, J. S. *et al.* Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3110–3115 (2014).
8. Rebouissou, S. *et al.* EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci. Transl. Med.* **6**, 244ra91 (2014).
9. Choi, W. *et al.* Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* **25**, 152–165 (2014).
10. Marzouka, N. *et al.* A validation and extended description of the Lund taxonomy for urothelial carcinoma using the TCGA cohort. *Sci. Rep.* **8**, 3737 (2018).

11. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
12. Sjödahl, G., Eriksson, P., Liedberg, F. & Höglund, M. Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. *J. Pathol.* **242**, 113–125 (2017).
13. Mo, Q. *et al.* Prognostic Power of a Tumor Differentiation Gene Signature for Bladder Urothelial Carcinomas. *J. Natl. Cancer Inst.* (2018). doi:10.1093/jnci/djx243
14. Hedegaard, J. *et al.* Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* **30**, 27–42 (2016).
15. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* **387**, 1909–1920 (2016).
16. Seiler, R. *et al.* Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *Eur. Urol.* **72**, 544–554 (2017).
17. Mariathasan, S. *et al.* TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
18. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
19. Eriksson, P. *et al.* Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems. *BMC Med. Genomics* **8**, 25 (2015).

20. Biton, A. *et al.* Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. *Cell Rep.* **9**, 1235–1245 (2014).
21. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
22. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
23. Castro, M. A. A. *et al.* Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* **48**, 12–21 (2016).
24. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
25. Böck, M. *et al.* Identification of ELF3 as an early transcriptional regulator of human urothelium. *Dev. Biol.* **386**, 321–330 (2014).
26. Choi, W. *et al.* Genetic Alterations in the Molecular Subtypes of Bladder Cancer: Illustration in the Cancer Genome Atlas Dataset. *Eur. Urol.* **72**, 354–365 (2017).
27. Ayers, M. *et al.* IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
28. Pawlik, T. M. & Keyomarsi, K. Role of cell cycle in mediating sensitivity to radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **59**, 928–942 (2004).
29. Horsman, M. R. & Overgaard, J. The impact of hypoxia and its modification of the outcome of radiotherapy. *J. Radiat. Res. (Tokyo)* **57**, i90–i98 (2016).

30. Aine, M., Eriksson, P., Liedberg, F., Höglund, M. & Sjödahl, G. On Molecular Classification of Bladder Cancer: Out of One, Many. *Eur. Urol.* **68**, 921–923 (2015).
31. Fishwick, C. *et al.* Heterarchy of transcription factors driving basal and luminal cell phenotypes in human urothelium. *Cell Death Differ.* **24**, 809–818 (2017).
32. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **advance online publication**, (2015).
33. Warrick, J. I. *et al.* Intratumoral Heterogeneity of Bladder Cancer by Molecular Subtypes and Histologic Variants. *Eur. Urol.* **0**, (2018).
34. Thomsen, M. B. H. *et al.* Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci. Rep.* **7**, 11702 (2017).
35. Tabernero, J. *et al.* Phase I Dose-Escalation Study of JNJ-42756493, an Oral Pan–Fibroblast Growth Factor Receptor Inhibitor, in Patients With Advanced Solid Tumors. *J. Clin. Oncol.* **33**, 3401–3408 (2015).
36. Nogova, L. *et al.* Evaluation of BGJ398, a Fibroblast Growth Factor Receptor 1-3 Kinase Inhibitor, in Patients With Advanced Solid Tumors Harboring Genetic Alterations in Fibroblast Growth Factor Receptors: Results of a Global Phase I, Dose-Escalation and Dose-Expansion Study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **35**, 157–165 (2017).
37. Schuler, M. *et al.* 859PAnti-tumor activity of the pan-FGFR inhibitor rogaratinib in patients with advanced urothelial carcinomas selected based on tumor FGFR mRNA expression levels. *Ann. Oncol.* **28**, (2017).
38. Pal, S. K. *et al.* Efficacy of BGJ398, a fibroblast growth factor receptor 1-3 inhibitor, in patients with previously treated advanced urothelial carcinoma with

FGFR3 alterations. *Cancer Discov.* CD-18-0229 (2018). doi:10.1158/2159-8290.CD-18-0229

39. Petrylak, D. P. *et al.* Ramucirumab plus docetaxel versus placebo plus docetaxel in patients with locally advanced or metastatic urothelial carcinoma after platinum-based therapy (RANGE): a randomised, double-blind, phase 3 trial. *The Lancet* **390**, 2266–2277 (2017).
40. Van Dongen, S. Graph Clustering Via a Discrete Uncoupling Process. *SIAM J. Matrix Anal. Appl.* **30**, 121–141 (2008).
41. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).

Figure legends

Figure 1: The six consensus classes and their relation to input molecular subtypes. (**a**) Clustered network by MCL clustering. The 6-consensus classes solution obtained with MCL clustering on the Cohen's Kappa-weighted network is represented by the 6 cliques surrounded by black dotted rectangles. The circles inside each clique symbolize the input subtypes associated with each consensus class and are coloured according to their matching classification system. Circle size is proportional to the number of samples assigned to the subtype. Edge width between subtypes is proportional to the Cohen's Kappa score, which assess the level of agreement between two classification schemes. (**b**) Input subtypes repartitioned among each consensus class. Consensus classes were predicted on 1750 MIBC samples using the single-sample classifier described in Methods. Here, the samples are grouped by their predicted consensus class label: LumP, LumN, LumU, Stroma-rich, Ba/Sq and Neuroendocrine (NE)-like. For each consensus class, a barplot shows the proportion of samples assigned in each input subtype of each input classification system.

Figure 2: Characterization of tumour and stroma signals using published mRNA signatures. (**a**) The 1750 mRNA expression profiles were used to compute (above, Biton) mean enrichment scores for specific gene signatures in each consensus class (based on a single-sample GSEA approach), or (below, Lund) mean expression of gene sets. Bladder cancer gene signatures include those related to the ICA components described in Biton *et* al[20] (see Methods), as well as other bladder cancer-specific signatures retrieved from the literature: urothelial differentiation,

keratinization and late cell-cycle signatures from Eriksson *et al*[42], and an *FGFR3* co-expressed signature from Sjödahl *et al*[5]. (Supplementary Table 2) (**b**) Tumour microenvironment characterization includes (above) an estimate of microenvironment immune and stromal cell subpopulations using MCPcounter[22] and (below) a more global measure of stromal and immune infiltrates by ESTIMATE[21].

Figure 3: Genomic alterations associated with consensus classes. (**a**) We used the available exome data from 388 TCGA samples to study the association between consensus classes and specific gene mutations. The panel displays the 23 genes with significant mutations (MutSig $P < 0.001$) that were either found in at least 10% of all tumours, or significantly overrepresented within one of the consensus classes (Fisher $P < 0.05$ and frequency within a consensus class > 10%). Gene mutations that were significantly enriched in one consensus class are marked by an asterisk. (**b**) Combined genomic alterations associated with seven bladder cancer-associated genes and statistical association with consensus classes. Upper panels: Main alteration types after aggregating CNV profiles from CIT (n=87), Iyer (n=58), Sjödahl (n=29), Stransky (n=22), and TCGA (n=404) data; exome profiles (n=388) and *FGFR3* and *PPARG* fusion data (n=404) from TCGA data; *CDKN2A* and *RB1* MLPA data from CIT(n=86; n=85) and Stransky (n=16; n=13) data; *FGFR3* mutation data from MDA (n=66), CIT (n=87), Iyer (n=39), Sjödahl (n=28), and Stransky (n=35) data; TP53 mutation data from MDA (n=66), CIT (n=87), Iyer (n=39), Sjödahl (n=28), and Stransky (n=19) data; and *RB1* mutation data from MDA (n=66), CIT (n=85), Iyer (n=39) and Stransky (n=13) data. Lower panels: Associations between each consensus class, each type of gene alteration, and the combined alterations were

evaluated by Fisher's exact tests. Consensus classes significantly enriched with alterations of these candidate genes are marked with a black asterisk.

Figure 4: Histopathological associations with consensus classes. (**a**) Histological variant overrepresentation within each consensus class. One-sided Fisher exact tests were performed for each class and histological pattern. Pathological review of histological variants was available for several cohorts: squamous differentiation was evaluated in CIT (n=75), MDA (n=46), Sjödahl2012 (n=23), Sjödahl2017 (n=239) and TCGA (n=406) cohorts; neuroendocrine variants were reviewed in CIT (n=75), MDA (n=46), Sjödahl2017 (n=243), and TCGA (n=406) cohorts; micropapillary variants were reviewed in CIT (n=75), MDA (n=46) and TCGA cohorts (n=118 FFPE tumour slides from TCGA were reviewed by Y.A. and J.F. for this study). Results are displayed on the heatmap as $-\log_{10}$(Fisher's *P*). (**b**) The proportion of samples with carcinoma *in situ* (CIS) associated within each consensus class, for 84 tumours from CIT cohort and 8 tumours from Dyrskjøt cohort. (**c**) The presence/absence of a papillary morphology, for 401 tumours from TCGA cohort and 47 tumours from CIT cohort. (**d**) Smooth muscle infiltration from images for 174 tumour slides from the TCGA cohort: 73 LumP, 18 LumNS, 16 LumU, 20 Stroma-rich and 46 Ba/Sq tumour samples. Each sample was assigned a semi-quantitative score ranging from 0 to 3 (0 = absent, 1 = low, 2 = moderate, 3 = high) to quantify the presence of large smooth muscle bundles. The barplot shows means and standard deviations.

Figure 5: Clinical characteristics and prognostic associations. (**a**) Association of consensus classes with gender (n=1554), clinical stage (n=1641), and age category

(n=1378). (**b**) 5-year overall survival stratified by consensus class. Kaplan-Meier curves were generated from 873 patients with available follow-up data. Patients who were marked as having received neoadjuvant chemotherapy were excluded from the survival analysis. (**c**) The 1750 mRNA expression profiles were used to compute per-class mean expression of gene sets that are clinically relevant for response to therapies (Supplementary Table 6). Gene sets are annotated with a plus (respectively minus) sign if high expression of the genes is associated with response (respectively resistance) to the category of therapies indicated on the left.

Figure 6: Summary of main characteristics of the consensus classes. Top to bottom: Proportion of consensus classes in the n=1750 tumour samples. Consensus classes names. Cellular schematics for tumour cells and their microenvironment (Immune cells, fibroblasts, and smooth muscle cells). Differentiation-based color scale showing the differentiation status associated with consensus classes, including a Luminal-to-basal gradient, and neuroendocrine differentiation status. Table of dominant characteristics: oncogenic mechanisms, mutations, stromal infiltrate, immune infiltrate, histological observations, clinical characteristics, and median overall survival.

Extended data figure legends

Extended data figure 1: Analytical workflow. We used mRNA classifiers provided by 6 teams involved in previously published classification systems to subtype 1750 mRNA profiles from 18 independent MIBC cohorts. A total number of 29 subtypes were considered when summing all classification systems. Using the subtyping results, we could build a 1750 x 29 binary matrix D where a sample s was given a value of 1 if assigned to the subtype m, and 0 otherwise. The matrix D was used to build a network interconnecting the 29 distinct subtypes. Edges between two subtypes were weighted using a Cohen's Kappa metric. We performed MCL clustering[40] on this network with 500 bootstrap iterations for several values of inflation factors and used stability scores as weights to calculate weighted silhouette width for each resulting cluster. We then used the mean weighted silhouette width as a performance measure to select an inflation factor yielding a robust consensus clustering solution. An optimal consensus solution was reached for 6 consensus classes, and this solution also defined a set of 1084 'core' consensus samples with subtype labels that were highly concordant among the consensus classes ($P < 0.001$, hypergeometric test). We used these 1084 core samples to build a nearest-centroid, single-sample classifier based on Pearson's correlation coefficient, then used the resulting classifier to predict consensus classes on all 1750 MIBC samples. We further characterized the consensus classes using molecular, histological and clinical data.

Extended data figure 2: **Regulon activity within consensus classes**. We computed regulon activity profiles (RAPs) as described in Robertson *et al*[2], for 23

bladder cancer regulators. (**a**) Heatmap of RAPs at the sample level. RAPs were computed on each of the 18 datasets independently and pooled for the heatmap visualization. (**b**) Summary showing the mean RAP for each consensus class. (**c**) Association of a regulon's active or inactive status with each class, indicated by –$\log_{10}$(Fisher *P* value). Fisher exact tests were done using RAPs that had been discretized by status (1 for active regulon status, 0 for neutral status, -1 for inactive regulon status).

Extended data figure 3: Distributions of SCNA, and total somatic and APOBEC mutation loads across consensus classes (**a**) Distribution of Somatic Copy Number Alteration (SCNA) counts across consensus classes. SCNA counts are defined as the number of genes with copy number changes, as estimated by GISTIC2[24] over 600 MIBC CNV profiles from datasets from CIT (n=87), Iyer (n=58), Sjödahl2012 (n =29), Stransky (n=22) and TCGA (n=404). (**b**) Distribution of nonsynonymous somatic mutation events across consensus classes. (**c**) Enrichment of APOBEC-induced mutation within consensus classes. The minimum estimate of the number of APOBEC-induced mutations was computed for 388 samples of TCGA MIBC cohort and discretized into categorical values : "No" : estimate = 0; "Low": estimate ≤ median of non-zero values (median was 61.5); "High": estimate > median of non-zero values.

Extended data figure 4: Response to neoadjuvant chemotherapy and PD-L1 blockade.

To further explore the association of the consensus classification with therapeutic response, we analysed overall survival and response data from patients who had

received neoadjuvant chemotherapy[9,16] (NAC) and patients treated with the anti PD-L1 atezolizumab[17] (IMvigor210). The pre-treatment tumour samples from these patients were classified according to the consensus molecular classification. To better evaluate the effect of NAC on overall survival we selected a set of NAC-free patients and compared the class-associated overall survival of these patients with survival of patients receiving NAC. (**a**) Overall survival and response data to neoadjuvant chemotherapy (NAC). For the analysis of overall survival, NAC-free patients were selected from MDA (n=46), Sjödahl (n=51) and TCGA (n=394) cohorts, patients treated with NAC from Seiler (n=273), MDA MVAC (n=22, GSE70691), and MDA DDMVAC (n=38, GSE69795) cohorts.

Pathological response to NAC was obtained from MDA MVAC (n=23), MDA DDMVAC (n=34) and Seiler (n=43) cohorts. **(b)** Overall survival and response to PD-L1 blockade (atezolizumab), from IMvigor210 trial (Mariathasan et al). Consensus classes were predicted for all MIBC samples included in IMvigor210 dataset using the single-sample classifier. Consensus classes associated (Fisher $P<0.05$) with positive response to atezolizumab, i.e. complete (CR) or partial responders (PR), are indicated by a black asterisk.

Bladder Cancer Molecular Taxonomy Group and affiliations

Mattias Aine, Division of Molecular Hematology, Department of Laboratory Medicine, Faculty of Medicine, Lund University, Lund, Sweden

Hikmat Al-Ahmadie, Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Yves Allory, Department of Pathology, Institut Curie Hospital Group, Paris, France

Joaquim Bellmunt, Bladder Cancer Center, Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, MA, 02215, USA

Isabelle Bernard-Pierrot, Oncologie Moléculaire, CNRS UMR 144, Institut Curie, Paris, France

Peter C. Black, Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Mauro A. A. Castro, Bioinformatics and Systems Biology Laboratory, Federal University of Paraná, Polytechnic Center, Curitiba, Brazil

Keith S. Chan, Molecular & Cellular Biology/Scott Department of Urology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Woonyoung Choi, Johns Hopkins Greenberg Bladder Cancer Institute and Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA

Bogdan Czerniak, Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

Colin P. Dinney, Department of Urology and Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Lars Dyrskjøt, Department of Molecular Medicine, Aarhus University Hospital, Aarhus 8200, Denmark

Pontus Eriksson, Division of Oncology and Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden

Jacqueline Fontugne, Department of Pathology, Institut Curie Hospital Group, Paris, France

Ewan A. Gibb, GenomeDx Biosciences Inc., Vancouver, BC Canada

Clarice S. Groeneveld, Bioinformatics and Systems Biology Laboratory, Federal University of Paraná, Polytechnic Center, Curitiba, Brazil

Arndt Hartmann, Institute of Pathology, University Erlangen-Nürnberg, Krankenhausstr 8-10, Erlangen, Germany

Katherine A. Hoadley, Department of Genetics, Department of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Mattias Höglund, Division of Oncology and Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden

Aurélie Kamoun, Cartes d'Identité des Tumeurs Program, Ligue Nationale Contre le Cancer, 75013 Paris, France

Jaegil Kim, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

William Y. Kim, Department of Genetics, Department of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

David Kwiatkowski, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

Thierry Lebret, Department of Urology, University of Versailles-Saint-Quentin-en-Yvelines, Foch Hospital, Suresnes, France.

Seth P. Lerner, Scott Department of Urology, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA

Fredrik Liedberg, Department of Translational Medicine, Lund University, Skåne University Hospital, Malmö, Sweden

Núria Malats, Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

David J. McConkey, Johns Hopkins Greenberg Bladder Cancer Institute and Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA

Qianxing Mo, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

Thomas Powles, Barts Cancer Institute ECMC, Barts Health and the Royal Free NHS Trust, Queen Mary University of London, London, UK

François Radvanyi, Oncologie Moléculaire, CNRS UMR 144, Institut Curie, Paris, France

Francisco X. Real, Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

Aurélien de Reyniès, Cartes d'Identité des Tumeurs Program, Ligue Nationale Contre le Cancer, 75013 Paris, France

A. Gordon Robertson, Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, Canada

Arlene Siefker-Radtke, Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Nanor Sirab, Department of Pathology, Institut Curie Hospital Group, Paris, France

Roland Seiler, Department of Urology, Bern University Hospital, Switzerland

Gottfrid Sjödahl, Division of Urological Research, Department of Translational Medicine, Lund University, Skåne University Hospital Malmö, Sweden

Ann Taber, Department of Molecular Medicine, Aarhus University Hospital, Aarhus 8200, Denmark

John Weinstein, Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA;

Alexandre Zlotta, Department of Surgery, Division of Urology, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, ON, Canada