Three transcriptional clusters of bladder carcinoma: a systematic approach to molecular classification of urothelial carcinoma

Figures, Tables, Methods

Department of Urology, Medial University of Innsbruck

2026-02-05

# Methods

Details on software, data sources, data processing and management, statistics, and bioinformatics are provided in **Supplementary Methods**, and **Supplementary Tables S1** and **S2**.

## Software and data sources

Data management, statistical and bioinformatic analyses were done with R version 4.2.3 (R Foundation for Statistical Computing).

The [cBioportal](https://www.cbioportal.org/), [Gene Expression Omnibus (GEO)](https://www.ncbi.nlm.nih.gov/geo/) and [Array Express](https://www.ebi.ac.uk/biostudies/arrayexpress) repositories were screened for transcriptomic and proteomic studies on urothelial cancers. Herein, 19 bulk cancer transcriptome cohorts with at least 80 specimens and 18000 quantified genes (TCGA BLCA (1,2), IMvigor (3,4), GSE13507 (5), GSE32548 (6), GSE48075 (7), GSE48276 (7), GSE83586 (8), GSE86411 (9), GSE87304 (10), GSE120736 (11), GSE124305 (12), GSE128192 (13), GSE128701 (14), GSE128959 (15), GSE198269 (16), GSE203149 (17), E-MTAB-4321 (18), Groeneveld 2024 (19), BCAN (20)) and three bulk cancer proteomes with 1600 to 9265 quantified proteins (Groeneveld 2024 (19), Dressler 2024 (21), Stroggilos 2020 (22)) were analyzed (**Table 1** and **2**, **Supplementary Tables S3** and **S4**). Additionally, gene expression and CRISPR-Cas9 knockout screening data for 33 urothelial cancer cell lines with reliable assignment to bladder cancer clusters were imported from the [DepMap portal](https://depmap.org/portal/) (23).

Normalized gene and protein expression, genetic alteration data, and clinical information provided by the study authors were analyzed. Gene and protein expression values were analyzed in a -transformed form. Missing gene expression Z-scores (BCAN cohort) and protein expression values (Groeneveld 2024 and Dressler 2024) for variables with less than 50% missing entries were imputed with k-nearest neighbor regressors (R package *impute*) (24). For some analyses (e.g. clustering, network analyses, prediction of drug resistance), bulk cancer transcriptome and proteome data were processed with ComBat to adjust for the between cohort variability (25). Genetic alteration data were analyzed in a binarized form (0: wild-type or two copies of a gene; 1: one or more mutations, at least one copy deletion, at least one additional copy).

Single sample gene set enrichment scores of metagenes, signatures of Reactome Pathways (extracted from the [MSig Database](https://www.gsea-msigdb.org/gsea/msigdb/index.jsp)) and cell proliferation (26–32) were computed with the GSVA algorithm (33). Levels of infiltrating immune and stromal cells in bulk cancer transcriptome samples were obtained by immunedeconvolution with xCell (34) and QuanTIseq algorithms (35). Predictions of drug resistance with transcriptome data were done by RIDGE linear models trained with in vitro drug screening data of GDSC, CTRP2, and PRISM experiments (package [*htGLMNET*](https://github.com/PiotrTymoszuk/htGLMNET)) (36–39). Linear modeling (LM) scores of activity of collecTRI regulons and PROGENy signaling pathways were predicted for bulk cancer transcriptome samples with linear modeling tools of *decoupleR* (40–42).

Bulk cancer transcriptome specimens were assigned to UROMOL classes of NMIBC (18) and consensus classes of MIBC (43) by nearest centroid classifiers from R packages [*classifyNMIBC*](https://github.com/sialindskrog/classifyNMIBC) and [*consensusMIBC*](https://github.com/cit-bioinfo/consensusMIBC).

## Harmonization of clinical and pathological information

Information on muscle invasiveness, if not provided in clinical meta-data accompanying the molecular data of bulk cancer transcriptome and proteome cohorts or available in the genuine publications, was derived from pathological staging with Ta, T1, Tis and CIS stage tumors classified as NMIBC and T2 - T4 specimens as MIBC.

To harmonize highly variable and partly incomplete classification of cancer tissue (primary, recurrent, metastatic) and anatomical location between the bulk cancer cohorts, tissue source was subsumed under “bladder”, “urinary organ, non-bladder”, and “non-urinary organ, distant metastasis” based on clinical/pathological metadata and genuine publications.

Survival outcomes were harmonized as follows:

* death and overall survival: any-cause death during the study follow-up
* tumor death and tumor-specific survival: death due to urothelial cancer or its metastases during the study follow-up
* relapse and relapse-free survival: any recurrence or progression of urothelial cancer after initial treatment during the study follow-up
* MIBC progression: progression of primary NMIBC to MIBC during the study follow-up. Please note that progression of MIBC has not been analyzed.

## Descriptive statistics, statistical hypothesis testing

Numeric variables are referenced as medians with interquartile ranges, ranges, and numbers of complete cases. Qualitative and ordinal features are provided as counts of observations and percentages of the categories, and numbers of complete cases. For survival, Kaplan-Meier estimates of survival times for the 25%, 50%, and 75% survival quantiles with 95% confidence intervals (95% CI) are provided. Estimates and meta-estimates obtained in statistically hypothesis testing and modeling are presented as estimates of expected values, means, or coefficients with 95% CI. Normality and equality of variance were investigated by Shapiro-Wilk and Levene tests

Statistical significance of differences in numeric variables was determined by T, Mann-Whitney, one-way ANOVA, or Kruskal-Wallis tests with Cohen’s d, biserial r, or effect size statistics. Statistical significance of differences in distributions of categorical variables was assessed by test with Cramer’s V effect size metric. Differences in survival were evaluated by Peto-Peto tests, and modeled by univariable Cox proportional hazard models and multivariable Cox proportional hazard models adjusted for clinical predictors (packages *survival* and *survminer*) (44,45).

P values were adjusted for multiple testing with the false discovery rate (FDR) (46) separately for each analytic task and cohort. If not indicated otherwise, effects with FDR-corrected p < 0.05 were considered significant. As additional control for false positive effects, significant effects shared by multiple cohorts are discussed in the reports and subjected to meta-analyses with DerSimonian-Lair algorithm (47) (differentially regulated genes, gene signatures, drug resistance estimates, regulon and signaling activity scores, metabolic reaction activity estimates: 10 cohorts; proteins and protein signatures: 2 cohorts).

## Definition and technical validation of bladder cancer clusters, assignment of transcriptomes and proteomes to bladder cancer clusters

Bladder cancer clusters were defined in the TCGA BLCA cohort with a two-step procedure. In the first step, metagenes, i.e. micro-clusters of tightly co-regulated genes, were defined by a toroidal 16 16 unit self-organizing map (SOM) with squared Euclidean distance between the genes and nodes (R packages *cohonen* and [*clustTools*](https://github.com/PiotrTymoszuk/clustTools)) (48,49). In the second step, the TCGA BLCA cancer specimens were clustered by their ssGSEA metagene scores with the hard-threshold regularized KMEANS algorithm (HTKmeans, package [*clustTools*](https://github.com/PiotrTymoszuk/clustTools)) (50), yielding three bladder cancer clusters #1, #2, and #3. Potential over-fitting by the SOM and HTKmeans algorithm were addressed by, respectively, holdout validation and cross-validation. Hyper-parameters of the SOM (topology, distance measure) were chosen by comparing metrics of explained clustering variance, topology error, and neighborhood preservation (48,51,52) in the training (75% observations) and test (25% observations not used in SOM development) subsets of the TCGA BLCA cohort (**Supplementary Figure S1**). Hyper-parameters of the HTKmeans algorithm (number of clusters, shrinkage parameter ) were selected by assessing cluster separation and potential misclassification (silhouette method), explained clustering variance, and neighborhood preservation (50,52,53) in the entire TCGA BLCA data set and five-fold cross-validation (54) (**Supplementary Figure S2**).

Bulk cancer transcriptome specimens in validation cohorts (IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, and Groeneveld 2024) were assigned to the clusters by an inverse distance-weighted nearest neighbor classifier (55). Quality of the clustering structure in the TCGA BLCA and validation collectives was assessed by numeric statistics (52,53), visualizations of UMAP embeddings (uniform manifold approximation and projection) (56), comparison of ssGSEA scores of the cluster-defining metagenes, and analysis of squared Euclidean distances between the clusters (**Supplementary Tables S5** - **S10**).

Bulk cancer transcriptome samples in the BCAN cohort, bulk cancer proteome samples, and DepMap urothelial cancer cell lines were assigned to bladder cancer clusters by multinomial Elastic Net classifiers (**Supplementary Tables S11** - **S14**).

Maker transcripts and proteins in bladder cancer clusters of bulk cancer transcriptomes, bulk cancer proteomes, and DepMap urothelial cancer cell lines were identified by FDR-adjusted one-way ANOVA and receiver-operating characteristic (ROC) with area under the ROC curve AUC 0.714 as expected for substantial markers according to Rice at al. (57).

## Genetic alterations, differential expression of genes, proteins and signatures, regulon and signaling activity, predicted drug resistance

Enrichment of genetic alterations in bladder cancer clusters of the TCGA BLCA, IMvigor, and BCAN cohorts as compared with expected frequency was assessed by weighted permutation testing with [*perich*](https://github.com/PiotrTymoszuk/perich) package as described in (58) and **Supplementary Methods**.

Differences in expression levels of genes and proteins, ssGSEA scores of gene and protein signature, LM activity scores of regulons and signaling pathways (bulk cancer transcriptomes), and predicted drug resistance metrics in bladder clusters were evaluated with a combined procedure. First, significance of differences between the clusters was assessed by one-way ANOVA with effect size statistic. Next, differences between the clusters and the cohort means were investigated by post-hoc one-sample T tests. Significantly differentially regulated features were identified by pFDR(ANOVA) < 0.05, 0.06, and pFDR(T test) < 0.05 (R packages [*fastTest*](https://github.com/PiotrTymoszuk/fastTest) and [*microViz*](https://github.com/PiotrTymoszuk/microViz)).

## Network analyses

Correlation and co-expression networks were build with matrices of Spearman’s pairwise correlation coefficients (59,60). Differentially regulated genes, proteins, gene/protein signatures, or regulons served as network vertices. The network edges were defined by pairwise correlations with 0.5 or 0.3, and were weighted by values. Similarity networks of differentially regulated Reactome Pathway gene signatures were constructed with matrices of Jaccard’s similarity coefficients J which measure overlaps between member genes/proteins of signature pairs. The network edges were defined by J 0.3 and were weighted by J values (R packages *igraph* and [*graphExtra*](https://github.com/PiotrTymoszuk/graphExtra)) (60). Communities of the networks were identified with the Leiden algorithm (61) and named after their characteristic biological features.

## Modeling of metabolism

Activity of metabolic reactions of the RECON2 knowledge model (62) was assessed by evaluation of the model’s gene/protein - reaction association rules in a Monte Carlo simulation (63,64). In this simulation, estimates of differential expression for all available genes or proteins were used. Enrichment of RECON metabolic subsystems with significantly activated and inhibited reactions, was investigated by comparing the subsystem’s reaction frequency in the activated or inhibited reaction set with 10000 random draws from the entire reaction pool. Metabolic reaction modeling and enrichment analyses were done with *BiGGR* and [*biggrExtra*](https://github.com/PiotrTymoszuk/biggrExtra).

## CRISPR-Cas9 gene effects

Effects of gene KO were measured by Chronos scores. The Chronos scores were transformed by multiplication by -1, hence, positive values correspond to growth inhibition attributed to the gene KO. Biologically relevant KO effects were assumed for Chronos scores > 0.5 (65,66). Biological effects of gene KO in the clusters and differences in gene KO effects between the clusters were assessed by bootstrap tests (R package *boot*) (67).

# Data and code availability

The analyzed data sets are freely available. The analysis pipelines are available as GitHub repositories [BLCA-cluster-paper](https://github.com/PiotrTymoszuk/BLCA-cluster-paper) and [BLCA\_subset\_schemes](https://github.com/PiotrTymoszuk/BLCA_subset_schemes).

An interactive HTML presentation with Supplementary Material is available from a [dedicated GitHub repository](https://github.com/PiotrTymoszuk/BLCA_cluster_supplements).

# Tables

Table 1: Demographic and clinical characteristic of urothelial cancer patients analyzed in the current study. Qualitative variables are presented as percentages and counts of the categories within the complete observation set. Qualitative variables are shown as medians with interquartile ranges and ranges. Statistical significance of differences between non-muscle invasive (NMIBC) and muscle invasive cancers (MIBC) was assessed by chi-square test with Cramer's V effect size statistic (qualitative variables), Mann-Whitney test with r effect size statistic (quantitative variables), or Peto-Peto test (survival). P values were corrected for multiple testing with the false discovery rate method.

| **Variablea** | **All patientsb** | **NMIBCc** | **MIBCd** | **Test typee** | **Significancee** | **Effect sizee** |
| --- | --- | --- | --- | --- | --- | --- |
| Patients, N | 4298 | 1380 | 2817 |  |  |  |
| Sex | female: 11% (468) male: 39% (1670) unknown: 50% (2160) | female: 11% (148) male: 47% (649) unknown: 42% (583) | female: 11% (320) male: 36% (1021) unknown: 52% (1476) | χ² | p = 0.009 | V = 0.062 |
| Age, years | 67 [IQR: 59 - 75] range: 22 - 96 complete: n = 1914 | 68 [IQR: 61 - 76] range: 24 - 96 complete: n = 616 | 67 [IQR: 59 - 74] range: 22 - 94 complete: n = 1298 | Mann-Whitney | p = 0.023 | r = 0.055 |
| Race | Asian: 1.3% (55) Black or African American: 1.1% (47) White: 17% (724) Other: 0.72% (31) unknown: 80% (3441) | not provided | Asian: 2% (55) Black or African American: 1.7% (47) White: 26% (724) Other: 1.1% (31) unknown: 70% (1960) |  |  |  |
| Body mass index, kg/m² | 26 [IQR: 23 - 30] range: 15 - 68 complete: n = 357 | not provided | 26 [IQR: 23 - 30] range: 15 - 68 complete: n = 357 |  |  |  |
| Body mass class | normal: 3.4% (148) overweight: 2.9% (124) obesity: 2% (85) unknown: 92% (3941) | not provided | normal: 5.3% (148) overweight: 4.4% (124) obesity: 3% (85) unknown: 87% (2460) |  |  |  |
| Smoking history | never: 5.4% (230) current or previous: 13% (557) unknown: 82% (3511) | not provided | never: 8.2% (230) current or previous: 20% (557) unknown: 72% (2030) |  |  |  |
| History of BCG treatment | no: 19% (813) yes: 4.2% (180) unknown: 77% (3305) | no: 27% (372) yes: 6.4% (88) unknown: 67% (920) | no: 16% (441) yes: 3.3% (92) unknown: 81% (2284) | χ² | ns (p = 0.56) | V = 0.024 |
| Neoadjuvant systemic chemotherapy | no: 21% (903) yes: 7.8% (335) unknown: 71% (3060) | not provided | no: 32% (903) yes: 12% (335) unknown: 56% (1579) |  |  |  |
| Cystectomy | no: 14% (621) yes: 7.7% (329) unknown: 78% (3348) | no: 37% (510) yes: 2.7% (37) unknown: 60% (833) | no: 3.9% (111) yes: 10% (292) unknown: 86% (2414) | χ² | p < 0.001 | V = 0.68 |
| Adjuvant systemic chemotherapy | no: 11% (477) yes: 19% (827) unknown: 70% (2994) | no: 4.9% (68) yes: 0% (0) unknown: 95% (1312) | no: 15% (409) yes: 29% (827) unknown: 56% (1581) | χ² | p < 0.001 | V = 0.31 |
| Death during follow-up | no: 18% (770) yes: 16% (685) unknown: 66% (2843) | no: 17% (229) yes: 6.8% (94) unknown: 77% (1057) | no: 19% (541) yes: 21% (591) unknown: 60% (1685) | χ² | p < 0.001 | V = 0.19 |
| Follow-up time, days | 660 [IQR: 330 - 1600] range: 0 - 6400 complete: n = 1443 | 1800 [IQR: 800 - 1800] range: 0 - 4600 complete: n = 323 | 570 [IQR: 270 - 1100] range: 0 - 6400 complete: n = 1120 | Peto-Peto | p < 0.001 |  |
| Relapse during follow-up | no: 8.7% (375) yes: 5.6% (241) unknown: 86% (3682) | no: 7% (96) yes: 3.9% (54) unknown: 89% (1230) | no: 9.9% (279) yes: 6.6% (187) unknown: 83% (2351) | χ² | ns (p = 0.54) | V = 0.036 |
| Relapse-free survial, days | 470 [IQR: 210 - 730] range: 0 - 5000 complete: n = 622 | 590 [IQR: 200 - 730] range: 4.1 - 730 complete: n = 154 | 460 [IQR: 220 - 730] range: 0 - 5000 complete: n = 468 | Peto-Peto | ns (p = 0.86) |  |
| MIBC progression during follow-up | no: 13% (561) yes: 1.2% (53) unknown: 86% (3684) | no: 41% (561) yes: 3.8% (53) unknown: 56% (766) | not provided |  |  |  |
| MIBC progression-free survival, days | 820 [IQR: 560 - 1200] range: 0 - 2300 complete: n = 614 | 820 [IQR: 560 - 1200] range: 0 - 2300 complete: n = 614 | not provided |  |  |  |
| aBCG: Bacillus Calmette-Guerin; Relapse: recurrence or progression of the disease after initial treatment; MIBC progression: progression of primary NMIBC to MIBC. | | | | | | |
| bBCAN: n = 174, Dressler 2024: n = 242, E-MTAB-4321: n = 476, Groeneveld 2024: n = 198, GSE120736: n = 145, GSE124305: n = 133, GSE128192: n = 112, GSE128701: n = 136, GSE128959: n = 70, GSE13507: n = 171, GSE198269: n = 394, GSE203149: n = 170, GSE32548: n = 131, GSE48075: n = 142, GSE48276: n = 116, GSE83586: n = 307, GSE86411: n = 132, GSE87304: n = 305, IMvigor: n = 220, Stroggilos 2020: n = 117, TCGA BLCA: n = 407 | | | | | | |
| cDressler 2024: n = 167, E-MTAB-4321: n = 460, Groeneveld 2024: n = 87, GSE120736: n = 84, GSE128959: n = 64, GSE13507: n = 68, GSE32548: n = 92, GSE48075: n = 70, GSE83586: n = 58, GSE86411: n = 132, Stroggilos 2020: n = 98 | | | | | | |
| dBCAN: n = 174, Dressler 2024: n = 75, E-MTAB-4321: n = 16, Groeneveld 2024: n = 82, GSE120736: n = 61, GSE124305: n = 133, GSE128192: n = 112, GSE128701: n = 136, GSE128959: n = 3, GSE13507: n = 41, GSE198269: n = 394, GSE203149: n = 170, GSE32548: n = 38, GSE48075: n = 72, GSE48276: n = 116, GSE83586: n = 243, GSE87304: n = 305, IMvigor: n = 220, Stroggilos 2020: n = 19, TCGA BLCA: n = 407 | | | | | | |
| eComparison of the NMIBC and MIBC groups, the 'unknown' category was excluded. | | | | | | |

Table 2: Pathological characteristic of transcriptomic and proteomic bulk cancer samples analyzed in the current study. Qualitative variables are presented as percentages and counts of the categories within the complete observation set. Qualitative variables are shown as medians with interquartile ranges and ranges. Statistical significance of differences between non-muscle invasive (NMIBC) and muscle invasive cancers (MIBC) was assessed by chi-square test with Cramer's V effect size statistic (qualitative variables), Mann-Whitney test with r effect size statistic (quantitative variables), or Peto-Peto test (survival). P values were corrected for multiple testing with the false discovery rate method.

| **Variable** | **All samplesa** | **NMIBCb** | **MIBCc** | **Test typed** | **Significanced** | **Effect sized** |
| --- | --- | --- | --- | --- | --- | --- |
| Samples, N | 4439 | 1505 | 2872 |  |  |  |
| Cancer tissue type | bladder: 95% (4236) urinary organ, non-bladder: 4.1% (183) non-urinary organ, distant metastasis: 0.45% (20) | bladder: 100% (1505) urinary organ, non-bladder: 0% (0) non-urinary organ, distant metastasis: 0% (0) | bladder: 93% (2669) urinary organ, non-bladder: 6.4% (183) non-urinary organ, distant metastasis: 0.7% (20) | χ² | p < 0.001 | V = 0.16 |
| Invasiveness | non-muscle invasive: 34% (1505) muscle invasive: 65% (2872) unknown: 1.4% (62) | non-muscle invasive: 100% (1505) muscle invasive: 0% (0) unknown: 0% (0) | non-muscle invasive: 0% (0) muscle invasive: 100% (2872) unknown: 0% (0) |  |  |  |
| Pathological tumor stage | T2: 21% (952) T3: 14% (600) T4: 4.9% (219) T1/Ta/Tis: 31% (1373) T2-4: 0.36% (16) T3-4: 0.56% (25) unknown: 28% (1254) | T2: 0% (0) T3: 0% (0) T4: 0% (0) T1/Ta/Tis: 91% (1373) T2-4: 0% (0) T3-4: 0% (0) unknown: 8.8% (132) | T2: 33% (952) T3: 21% (600) T4: 7.6% (219) T1/Ta/Tis: 0% (0) T2-4: 0.56% (16) T3-4: 0.87% (25) unknown: 37% (1060) |  |  |  |
| Pathological node stage | N0: 25% (1113) N1: 4.6% (205) N2: 2.9% (127) N3: 0.38% (17) unknown: 67% (2977) | N0: 26% (387) N1: 0% (0) N2: 0% (0) N3: 0% (0) unknown: 74% (1118) | N0: 25% (726) N1: 7.1% (205) N2: 4.4% (127) N3: 0.59% (17) unknown: 63% (1797) | χ² | p < 0.001 | V = 0.34 |
| Pathological metastasis stage | M0: 15% (683) M1: 1.3% (58) unknown: 83% (3698) | M0: 16% (237) M1: 0% (0) unknown: 84% (1268) | M0: 16% (446) M1: 2% (58) unknown: 82% (2368) | χ² | p < 0.001 | V = 0.2 |
| Pathological grade | G1: 1.4% (61) G2: 3.1% (139) G3: 12% (511) unknown: 84% (3728) | G1: 4.1% (61) G2: 8.4% (126) G3: 14% (213) unknown: 73% (1105) | G1: 0% (0) G2: 0.42% (12) G3: 10% (292) unknown: 89% (2568) | χ² | p < 0.001 | V = 0.47 |
| Histological grade | low grade: 10% (443) high grade: 16% (730) unknown: 74% (3266) | low grade: 27% (402) high grade: 15% (232) unknown: 58% (871) | low grade: 1.4% (41) high grade: 17% (498) unknown: 81% (2333) | χ² | p < 0.001 | V = 0.57 |
| aBCAN: n = 174, Dressler 2024: n = 242, E-MTAB-4321: n = 476, Groeneveld 2024: n = 198, GSE120736: n = 145, GSE124305: n = 133, GSE128192: n = 112, GSE128701: n = 136, GSE128959: n = 192, GSE13507: n = 188, GSE198269: n = 394, GSE203149: n = 171, GSE32548: n = 131, GSE48075: n = 142, GSE48276: n = 116, GSE83586: n = 307, GSE86411: n = 132, GSE87304: n = 305, IMvigor: n = 221, Stroggilos 2020: n = 117, TCGA BLCA: n = 407 | | | | | | |
| bDressler 2024: n = 167, E-MTAB-4321: n = 460, Groeneveld 2024: n = 87, GSE120736: n = 84, GSE128959: n = 154, GSE13507: n = 103, GSE32548: n = 92, GSE48075: n = 70, GSE83586: n = 58, GSE86411: n = 132, Stroggilos 2020: n = 98 | | | | | | |
| cBCAN: n = 174, Dressler 2024: n = 75, E-MTAB-4321: n = 16, Groeneveld 2024: n = 82, GSE120736: n = 61, GSE124305: n = 133, GSE128192: n = 112, GSE128701: n = 136, GSE128959: n = 35, GSE13507: n = 62, GSE198269: n = 394, GSE203149: n = 171, GSE32548: n = 38, GSE48075: n = 72, GSE48276: n = 116, GSE83586: n = 243, GSE87304: n = 305, IMvigor: n = 221, Stroggilos 2020: n = 19, TCGA BLCA: n = 407 | | | | | | |
| dComparison of the NMIBC and MIBC groups, the 'unknown' category was excluded. | | | | | | |

Table 3: Demographic, clinical, and pathological characteristic of bladder cancer clusters in a pooled collective of urothelial cancers. Qualitative variables are presented as percentages and counts of complete observations in a cluster. Quantitative variables are shown as medians with interquartile ranges and ranges. Statistical significance of differences between bladder cancer clusters was determined by chi-square test (qualitative variables) or Kruskal-Wallis test. Cramer's V and eta-square served as effect size statistics. P values were corrected for multiple testing with the false discovery rate method.

| **Variablea** | **#1b** | **#2c** | **#3d** | **Significancee** | **Effect sizee** |
| --- | --- | --- | --- | --- | --- |
| samples, N | 1383 | 1041 | 2078 |  |  |
| age, years | 68 [IQR: 60 - 75] range: 22 - 96 complete: n = 638 | 69 [IQR: 61 - 76] range: 35 - 94 complete: n = 493 | 66 [IQR: 58 - 74] range: 24 - 90 complete: n = 898 | p = 0.0014 | η² = 0.0061 |
| sex | female: 11% (148) male: 42% (575) unknown: 48% (660) | female: 14% (141) male: 35% (366) unknown: 51% (534) | female: 9.7% (201) male: 40% (823) unknown: 51% (1054) | p = 0.0015 | V = 0.079 |
| race | Asian: 0.43% (6) Black or African American: 0.87% (12) White: 19% (261) Other: 0.87% (12) unknown: 79% (1092) | Asian: 0.58% (6) Black or African American: 1.5% (16) White: 19% (202) Other: 0.67% (7) unknown: 78% (810) | Asian: 2.1% (43) Black or African American: 0.91% (19) White: 13% (262) Other: 0.58% (12) unknown: 84% (1742) | p < 0.001 | V = 0.11 |
| smoking history | never: 5.1% (71) current or previous: 13% (185) unknown: 81% (1127) | never: 5.9% (61) current or previous: 15% (152) unknown: 80% (828) | never: 4.7% (98) current or previous: 11% (221) unknown: 85% (1759) | ns (p = 0.8) | V = 0.027 |
| cancer tissue type | bladder: 96% (1322) urinary organ, non-bladder: 4% (55) non-urinary organ, distant metastasis: 0.43% (6) | bladder: 95% (992) urinary organ, non-bladder: 4% (42) non-urinary organ, distant metastasis: 0.67% (7) | bladder: 96% (1985) urinary organ, non-bladder: 4.1% (86) non-urinary organ, distant metastasis: 0.34% (7) | ns (p = 0.85) | V = 0.0087 |
| pathology grade | G1: 0.87% (12) G2: 1.7% (24) G3: 13% (176) unknown: 85% (1171) | G1: 0.29% (3) G2: 1.3% (14) G3: 12% (122) unknown: 87% (902) | G1: 2.2% (46) G2: 4.9% (101) G3: 10% (213) unknown: 83% (1718) | p < 0.001 | V = 0.14 |
| histology grade | low grade: 8.5% (118) high grade: 16% (224) unknown: 75% (1041) | low grade: 4.8% (50) high grade: 19% (200) unknown: 76% (791) | low grade: 13% (275) high grade: 15% (306) unknown: 72% (1497) | p < 0.001 | V = 0.22 |
| invasiveness | non-muscle invasive: 28% (393) muscle invasive: 70% (971) unknown: 1.4% (19) | non-muscle invasive: 19% (193) muscle invasive: 80% (832) unknown: 1.5% (16) | non-muscle invasive: 45% (940) muscle invasive: 53% (1107) unknown: 1.5% (31) | p < 0.001 | V = 0.24 |
| pT stage | T2: 22% (310) T3: 16% (224) T4: 6.4% (89) Ta/Tis/T1: 25% (349) T3/T4: 1.1% (15) unknown: 29% (396) | T2: 27% (285) T3: 20% (205) T4: 7% (73) Ta/Tis/T1: 16% (164) T3/T4: 0.77% (8) unknown: 29% (306) | T2: 18% (374) T3: 8.9% (184) T4: 3.1% (65) Ta/Tis/T1: 42% (875) T3/T4: 0.096% (2) unknown: 28% (578) | p < 0.001 | V = 0.16 |
| pN stage | N0: 23% (315) N1+: 11% (146) unknown: 67% (922) | N0: 26% (272) N1+: 9.9% (103) unknown: 64% (666) | N0: 27% (559) N1+: 5.4% (112) unknown: 68% (1407) | p < 0.001 | V = 0.16 |
| pM stage | M0: 15% (202) M1: 2.2% (31) unknown: 83% (1150) | M0: 15% (155) M1: 2% (21) unknown: 83% (865) | M0: 18% (368) M1: 1% (21) unknown: 81% (1689) | p = 0.0029 | V = 0.12 |
| neoadjuvant systemic chemotherapy | no: 19% (269) yes: 8.2% (113) unknown: 72% (1001) | no: 25% (261) yes: 9% (94) unknown: 66% (686) | no: 18% (373) yes: 6.2% (129) unknown: 76% (1576) | ns (p = 0.51) | V = 0.036 |
| history of BCG therapy | no: 18% (253) yes: 4.4% (61) unknown: 77% (1069) | no: 18% (187) yes: 3.1% (32) unknown: 79% (822) | no: 18% (374) yes: 4.2% (87) unknown: 78% (1617) | ns (p = 0.41) | V = 0.047 |
| cystectomy | no: 14% (192) yes: 10% (143) unknown: 76% (1048) | no: 10% (108) yes: 11% (111) unknown: 79% (822) | no: 17% (352) yes: 5% (103) unknown: 78% (1623) | p < 0.001 | V = 0.25 |
| adjuvant systemic chemotherapy | no: 9.9% (137) yes: 21% (295) unknown: 69% (951) | no: 13% (134) yes: 19% (200) unknown: 68% (707) | no: 12% (255) yes: 16% (340) unknown: 71% (1483) | p = 0.002 | V = 0.099 |
| apT stage: pathological tumor stage; pN stage: pathological lymph node stage; pM stage: pathological metastasis stage; BCG: Bacillus Calmette-Guerin. | | | | | |
| bBCAN: n = 64, Dressler 2024: n = 69, E-MTAB-4321: n = 153, Groeneveld 2024: n = 60, NA: n = 23, GSE120736: n = 36, GSE124305: n = 44, GSE128192: n = 34, GSE128701: n = 50, GSE128959: n = 75, GSE13507: n = 51, GSE198269: n = 131, GSE203149: n = 52, GSE32548: n = 34, GSE48075: n = 34, GSE48276: n = 36, GSE83586: n = 82, GSE86411: n = 41, GSE87304: n = 79, IMvigor: n = 75, Stroggilos 2020: n = 40, TCGA BLCA: n = 120 | | | | | |
| cBCAN: n = 47, Dressler 2024: n = 50, E-MTAB-4321: n = 81, Groeneveld 2024: n = 37, NA: n = 18, GSE120736: n = 28, GSE124305: n = 44, GSE128192: n = 37, GSE128701: n = 34, GSE128959: n = 24, GSE13507: n = 31, GSE198269: n = 88, GSE203149: n = 34, GSE32548: n = 23, GSE48075: n = 35, GSE48276: n = 33, GSE83586: n = 85, GSE86411: n = 28, GSE87304: n = 92, IMvigor: n = 56, Stroggilos 2020: n = 19, TCGA BLCA: n = 117 | | | | | |
| dBCAN: n = 63, Dressler 2024: n = 123, E-MTAB-4321: n = 242, Groeneveld 2024: n = 101, NA: n = 22, GSE120736: n = 81, GSE124305: n = 45, GSE128192: n = 41, GSE128701: n = 52, GSE128959: n = 93, GSE13507: n = 106, GSE198269: n = 175, GSE203149: n = 85, GSE32548: n = 74, GSE48075: n = 73, GSE48276: n = 47, GSE83586: n = 140, GSE86411: n = 63, GSE87304: n = 134, IMvigor: n = 90, Stroggilos 2020: n = 58, TCGA BLCA: n = 170 | | | | | |
| eComparison of bladder cancer clusters, the 'unknown' category was excluded. | | | | | |

Table 4: Sensitivity of UC cell lines representative for bladder cancer clusters to anti-cancer drugs. The sensitivity was measured as inhibition of the cell culture growth at four days following the drug treatment. Drug concentrations resulting in 50% growth inhibition (IC50, µM) were determined in three independent biological replicates, and are presented as medians and ranges.

| **Molecular target** | **Compound name** | **Cluster #1, UMUC3** | **Cluster #2, 5637** | **Cluster #3, RT112** |
| --- | --- | --- | --- | --- |
| ERBB | Gefitinib | 11 [9.7 - 12] | 8.8 [7.3 - 11] | 15 [14 - 16] |
| Neratinib | 4.9 [4.5 - 5.2] | 0.28 [0.15 - 0.75] | 8.7 [7 - 9.5] |
| Osimertinib | 7.2 [6.3 - 8.5] | 2.5 [2.1 - 4.3] | 4.4 [3.9 - 5.6] |
| Afatinib | 2.8 [2.2 - 3.2] | 2.1 [1.7 - 2.9] | 3.5 [2.1 - 4.8] |
| MEK | Cobimetinib | 2.2 [1.8 - 4.3] | 1.3 [1.1 - 1.9] | 3 [1.5 - 6.5] |
| Trametinib | 12 [11 - 14] | 1.1 [0.71 - 1.9] | 8 [7.8 - 11] |
| DNA | Cisplatin | 3.5 [3.2 - 3.9] | 4.4 [4 - 4.7] | 6.9 [6.2 - 7.1] |
| Gemcitabine | 0.0033 [0.0014 - 0.0033] | 0.0084 [0.0061 - 0.0097] | 0.0096 [0.0057 - 0.01] |
| FGFR | Erdafitinib | 6.7 [5.3 - 6.8] | 7.4 [6.7 - 7.8] | 7.2 [7.1 - 7.2] |
| mTOR | Torin2 | 0.013 [0.011 - 0.015] | 0.016 [0.013 - 0.029] | 0.009 [0.004 - 0.029] |
| STAT3 | Stattic | 2.6 [2 - 2.9] | 2 [1.7 - 2.4] | 2.2 [2.1 - 2.2] |

# Figures



Figure 1: Analysis scheme, placeholder

**Figure 1. Analysis scheme, placeholder**

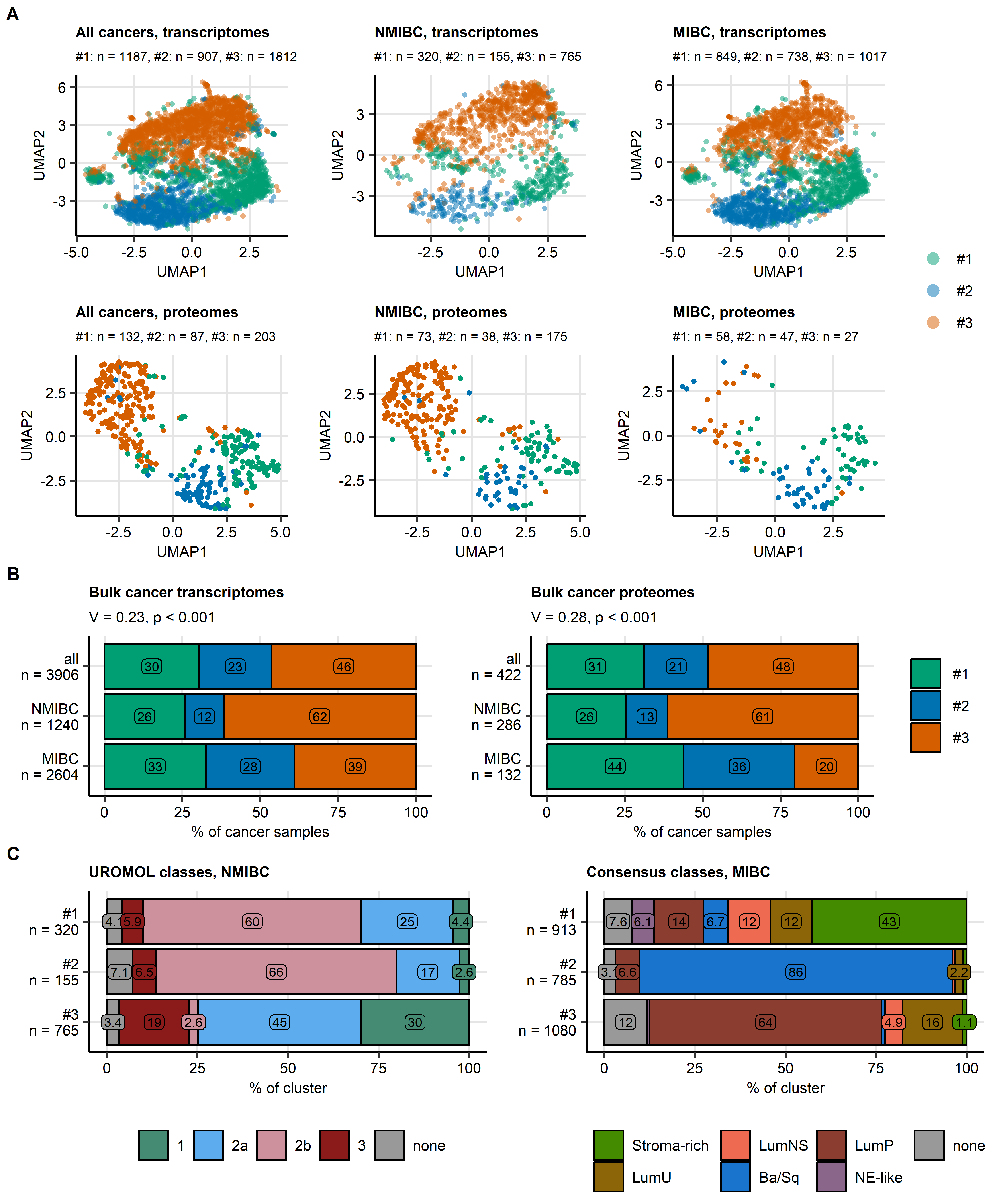


Figure 2: Bladder cancer clusters of bulk cancer transcriptomes and proteomes.

**Figure 2. Bladder cancer clusters of bulk cancer transcriptomes and proteomes.**

*Bladder cancer clusters were developed in the TCGA BLCA bulk cancer transcriptome cohort by regularized KMEANS clustering of metagenes defined by a self-organizing map.* *Bulk cancer samples in other transcriptomic cohorts were assigned to bladder cancer clusters by a k-nearest neighbor classifier trained in the TCGA BLCA cohort.* *Bulk cancer transcriptomes in the BCAN cohort and bulk cancer proteome samples were assigned to bladder cancer clusters by Elastic Net classifiers trained in the TCGA BLCA cohort.* *Predictions of NMIBC UROMOL and MIBC consensus classes for, respectively, non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC) bulk cancer transcriptome samples were made by nearest centroid classifiers (R packages classifyNMIBC and consensusMIBC) fed with ComBat batch effect-adjusted mRNA expression levels of 12394 genes measured in all collectives.*

*(A) Visualization of bladder cancer clusters via dimensionality reduction of mRNA and protein expression data. UMAP (uniform manifold approximation and projection) embeddings of scores of cluster-defining metagenes (256 metagenes), and ComBat-adjusted expression levels of proteins used by the Elastic Net classifier of bladder cancer cluster assignment (141 protein) were computed. The embeddings for all available bulk cancer samples, NMIBC and MIBC specimens are visualized as scatter plots. Each point represents a single cancer sample. Point color codes for cluster assignment. Numbers of samples in bladder cancer clusters are displayed in the plot captions.*

*(B) Size of bladder cancer clusters in all, NMIBC, and MIBC cancers. Differences of sizes of bladder cancer clusters between NMIBC and MIBC collectives were assessed by test with Cramer’s V effect size statistic. Percentages of samples assigned to bladder cancer clusters in all, NMIBC and MIBC bulk transcriptome and bulk proteome collectives are depicted in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of samples are indicated in the Y axes.*

*(C) Assignment of bulk cancer transcriptomes to bladder cancer clusters, NMIBC UROMOL classes, and MIBC consensus classes. Percentages of NMIBC and MIBC samples in, respectively, NMIBC UROMOL classes and MIBC consensus classes within bladder cancer clusters are presented in stack plots. Numbers of samples in bladder cancer clusters are indicated in the Y axes.*

*LumU: luminal unstable; LumNS: luminal non-specified; Ba/Sq: basal/squamous-like; LumP: luminal papillary; NE-like: neuroendocrine-like, none: not assigned.*

*Data sources for bulk transcriptome samples: TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, and Groeneveld 2024 cohorts.*

*Data sources for bulk proteome samples: Groeneveld 2024, Dressler 2024, and Stroggilos 2020 cohorts.*



Figure 3: Pathological staging and overall survival in bladder cancer clusters. Evolution of bladder cancer clusters during recurrence and progression.

**Figure 3. Pathological staging and overall survival in bladder cancer clusters. Evolution of bladder cancer clusters during recurrence and progression.**

*(A) Pathological stages of tumor (pT), lymph nodes (pN), and metastases (pM) in bladder cancer clusters of urothelial cancers. Percentages of the stages within bladder cancer clusters in the pooled collective of all bulk transcriptome and bulk proteome specimens are presented in stack plots. Numbers of samples in bladder cancer clusters are indicated in the Y axes. Data sources, bulk cancer transcriptomes: TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, Groeneveld 2024 mRNA, and BCAN cohorts. Data sources, bulk cancer proteomes: Groeneveld 2024 protein, Dressler 2024, and Stroggilos 2020 cohorts.*

*(B) Overall survival was compared between bladder cancer clusters of transcriptomic and proteomic cohorts with at least 80 patients with complete survival information (transcriptomic cohorts: TCGA BLCA, IMvigor, GSE13507, and Groeneveld 2024; proteomic cohorts: Dressler 2024). Statistical significance of global differences in overall survival between bladder cancer clusters was determined by Peto-Peto test. P values were corrected for multiple testing with the false discovery rate method (FDR). Fractions of surviving patients are presented in Kaplan-Meier plots for collectives, where significant global differences in survival between bladder cancer clusters were observed (pFDR < 0.05). FDR-corrected p values are displayed in the plots. Numbers of observations and deaths in bladder cancer clusters are indicated in the plot legends.*

*(C) Bulk mRNA sequencing results for patient-matched primary and recurrent cancers, and patient-matched primary non-muscle invasive (NMIBC) and muscle invasive bladder cancer (MIBC) progression specimens were available for the GSE13507 and GSE128959 cohorts. Changes of bladder cluster assignment between patient-matched primary cancer and first recurrence specimens, and between patient-matched primary NMIBC and MIBC progression samples are visualized in alluvial plots. Total numbers of samples are displayed in the plot captions.*

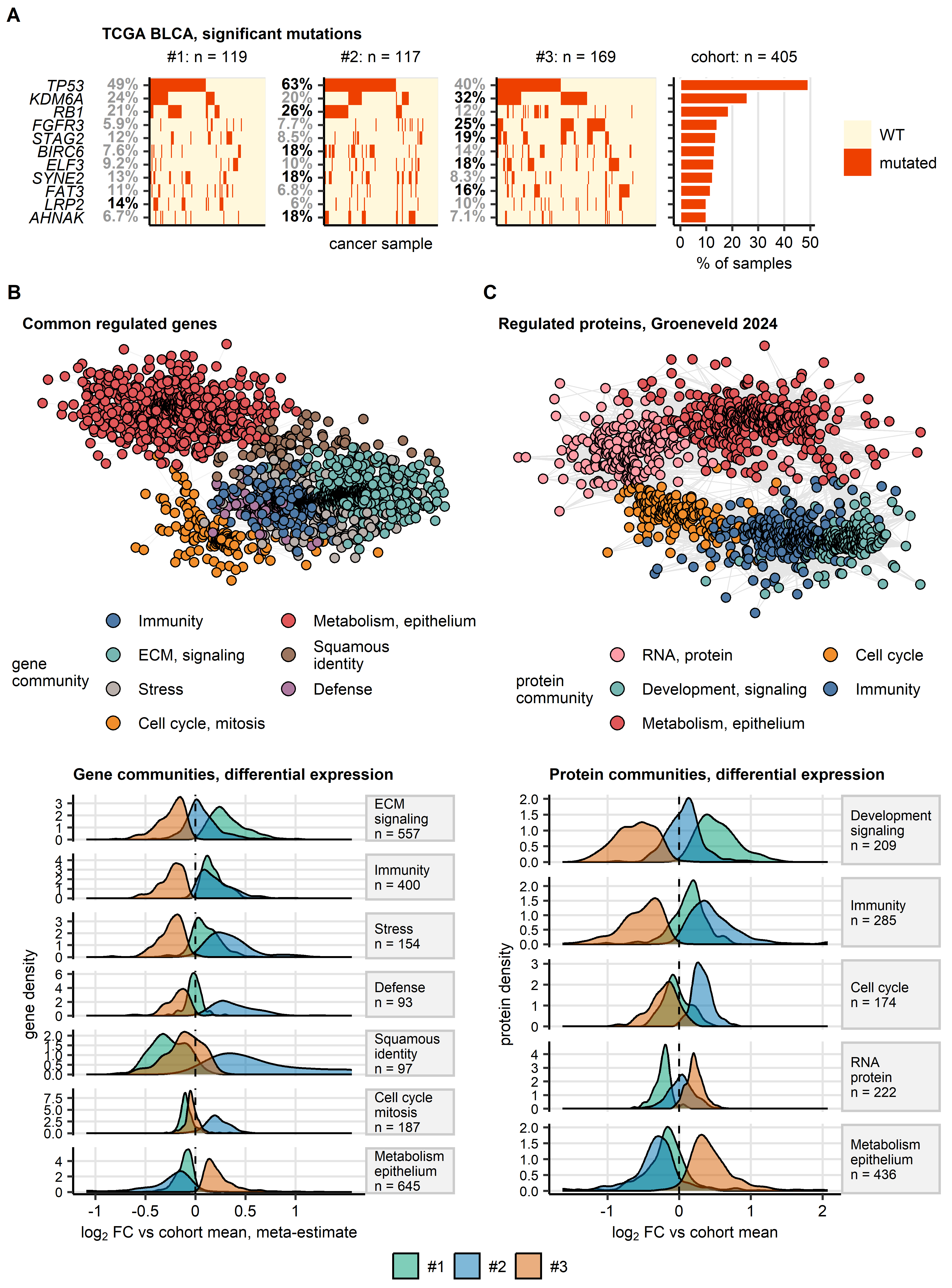


Figure 4: Genetics, transcriptome, and proteome of bladder cancer clusters.

**Figure 4. Genetics, transcriptome, and proteome of bladder cancer clusters.**

*(A) Enrichment analysis of somatic mutation in bladder cancer clusters. Enrichment of somatic mutations among bulk cancer samples assigned to bladder cancer clusters in TCGA BLCA, IMvigor, and BCAN cohorts was assessed by permutation tests. Mutation status of genes with significant enrichment (raw p < 0.05 without multiple testing correction) in at least one of the clusters and overall mutation rate 10% in the TCGA BLCA cohort is shown in oncoplots. Each oncoplot tile represents a cancer sample. Percentages of samples with mutations within the clusters are indicated in the Y axes of the oncoplots; significant enrichment is highlighted by bold font. The overall mutation percentages are presented in the bar plot. Numbers of specimens in bladder cancer cluster and in the entire cohort are displayed above the plots.*

*(BC) Co-regulation networks of genes (B) and proteins (C) differentially expressed in bladder cancer clusters as compared with the cohort means. -transformed gene and protein expression levels (no ComBat adjustment) were compared between bladder cancer clusters in transcriptomic and proteomic bulk cancer cohorts by one-way ANOVA with effect size statistic. Differences between mean expression levels in the clusters and the respective cohort means were evaluated by one-sample T test. P values were corrected for multiple comparisons with the false discovery rate (FDR) method. Significantly differentially regulated genes and proteins were defined by ANOVA pFDR < 0.05, 0.06, and T test pFDR < 0.05.* *Meta-estimates of fold differential gene expression for features differentially regulated between the clusters in at least 10 cohorts were computed with the DerSimonian-Lair inverse variance method.* *Co-expression patterns of the differentially regulated genes shared by at least 10 out of 19 transcriptomic cohorts and of the differentially regulated proteins in proteomic collectives were investigated by network analysis.* *The co-regulation transcriptomic network was build in a pooled data set of ComBat-adjusted gene expression values (total samples: n = 3906; sample sources TCGA BLCA: n = 407, IMvigor: n = 221, GSE13507: n = 188, GSE32548: n = 131, GSE48075: n = 142, GSE48276: n = 116, GSE83586: n = 307, GSE86411: n = 132, GSE87304: n = 305, GSE120736: n = 145, GSE124305: n = 133, GSE128192: n = 112, GSE128701: n = 136, GSE128959: n = 192, GSE198269: n = 394, GSE203149: n = 171, E-MTAB-4321: n = 476, and Groeneveld 2024: n = 198).* *The proteomic co-regulation networks were build separately for each proteomic cohort (sample numbers: Groeneveld 2024: n = 63, Dressler 2024: n = 242, and Stroggilos 2020: n = 117).* *The networks’ edges were defined by pairwise correlations of gene or protein levels with Spearman’s 0.5 and weighted by the corresponding values. Isolated vertices, i.e. genes or proteins without correlation partners were removed.* *Communities of co-regulated genes and proteins were identified by the Leiden algorithm and named after their key biological processes revealed by biological process gene ontology enrichment.* *The co-expression networks were visualized as force-directed graphs with the Fruchterman–Reingold algorithm (B: pooled transcriptome cohort; C: Groeneveld 2024 proteome cohort). Single genes are represented by points, correlations between genes with 0.5 are depicted as lines. Point color codes for the community assignment.* *Density of meta-estimates (B) and estimates (C) of fold changes of expression in the clusters as compared with the cohort mean for members of gene and protein communities are presented in density plots; numbers of genes and proteins in the communities are indicated in the plot facets.*

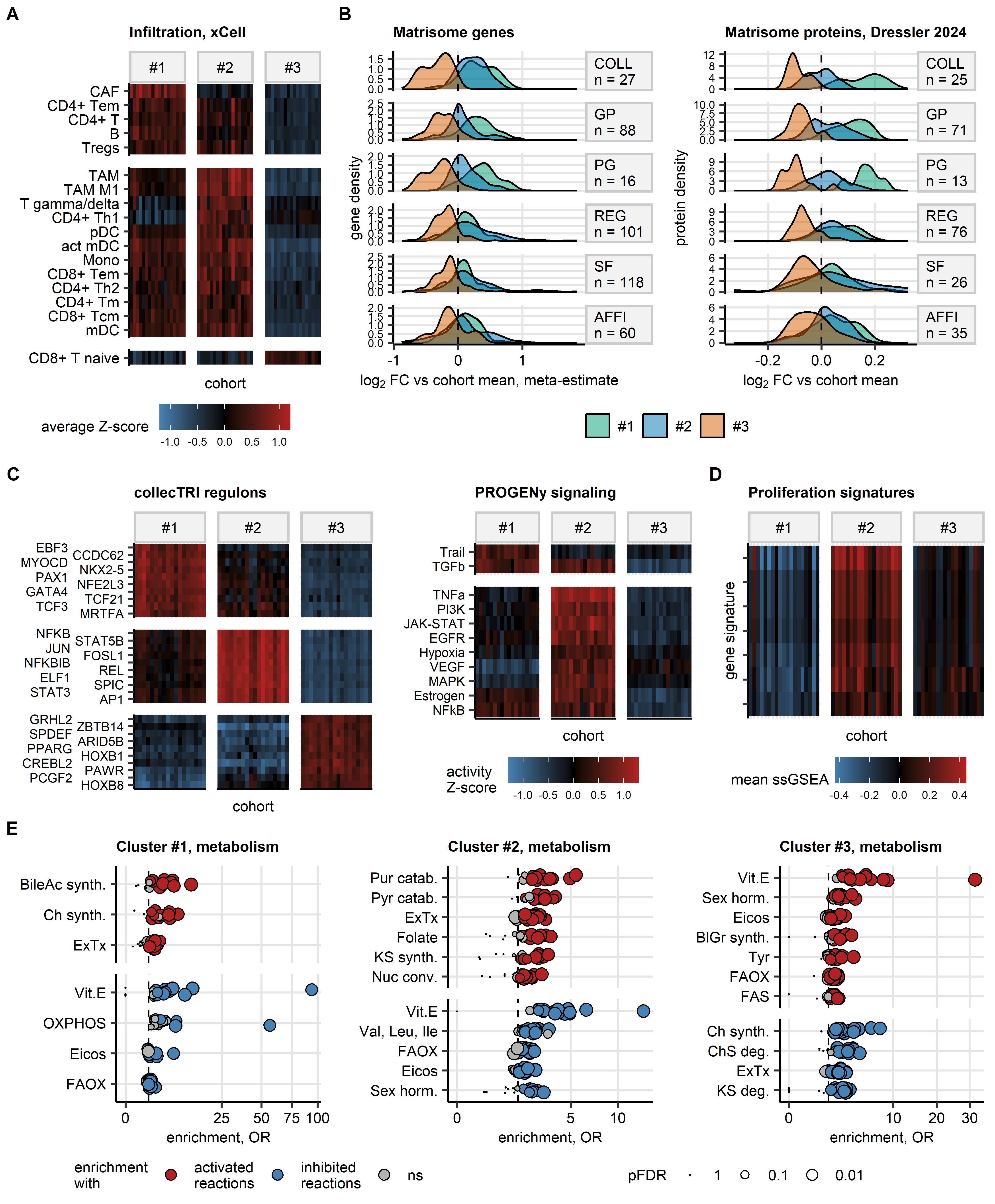


Figure 5: Tumor microenvironment, signaling, and metabolism of bladder cancer clusters.

**Figure 5. Tumor microenvironment, signaling, and metabolism of bladder cancer clusters.**

*(A) Fractions of infiltrating immune and stromal cells predicted by xCell immunedeconvolution of bulk cancer transcriptomes (n = 18 cohorts; TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, and Groeneveld 2024). The fractions of infiltrating cells were compared between bladder cancer clusters by Kruskal-Wallis test with effect size statistic. P values were corrected for multiple testing with the false discovery rate (FDR) method. Significant differences were considered for pFDR < 0.05 and 0.06. Levels of infiltrating cell populations found to differ between bladder cancer clusters in at least 10 cohorts are visualized in a summary heat map. The heat map tiles represent cohorts and clusters; tile color corresponds to average cell fraction Z-score in the cluster and cohort.*

*(B) Differential expression of matrisome genes and proteins in bladder cancer clusters. Differential gene and protein expression in bladder cancer clusters as compared with the cohort means was assessed by one-way ANOVA and one-sample post-hoc T test as presented in Figure 4BC. Meta-estimates of fold-regulation.* *Meta-estimates of fold differential gene expression for features differentially regulated between the clusters in at least 10 out of 19 transcriptomic cohorts (TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, Groeneveld 2024, and BCAN) were computed with the DerSimonian-Lair inverse variance method.* *Density of meta-estimates and estimates of fold changes of expression in the clusters as compared with the cohort mean for significantly differentially regulated matrisome genes and proteins shared by, respectively, at least 10 and two cohorts are presented in density plots. The differentially regulated gene and proteins were classified by matrisome categories (COLL: collagens; GP: glycoproteins; PR: proteglycans; REG: regulatory proteins; SF: secreted factors; AFFI: affiliated proteins); numbers of differentially regulated genes and proteins in the matrisome categories are displayed in the plot facets.*

*(C) Activity of transcriptional regulons and signaling pathways in bladder cancer cluster predicted with bulk cancer transcriptome data by decoupleR modeling algorithms (n = 18 cohorts; TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, and Groeneveld 2024). The activity linear modeling (LM) scores of regulons and signaling pathways were predicted for single bulk cancer transcriptome samples by, respectively, univariable and multivariable modeling with collecTRI and PROGENy knowledge models. LM scores were compared between bladder cancer clusters by one-way ANOVA with effect size statistic. Differences between mean LM scores in the clusters and the respective cohort means were evaluated by one-sample T test. P values were corrected for multiple comparisons with the FDR method. Significantly differentially regulated regulons and signaling pathways were defined by ANOVA pFDR < 0.05, 0.06, and T test pFDR < 0.05.* *Activities of the top 10 strongest activated regulons shared by at least 10 cohorts and activities of the differentially regulated signaling pathways shared by at 10 cohorts are visualized in summary heat maps. The heat map tiles represent cohorts and clusters; tile color corresponds to average cell fraction Z-score in the cluster and cohort.*

*(D) Gene signatures of cell proliferation in bladder cancer clusters. Single sample gene set enrichment analysis (ssGSEA) scores of 7 published gene signatures of cell proliferation were compared between bladder cancer clusters in 19 transcriptomic cohorts (TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, Groeneveld 2024, and BCAN) by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the FDR method. Significant differences were considered for pFDR < 0.05 and 0.06. ssGSEA scores of proliferation signatures found to differ significantly between the clusters in at least 10 cohorts are visualized in a summary heat map. The heat map tiles represent cohorts and clusters; tile color corresponds to average ssGSEA score in the cluster and cohort.*

*(E) Metabolism of bladder cancer clusters predicted for bulk cancer transcriptome data (n = 19 cohorts; TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, Groeneveld 2024, and BCAN). Activity of RECON metabolic reactions was modeled by Monte Carlo simulations with tools of BiGGR and biggrExtra packages fed with fold-regulation estimates and standard errors of differential expression for all measured genes. Enrichment of RECON metabolic subsystems with significantly activated and inhibited reactions was assessed by comparison with 10000 random draws from the entire reaction pool. Enrichment p values were corrected for multiple testing with the FDR method. Odds ratio (OR) served as a metric of enrichment magnitude. Significant enrichment with activated or inhibited reactions was considered for pFDR < 0.05 and OR 1.44. For the significant subsystems shared by at least 10 cohorts, OR values are presented in dot plots. Each point represents a single cohort. Point color codes for reaction activity status and significance, point size codes for FDR-corrected p value (BileAc synth.: bile acid synthesis; Ch synth.: chondroitin synthesis; ExTx: extracellular transport; Vit.E: vitamin E; OXPHOS: oxidative phosphorylation; Eicos: eicosanoid metabolism; FAOX: fatty acid oxidation; Pur catab.: purine catabolism; Pyr catab.: pyrimidine catabolism; KS synth.: keratan sulfate synthesis; Nuc conv.: nucleotide conversion; Val, Leu, Ile: valine, leucine, and isoleucine metabolism; Sex horm.: sex hormone metabolism; BlGr synth.: blood group synthesis; Tyr: tyrosine metabolism; FAS: fatty acid synthesis; ChS deg. chondroitin sulfate degradation).*

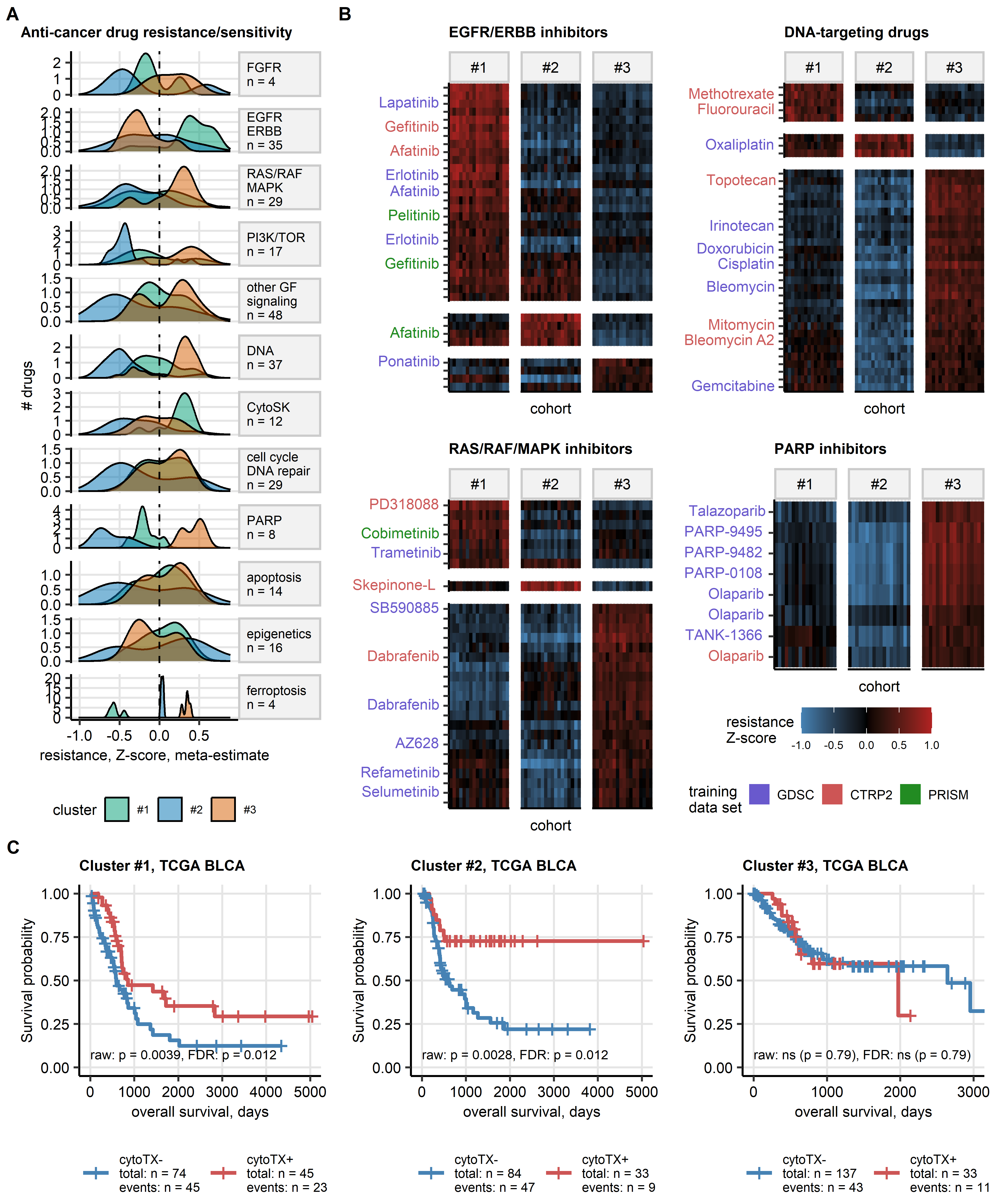


Figure 6: Predictions of resistance and sensitivity to anti-cancer drugs in bladder cancer clusters.

**Figure 6. Predictions of resistance and sensitivity to anti-cancer drugs in bladder cancer clusters.**

*Based on whole-transcriptome expression data, resistance to anti-cancer drugs was predicted for bulk cancer transcriptome samples (n = 18 cohorts; TCGA BLCA, IMvigor, GSE13507, GSE32548, GSE48075, GSE48276, GSE83586, GSE86411, GSE87304, GSE120736, GSE124305, GSE128192, GSE128701, GSE128959, GSE198269, GSE203149, E-MTAB-4321, and Groeneveld 2024) by RIDGE linear models trained with anti-cancer drug resistance data and baseline gene expression of epithelial cancer cell lines recorded in the GDSC, CTRP2 and PRISM drug screening experiments.* *Predicated drug resistance in form of Z-scores of IC50 (concentration resulting 50% growth inhibition) and AUC (area under the dose-response curve) was compared between bladder cancer clusters by one-way ANOVA with effect size statistic. Differences between mean resistance Z&-scores in the clusters and the respective cohort means were evaluated by one-sample T test. P values were corrected for multiple comparisons with the false discovery rate (FDR) method. Significant differences in predicted resistance were considered for ANOVA pFDR < 0.05, 0.06, and T test pFDR < 0.05.* *Meta-estimates of differences of resistance Z-scores the cluster and the cohort mean for compounds with significant differences in predicted resistance shared by at at least 10 cohorts were computed with the DerSimonian-Lair inverse variance method.*

*(A) Densities of meta-estimates of resistance Z-scores between bladder cancer clusters and the cohort mean for the compounds with significant differences in predicted resistance shared by at at least 10 cohorts. The compounds were categorized by their molecular targets (GF: growth factors; CytoSK: cytoskeleton). Note: density curves shifted to right are characteristic for resistance, density curves shifted to left denote predicted sensitivity. The compound numbers are displayed in the plot facets.*

*(B) Resistance Z-scores of compounds targeting EGFR/ERBB and RAS/RAF/MAPK signaling, DNA and PARP with significant differences between bladder cancer clusters shared by at least 10 cohorts. The resistance Z-scores are visualized in a summary heat map. The heat map tiles represent cohorts and clusters; tile color corresponds to average resistance Z-score in the cluster and cohort. Names of compounds of clinical relevance are indicated in Y axes of the heat maps. The compound name text color codes for the in vitro drug screening data used for resistance prediction (blue: GDSC; red: CTRP2; green: PRISM).*

*(C) Overall survival in bladder cancer clusters was compared between TCGA BLCA patients with systemic cytotoxic chemotherapy (cytoTX+, gemcitabine, platinum, topoisomerase inhibitors, alkylating agents, metothrexate) and without systemic cytotoxic chemotherapy (cytoTX-). Statistical significance was determined by Peto-Peto test corrected for multiple testing with the false discovery rate (FDR) method. Fractions of surviving patients are shown in Kaplan-Meier plots. Raw unadjusted p values and FDR-corrected p values of survival differences are displayed in the plots. Total numbers of observations and deaths are displayed in the plot captions. Numbers of patients and deaths in cytotoxic treatment subsets in bladder cancer clusters are indicated in the plot legends.*

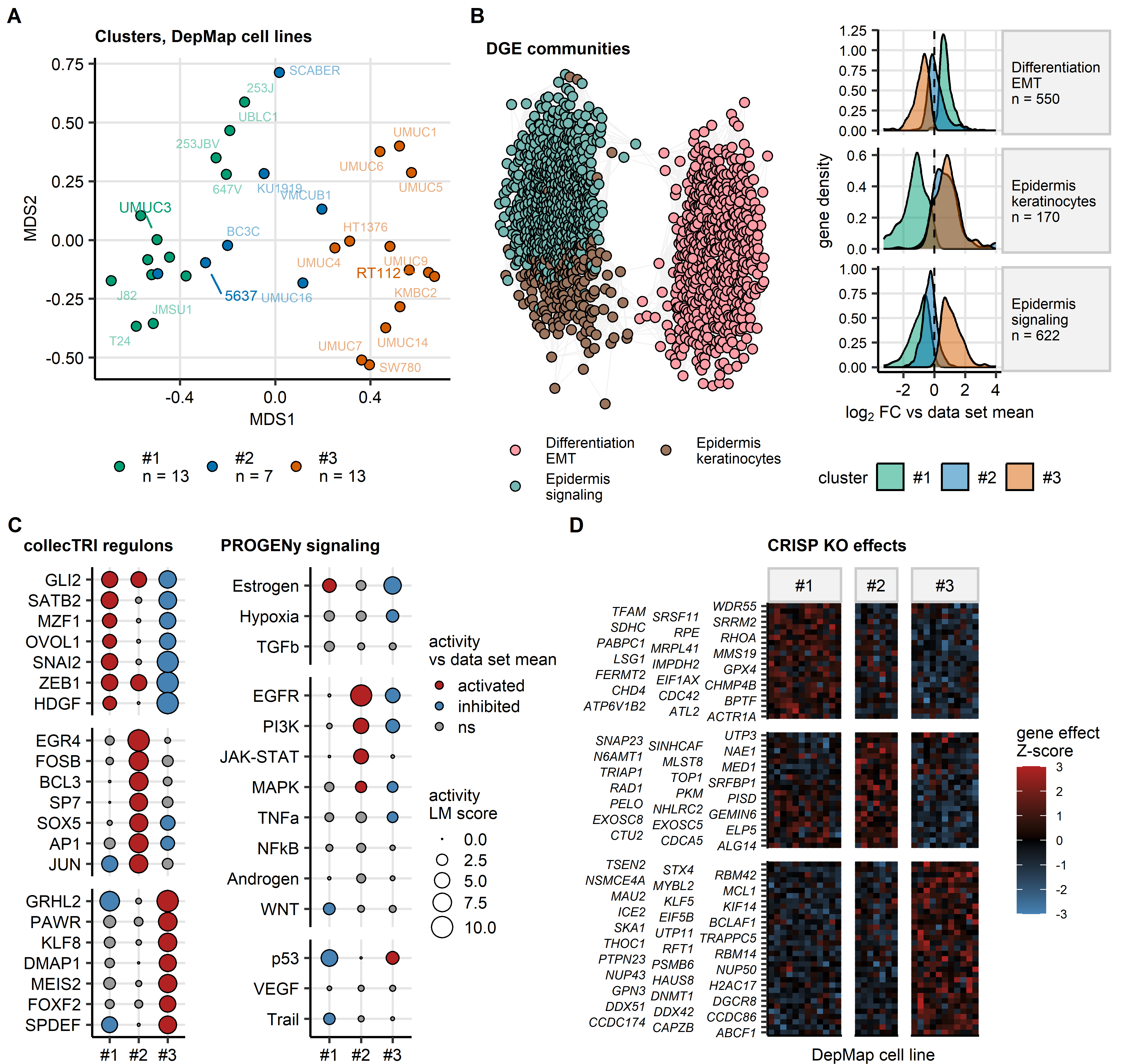


Figure 7: Phenotype of bladder cancer clusters of urothelial cancer cell lines.

**Figure 7. Phenotype of bladder cancer clusters of urothelial cancer cell lines.**

*DepMap urothelial cell lines (n = 33) were assigned to bladder cancer clusters by an Elastic Net classifier trained in the TCGA BLCA bulk cancer cohort.*

*(A) Visualization of bladder cancer clusters of DepMap cell lines via multi-dimensional scaling (MDS). The MDS embeddings of transcript levels used by the Elastic Net classifier are visualized as a scatter plot. Each point represents a single cell line. Point color codes for cluster assignment; points are labeled with cell line names. Numbers of cell lines in bladder cancer clusters are indicated in the plot legend.*

*(B) Co-regulation network of genes differentially expressed between bladder cancer clusters of DepMap cell lines. -transformed gene expression levels were compared between bladder cancer clusters by one-way ANOVA with effect size statistic. Differences between mean expression levels in the clusters and the respective DepMap collection means were evaluated by one-sample T test. P values were corrected for multiple comparisons with the false discovery rate (FDR) method. Significantly differentially regulated genes and proteins were defined by ANOVA pFDR < 0.05, 0.06, and T test pFDR < 0.05.* *Co-expression patterns of the differentially regulated genes were investigated by a network analysis. The networks edges were defined by pairwise correlations of gene or protein levels with Spearman’s 0.5 and weighted by the corresponding values. Isolated vertices, i.e. genes without correlation partners were removed. Communities of co-regulated genes were identified by the Leiden algorithm and named after their key biological processes revealed by biological process gene ontology enrichment.* *The co-expression network was visualized as a force-directed graph with the Fruchterman–Reingold algorithm with single genes represented by points, and correlations between genes with 0.5 depicted as edges. Point color codes for the community assignment.* *Density of estimates of fold changes of expression in the clusters as compared with the DepMap collection mean for members of gene and protein communities are presented in density plots; numbers of genes in the communities are indicated in the plot facets.*

*(C) Activity of transcriptional regulons and signaling in bladder cancer clusters of DepMap cell lines. Activity of collecTRI transcriptional regulons and PROGENy signaling pathways in bladder cancer clusters as compared with the DepMap collection means was modeled by, respectively, univariable and multivariable linear decoupleR modeling algorithms. The algorithms were fed with p value-weighted fold-change estimates of differential expression for all available genes. P values were corrected for multiple testing with the FDR method. Linear modeling (LM) score served as a metric of activity in a cluster as compared with the DepMap collection mean. The LM scores for the top 7 strongest cluster-specific regulons and all investigated signaling pathways are presented in bubble plots. Point sizes code for absolute values of LM scores. Point colors code for the activity status and significance.*

*(D) Identification of essential genes specific for bladder cancer clusters by analysis of the DepMap CRISPR-Cas9 knockout screen. Distribution of Chronos scores (positive values: growth inhibition) of CRISPR-Cas9 knockouts in bladder cancer clusters of DepMap urothelial cancer cell lines (#1: n = 12, #2: n = 7, #3: n = 11) was modeled by bootstrapping. For n = 77 genes, biologically significant knockout effects with Chronos score > 0.5 in at least one cluster, and significant differences in Chronos scores between the clusters were discerned. Their Chronos scores are presented in a heat map.*

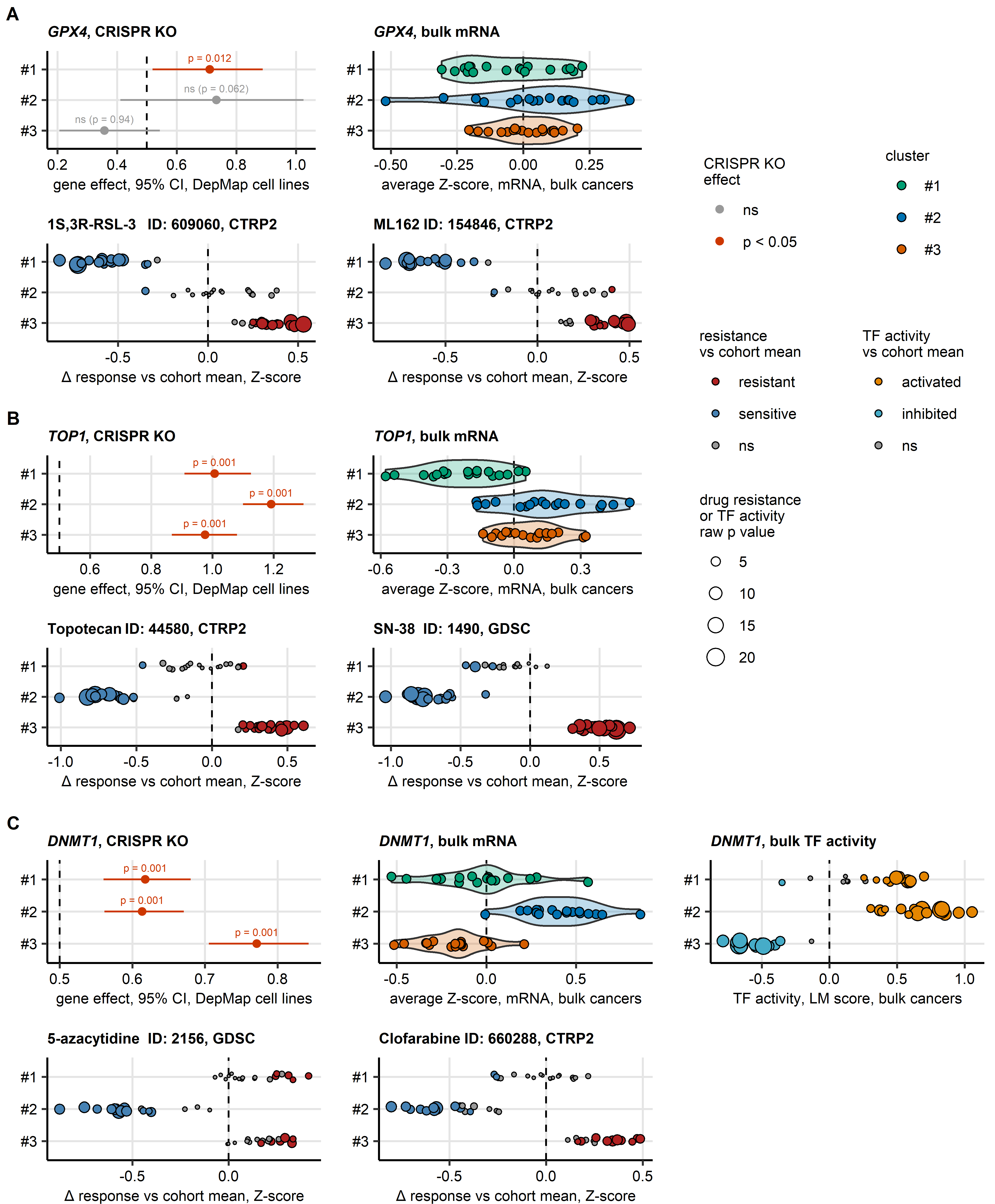


Figure 8: Identification of druggable essential genes for bladder cancer data by integration of cell line and bulk cancer information.

**Figure 8. Identification of druggable essential genes for bladder cancer data by integration of cell line and bulk cancer information.**

*Candidate essential genes of bladder cancer clusters with existing pharmacological inhibitors or modulators were identified by inspection of the CRISPR-Cas9 knockout screen results of DepMap urothelial cell lines (Figure 7D, #1: n = 12, #2: n = 7, #3: n = 11 cell lines), differential gene expression in bulk cancer transcriptomes (Figure 4B, n = 19 cohorts), activity of transcriptional regulons predicted for bulk cancer transcriptomes by collecTRI/decoupleR (Figure 5C, n = 18 cohorts), and drug response predictions made for bulk cancer transcriptomes by RIDGE machine learning (Figure 6, n = 18 cohorts).*

*Estimates of distribution of Chronos score of effects of CRISPR-Cas9 knockouts in bladder cancer cluster of the cell lines were obtained by bootstrapping. The estimated means Chronos scores with 95% confidence intervals (95% CI) in the clusters of DepMap cell lines are shown in Forest plots (points: means, whiskers: 95% CI; significant biological effects with Chronos scores > 0.5 labeled in red).* *Distributions of average Z-scores of -transformed mRNA levels in bulk cancers are depicted in violin plots, with averages of single cohorts and clusters represented by points.* *Linear modeling (LM) scores of activity of collecTRI regulons (TF: transcription factors) are presented in dot plots. Point sizes in these dot plots correspond to -log10-transformed p values; point colors code for activity status and significance.* *Estimates of differences in drug resistance Z-scores between bladder cancer clusters and the cohort means are presented in dot plots. Point sizes in these dot plots correspond to -log10-transformed p values; point colors code for resistance status and significance. Compound name and identifier, and name of the data set used for prediction of resistance are indicated in the plot titles.*

\_(A) *GPX4 as a candidate essential gene specific for cluster #1, and GPX4 inhibitors RSL-3 and ML162 and ferroptosis inducers as candidate therapy of cluster #1 cancers.*

*(B) Topoisomerase-coding TOP1 as a candidate essential gene for cluster #2, and topoisomerase I inhibitors Topotecan and SN-38 as candidate therapy for cluster #2 cancers.*

*(C) DNA methylase-coding DNMT1 as a candidate essential gene for cluster #2. Note the substantial biological effects of DNMT1 knockout in all clusters, and upregulation of DNMT1 expression and activity in cluster #2. DNMT1 inhibitors Azacytidine and Clofarabine are proposed as candidate therapeutic options for cluster #2 cancers.*

# References

1. 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)

2. 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)

3. 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)

4. 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)

5. Kim WJ, Kim EJ, Kim SK, Kim YJ, Ha YS, Jeong PP, Kim MJ, Yun SJ, Lee KM, Moon SK, et al. Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. *Molecular cancer* (2010) 9: doi: [10.1186/1476-4598-9-3](https://doi.org/10.1186/1476-4598-9-3)

6. Lindgren D, Sjödahl G, Lauss M, Staaf J, Chebil G, Lövgren K, Gudjonsson S, Liedberg F, Patschan O, Månsson W, et al. Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PloS one* (2012) 7: doi: [10.1371/JOURNAL.PONE.0038863](https://doi.org/10.1371/JOURNAL.PONE.0038863)

7. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, Roth B, Cheng T, Tran M, Lee IL, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer cell* (2014) 25:152–165. doi: [10.1016/J.CCR.2014.01.009](https://doi.org/10.1016/J.CCR.2014.01.009)

8. Sjödahl G, Eriksson P, Liedberg F, Höglund M. Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. *The Journal of pathology* (2017) 242:113–125. doi: [10.1002/PATH.4886](https://doi.org/10.1002/PATH.4886)

9. Guo CC, Bondaruk J, Yao H, Wang Z, Zhang L, Lee S, Lee JG, Cogdell D, Zhang M, Yang G, et al. Assessment of Luminal and Basal Phenotypes in Bladder Cancer. *Scientific reports* (2020) 10: doi: [10.1038/S41598-020-66747-7](https://doi.org/10.1038/S41598-020-66747-7)

10. Seiler R, Ashab HAD, Erho N, Rhijn BWG van, Winters B, Douglas J, Van Kessel KE, Fransen van de Putte EE, Sommerlad M, Wang NQ, et al. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *European urology* (2017) 72:544–554. doi: [10.1016/J.EURURO.2017.03.030](https://doi.org/10.1016/J.EURURO.2017.03.030)

11. Song BN, Kim SK, Mun JY, Choi YD, Leem SH, Chu IS. Identification of an immunotherapy-responsive molecular subtype of bladder cancer. *EBioMedicine* (2019) 50:238–245. doi: [10.1016/J.EBIOM.2019.10.058](https://doi.org/10.1016/J.EBIOM.2019.10.058)

12. Seiler R, Gibb EA, Wang NQ, Oo HZ, Lam HM, Kessel KE van, Voskuilen CS, Winters B, Erho N, Takhar MM, et al. Divergent Biological Response to Neoadjuvant Chemotherapy in Muscle-invasive Bladder Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* (2019) 25:5082–5093. doi: [10.1158/1078-0432.CCR-18-1106](https://doi.org/10.1158/1078-0432.CCR-18-1106)

13. Guo CC, Majewski T, Zhang L, Yao H, Bondaruk J, Wang Y, Zhang S, Wang Z, Lee JG, Lee S, et al. Dysregulation of EMT Drives the Progression to Clinically Aggressive Sarcomatoid Bladder Cancer. *Cell reports* (2019) 27:1781–1793.e4. doi: [10.1016/J.CELREP.2019.04.048](https://doi.org/10.1016/J.CELREP.2019.04.048)

14. Efstathiou JA, Mouw KW, Gibb EA, Liu Y, Wu CL, Drumm MR, Costa JB da, Plessis M du, Wang NQ, Davicioni E, et al. Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. *European urology* (2019) 76:59–68. doi: [10.1016/J.EURURO.2019.01.011](https://doi.org/10.1016/J.EURURO.2019.01.011)

15. Sjödahl G, Eriksson P, Patschan O, Marzouka NAD, Jakobsson L, Bernardo C, Lövgren K, Chebil G, Zwarthoff E, Liedberg F, et al. Molecular changes during progression from nonmuscle invasive to advanced urothelial carcinoma. *International journal of cancer* (2020) 146:2636–2647. doi: [10.1002/IJC.32737](https://doi.org/10.1002/IJC.32737)

16. Heijden MS van der, Powles T, Petrylak D, Wit R de, Necchi A, Sternberg CN, Matsubara N, Nishiyama H, Castellano D, Hussain SA, et al. Predictive biomarkers for survival benefit with ramucirumab in urothelial cancer in the RANGE trial. *Nature communications* (2022) 13: doi: [10.1038/S41467-022-29441-Y](https://doi.org/10.1038/S41467-022-29441-Y)

17. Smith T, Lane B, More, E, Valentine H, Lunj S, Irlam-Jones J, Reeves K, Hoskin P, Choudhury A, West C. GSE203149: Gene expression data from muscle-invasive bladder cancer samples. (2022) <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE203149> [Accessed July 1, 2025]

18. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, Vang S, Reinert T, Hermann GG, Mogensen K, et al. Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* (2016) 30:27–42. doi: [10.1016/j.ccell.2016.05.004](https://doi.org/10.1016/j.ccell.2016.05.004)

19. Groeneveld CS, Sanchez-Quiles V, Dufour F, Shi M, Dingli F, Nicolle R, Chapeaublanc E, Poullet P, Jeffery D, Krucker C, et al. Proteogenomic Characterization of Bladder Cancer Reveals Sensitivity to Apoptosis Induced by Tumor Necrosis Factor–related Apoptosis-inducing Ligand in FGFR3-mutated Tumors. *European Urology* (2024) 85:483–494. doi: [10.1016/J.EURURO.2023.05.037](https://doi.org/10.1016/J.EURURO.2023.05.037)

20. Damrauer JS, Beckabir W, Klomp J, Zhou M, Plimack ER, Galsky MD, Grivas P, Hahn NM, O’Donnell PH, Iyer G, et al. Collaborative study from the Bladder Cancer Advocacy Network for the genomic analysis of metastatic urothelial cancer. *Nature communications* (2022) 13: doi: [10.1038/S41467-022-33980-9](https://doi.org/10.1038/S41467-022-33980-9)

21. Dressler FF, Diedrichs F, Sabtan D, Hinrichs S, Krisp C, Gemoll T, Hennig M, Mackedanz P, Schlotfeldt M, Voß H, et al. Proteomic analysis of the urothelial cancer landscape. *Nature Communications* (2024) 15: doi: [10.1038/S41467-024-48096-5](https://doi.org/10.1038/S41467-024-48096-5)

22. Stroggilos R, Mokou M, Latosinska A, Makridakis M, Lygirou V, Mavrogeorgis E, Drekolias D, Frantzi M, Mullen W, Fragkoulis C, et al. Proteome-based classification of Nonmuscle Invasive Bladder Cancer. *International journal of cancer* (2020) 146:281–294. doi: [10.1002/IJC.32556](https://doi.org/10.1002/IJC.32556)

23. DepMap. Current DepMap Release data, including CRISPR Screens, PRISM Drug Screens, Copy Number, Mutation, Expression, and Fusions. *Figshare+* (2024) Broad: doi: [10.25452/figshare.plus.25880521.v1](https://doi.org/10.25452/figshare.plus.25880521.v1)

24. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: Imputation for microarray data. (2023) <https://bioconductor.org/packages/release/bioc/html/impute.html>

25. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882. doi: [10.1093/BIOINFORMATICS/BTS034](https://doi.org/10.1093/BIOINFORMATICS/BTS034)

26. Ramaker RC, Lasseigne BN, Hardigan AA, Palacio L, Gunther DS, Myers RM, Cooper SJ. RNA sequencing-based cell proliferation analysis across 19 cancers identifies a subset of proliferation-informative cancers with a common survival signature. *Oncotarget* (2017) 8:38668. doi: [10.18632/ONCOTARGET.16961](https://doi.org/10.18632/ONCOTARGET.16961)

27. Li X, Luo S, Fu W, Huang M, Huang X, Kang S, Zhang J, Wang Q, Song C. Discovery of a proliferation essential gene signature and actin-like 6A as potential biomarkers for predicting prognosis and neoadjuvant chemotherapy response in triple-positive breast cancer. *Cancer* (2024) 130:1435–1448. doi: [10.1002/CNCR.35228](https://doi.org/10.1002/CNCR.35228)

28. Altorki NK, Bhinder B, Borczuk AC, Elemento O, Mittal V, McGraw TE. A signature of enhanced proliferation associated with response and survival to anti-PD-L1 therapy in early-stage non-small cell lung cancer. *Cell Reports Medicine* (2024) 5:101438. doi: [10.1016/J.XCRM.2024.101438](https://doi.org/10.1016/J.XCRM.2024.101438)

29. Locard-Paulet M, Palasca O, Jensen LJ. Identifying the genes impacted by cell proliferation in proteomics and transcriptomics studies. *PLOS Computational Biology* (2022) 18:e1010604. doi: [10.1371/JOURNAL.PCBI.1010604](https://doi.org/10.1371/JOURNAL.PCBI.1010604)

30. Pang KL, Li P, Yao XR, Xiao WT, Ren X, He JY. Deciphering a proliferation-essential gene signature based on CRISPR-Cas9 screening to predict prognosis and characterize the immune microenvironment in HNSCC. *BMC Cancer* (2025) 25:1–20. doi: [10.1186/S12885-025-14181-1/FIGURES/14](https://doi.org/10.1186/S12885-025-14181-1/FIGURES/14)

31. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Computational Biology* (2011) 7: doi: [10.1371/journal.pcbi.1002240](https://doi.org/10.1371/journal.pcbi.1002240)

32. Ingebriktsen LM, Finne K, Akslen LA, Wik E. A novel age-related gene expression signature associates with proliferation and disease progression in breast cancer. *British Journal of Cancer 2022 127:10* (2022) 127:1865–1875. doi: [10.1038/s41416-022-01953-w](https://doi.org/10.1038/s41416-022-01953-w)

33. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* (2013) 14:7. doi: [10.1186/1471-2105-14-7](https://doi.org/10.1186/1471-2105-14-7)

34. Aran D, Hu Z, Butte AJ. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* (2017) 18:220. doi: [10.1186/s13059-017-1349-1](https://doi.org/10.1186/s13059-017-1349-1)

35. Finotello F, Mayer C, Plattner C, Laschober G, Rieder Di, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine* (2019) 11:34. doi: [10.1186/s13073-019-0638-6](https://doi.org/10.1186/s13073-019-0638-6)

36. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research* (2013) 41:D955–D961. doi: [10.1093/NAR/GKS1111](https://doi.org/10.1093/NAR/GKS1111)

37. Seashore-Ludlow B, Rees MG, Cheah JH, Coko M, Price EV, Coletti ME, Jones V, Bodycombe NE, Soule CK, Gould J, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer discovery* (2015) 5:1210–1223. doi: [10.1158/2159-8290.CD-15-0235](https://doi.org/10.1158/2159-8290.CD-15-0235)

38. Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, Humeidi R, Peck D, Wu X, Tang AA, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer 2020 1:2* (2020) 1:235–248. doi: [10.1038/s43018-019-0018-6](https://doi.org/10.1038/s43018-019-0018-6)

39. Geeleher P, Cox N, Stephanie Huang R. pRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLOS ONE* (2014) 9:e107468. doi: [10.1371/JOURNAL.PONE.0107468](https://doi.org/10.1371/JOURNAL.PONE.0107468)

40. Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-i-Mompel P, Fallegger R, Türei D, Lægreid A, Saez-Rodriguez J. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic acids research* (2023) 51:10934–10949. doi: [10.1093/NAR/GKAD841](https://doi.org/10.1093/NAR/GKAD841)

41. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, Garnett MJ, Blüthgen N, Saez-Rodriguez J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications 2017 9:1* (2018) 9:1–11. doi: [10.1038/s41467-017-02391-6](https://doi.org/10.1038/s41467-017-02391-6)

42. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* (2022) 2: doi: [10.1093/BIOADV/VBAC016](https://doi.org/10.1093/BIOADV/VBAC016)

43. 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)

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

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

46. 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)

47. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health* (2019) 22:153–160. doi: [10.1136/ebmental-2019-300117](https://doi.org/10.1136/ebmental-2019-300117)

48. Kohonen T. *Self-Organizing Maps*. Berlin, Heidelberg: Springer Berlin Heidelberg (1995). doi: [10.1007/978-3-642-97610-0](https://doi.org/10.1007/978-3-642-97610-0)

49. Wehrens R, Kruisselbrink J. Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software* (2018) 87:1–18. doi: [10.18637/jss.v087.i07](https://doi.org/10.18637/jss.v087.i07)

50. Raymaekers J, Zamar RH. Regularized K-means Through Hard-Thresholding. *Journal of Machine Learning Research* (2022) 23:1–48. <http://jmlr.org/papers/v23/21-0052.html>

51. Vesanto J, Alhoniemi E. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks* (2000) 11:586–600. doi: [10.1109/72.846731](https://doi.org/10.1109/72.846731)

52. Venna J, Kaski S. Neighborhood preservation in nonlinear projection methods: An experimental study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2001) 2130:485–491. doi: [10.1007/3-540-44668-0\_68](https://doi.org/10.1007/3-540-44668-0_68)

53. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* (1987) 20:53–65. doi: [10.1016/0377-0427(87)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

54. Lange T, Roth V, Braun ML, Buhmann JM. Stability-based validation of clustering solutions. *Neural Computation* (2004) 16:1299–1323. doi: [10.1162/089976604773717621](https://doi.org/10.1162/089976604773717621)

55. Leng M, Wang J, Cheng J, Zhou H, Chen X. Adaptive semi-supervised clustering algorithm with label propagation. *Journal of Software Engineering* (2014) 8:14–22. doi: [10.3923/jse.2014.14.22](https://doi.org/10.3923/jse.2014.14.22)

56. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018) <https://arxiv.org/abs/1802.03426v3 http://arxiv.org/abs/1802.03426>

57. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen’s d, and r. *Law and Human Behavior* (2005) 29:615–620. doi: [10.1007/S10979-005-6832-7](https://doi.org/10.1007/S10979-005-6832-7)

58. Pichler R, Creij NCH van, Mertens LS, Giudice F del, Koll F, Soria F, Subiela JD, Plage H, Tymoszuk P, Mayr R, et al. FGFR1/3 Signaling as an Achilles’ Heel of Phenotypic Diversity in Urothelial Carcinoma. *European urology oncology* (2025) doi: [10.1016/J.EUO.2025.07.005](https://doi.org/10.1016/J.EUO.2025.07.005)

59. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* (2005) 4: doi: [10.2202/1544-6115.1128/MACHINEREADABLECITATION/RIS](https://doi.org/10.2202/1544-6115.1128/MACHINEREADABLECITATION/RIS)

60. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* (2006) Complex Sy:1695. <https://igraph.org>

61. Traag VA, Waltman L, Eck NJ van. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports 2019 9:1* (2019) 9:1–12. doi: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z)

62. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, et al. A community-driven global reconstruction of human metabolism. *Nature Biotechnology 2013 31:5* (2013) 31:419–425. doi: [10.1038/nbt.2488](https://doi.org/10.1038/nbt.2488)

63. Gavai AK, Supandi F, Hettling H, Murrell P, Leunissen JAM, Van Beek JHGM. Using Bioconductor Package BiGGR for Metabolic Flux Estimation Based on Gene Expression Changes in Brain. *PLOS ONE* (2015) 10:e0119016. doi: [10.1371/JOURNAL.PONE.0119016](https://doi.org/10.1371/JOURNAL.PONE.0119016)

64. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* (2016) 44:D515–D522. doi: [10.1093/NAR/GKV1049](https://doi.org/10.1093/NAR/GKV1049)

65. Dempster JM, Pacini C, Pantel S, Behan FM, Green T, Krill-Burger J, Beaver CM, Younger ST, Zhivich V, Najgebauer H, et al. Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nature Communications* (2019) 10:5817. doi: [10.1038/S41467-019-13805-Y](https://doi.org/10.1038/S41467-019-13805-Y)

66. Dempster JM, Boyle I, Vazquez F, Root DE, Boehm JS, Hahn WC, Tsherniak A, McFarland JM. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biology* (2021) 22:1–23. doi: [10.1186/S13059-021-02540-7/FIGURES/6](https://doi.org/10.1186/S13059-021-02540-7/FIGURES/6)

67. Canty A, Ripley B, Brazzale A. boot: Bootstrap Functions. (2024) doi: [10.32614/CRAN.PACKAGE.BOOT](https://doi.org/10.32614/CRAN.PACKAGE.BOOT)