

SCIENTIFIC REPORTS



OPEN

Urothelial cancer proteomics provides both prognostic and functional information

Guillermo de Velasco¹, Lucia Trilla-Fuertes^{2,3}, Angelo Gamez-Pozo^{2,3}, Maria Urbanowicz⁴, Gustavo Ruiz-Ares¹, Juan M. Sepúlveda¹, Guillermo Prado-Vazquez², Jorge M. Arevalillo⁵, Andrea Zapater-Moros², Hilario Navarro⁵, Rocio Lopez-Vacas², Ray Manneh¹, Irene Otero¹, Felipe Villacampa^{6,8}, Jesus M. Paramio^{7,8}, Juan Angel Fresno Vara^{2,3,8} & Daniel Castellano^{1,8}

Traditionally, bladder cancer has been classified based on histology features. Recently, some works have proposed a molecular classification of invasive bladder tumors. To determine whether proteomics can define molecular subtypes of muscle invasive urothelial cancer (MIUC) and allow evaluating the status of biological processes and its clinical value. 58 MIUC patients who underwent curative surgical resection at our institution between 2006 and 2012 were included. Proteome was evaluated by high-throughput proteomics in routinely archive FFPE tumor tissue. New molecular subgroups were defined. Functional structure and individual proteins prognostic value were evaluated and correlated with clinicopathologic parameters. 1,453 proteins were quantified, leading to two MIUC molecular subgroups. A protein-based functional structure was defined, including several nodes with specific biological activity. The functional structure showed differences between subtypes in metabolism, focal adhesion, RNA and splicing nodes. Focal adhesion node has prognostic value in the whole population. A 6-protein prognostic signature, associated with higher risk of relapse (5 year DFS 70% versus 20%) was defined. Additionally, we identified two MIUC subtypes groups. Prognostic information provided by pathologic characteristics is not enough to understand MIUC behavior. Proteomics analysis may enhance our understanding of prognostic and classification. These findings can lead to improving diagnosis and treatment selection in these patients.

Urothelial cancer (UC) is responsible for approximately 165,000 deaths per year worldwide (GLOBOCAN 2012)¹. Pathological classification divides UC into two major subtypes according to the invasion depth: non-muscle invasive and muscle invasive urothelial carcinoma (MIUC) but not molecular categorization is clinically indicated. However, the outcome and prognosis may be different across subsets of patients within same staging.

MIUC is characterized by a high risk of relapse and metastasis. Despite radical cystectomy with neoadjuvant cisplatin-based chemotherapy, the current risk of recurrence as well as mortality is nearly 50%². In the adjuvant setting, chemotherapy is also associated with improved survival in patients with locally advanced bladder cancer³.

Pathological prognostic factors such as lymphovascular invasion, grade or molecular alterations are not currently modifying treatment choice. Large collaborative efforts have provided a more comprehensive view of the genomic landscape of MIUC identifying molecular subtypes that have yet to prove predictive value³⁻⁵. At present, no molecularly targeted drugs are approved for UC.

Before the genomic era, p53 was thought to be prognostic and predictive marker measured by immunohistochemistry in UC⁶. Several methodological issues questioned conflicting results including proteomics assessment⁷.

¹Department of Medical Oncology, University Hospital 12 de Octubre, i + 12, Madrid, Spain. ²Molecular Oncology & Pathology Lab, INGEMM, Instituto de Investigación Hospital La Paz-IdiPAZ, Madrid, Spain. ³Biomedica Molecular Medicina, Madrid, Spain. ⁴Department of Pathology, University Hospital 12 de Octubre, Madrid, Spain. ⁵Department of Statistics, Operational Research and Numerical Analysis, University Nacional Educacion a Distancia (UNED), Madrid, Spain. ⁶Department of Urology, University Hospital 12 de Octubre, Madrid, Spain. ⁷Molecular and Cell Oncology Group, Biomedical research Institute, University Hospital 12 de Octubre, i + 12, and Molecular Oncology Unit, CIEMAT, Madrid, Spain. ⁸CIBERONC, Madrid, Spain. Guillermo de Velasco, Lucia Trilla-Fuertes, Juan Angel Fresno Vara and Daniel Castellano contributed equally to this work. Correspondence and requests for materials should be addressed to G.V. (email: gvelasco.gdv@gmail.com)

Urothelial tumors	
Number of patients	58
Age (years)	
≤60	20(34,5%)
>60	38(65,5%)
Median(IQR)	68(60–71)
Range	45–78
Sex	
Male	51(88%)
Female	7(12%)
pT category	
pT2a	2(3.5%)
pT2b	10(17.3%)
pT3a	27(46.5%)
pT3b	8(13.8%)
pT4a	9(15.5%)
pT4b	1(1.7%)
Missing	1(1.7%)
pN category	
pN0	32(55%)
pN1	14(24%)
pN2	6(10%)
Missing	6(10%)
Highest G grade	
G1-2	8(14%)
G3	44(76%)

Table 1. Study population.

In the last years, proteomics approaches have been incorporated into the study of clinical samples, as a way to complement the information provided by classical factors and genomics. Mass spectrometry-based proteomics have emerged as preferred components of a strategy for discovering diagnostic and prognostic protein biomarkers and as well as new therapeutic targets⁸. These investigations are very encouraging^{9,10} and the potential of tumor biomarkers discovery is unclear¹¹.

Genomics advance in UC has not been translated into molecularly-based biomarker for treatment selection. Since few data is available with proteomics, we aimed to identify whether differentially expressed protein biomarkers in tumor tissue may predict different outcomes.

Results

Study Population. Fifty eight patients with a median age of 68 years (range 45–78 years) were included. Main characteristics are displayed in Table 1. After a median follow up of 38 months, 34 (58.6%) patients relapsed and 35 (60.4%) had died. Median follow-up of all patients was 34 months (range 3–114 months). Median distant disease free survival was 27.7 (27.2–45.1, 95%CI). Five- years-distant relapse free survival was: 75% in stage I/II, 45% in stage III and 25% in stage IV.

Protein preparation and mass spectrometry analysis. After mass spectrometry (MS) workflow, 58 urothelial tumors were analyzed. Raw data normalization was performed as described previously¹². 4,405 protein groups were identified using Andromeda, of which 1,453 presented at least two unique peptides and detectable expression in at least 75% of the samples. No decoy protein passed through these additional filters.

Protein expression analyses of urothelial tumors and identification of new molecular subtypes. Proteomics data from 58 MIUC tumors were analyzed using sparse k-means and random-forest in order to establish a consistent classification of our samples. Using these approaches, two different molecular groups were identified on the basis of 34 proteins differentially expressed between both groups (Supplementary Figure 1, Supplementary Table 1). From those, 20 proteins have higher expression in group 1, including EHD2, FLNA and TNS1. Gene ontology analyses showed that these proteins are mainly related with focal adhesion and extracellular matrix. On the other hand, 14 proteins have higher expression in group 2, including HSBP1. Gene ontology analyses showed that these proteins are mainly related with transcription processes and immune response. Group 1 showed better prognosis than Group 2, although these differences were not significant (Fig. 1). Contingency analyses showed that these two groups are independent of clinical factors such as stage, tumor size and lymph node status.

Network construction and functional node assignment. Protein expression data from all samples were used in the probabilistic graphical models analyses, with no other *a priori* information. The resulting graph was processed (Fig. 2) looking for a functional structure, i.e., if the proteins included in each branch of the tree

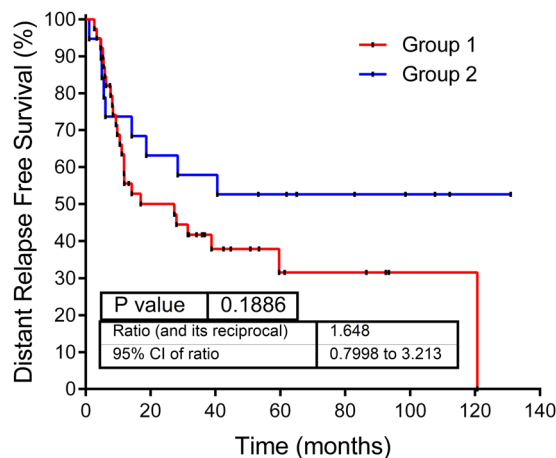


Figure 1. Kaplan–Meier survival curves obtained from high/low risk groups originated in our classification.

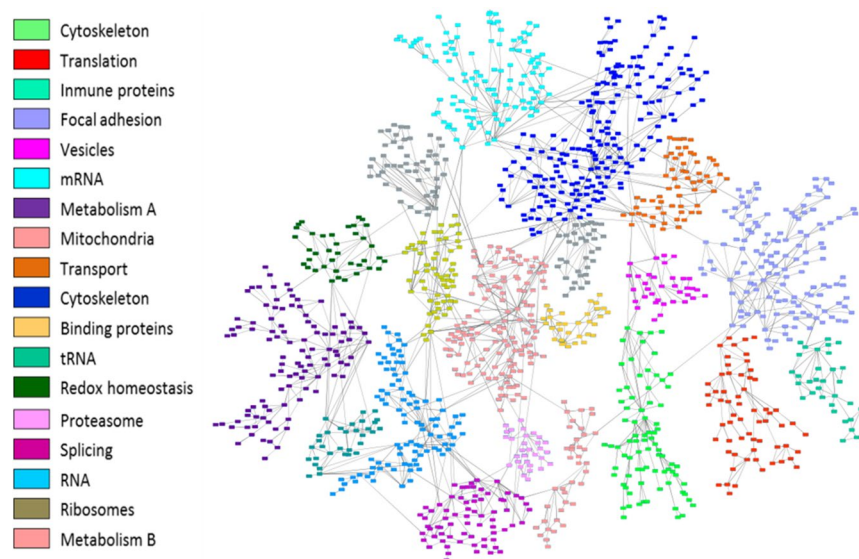


Figure 2. Probabilistic graphical model analysis unravels the functional organization of proteins in MIUC based on correlation. Grey nodes are nodes without any majority function assigned.

had some relationship regarding their function, as previously described¹². Thus, we divided our graph into eighteen branches, and performed gene ontology analyses. The structure of the probabilistic graphical model had a strong biological function basis. The next step was to calculate the activity for each branch with a specific biological function, i.e., a functional node, as previously described¹² (Supplementary Figure 2). Once calculated, we evaluated the prognostic value of each functional node activity in MIUC. Focal adhesion functional node activity splits the population into two groups with different prognosis ($p = 0.0241$, $HR = 0.44$ $IC_{95} = 0.234$ to 0.899) (Fig. 3). Afterwards, we assessed the differences in the functional nodes activities between Group 1 and Group 2 using class comparison analyses. Twelve nodes showed significant different activity between both groups. Focal adhesion, two cytoskeleton nodes, tRNA, ribosomes and metabolism A & B functional nodes showed increased activity in Group 1 tumors, whereas vesicles, transport, proteasome, RNA and splicing nodes showed increased activity in Group 2 tumors (Supplementary Figure 2).

Focal adhesion functional node. Focal adhesion functional node includes twenty six proteins related with extracellular matrix and focal adhesion. COL1A1, SOD3, COL6A1, COL6A2, CAPN2, MSN, STOM, PRELP, NID2, DAG1, LPP and GPI are highly expressed in group 1 while SFN and HDLBP are highly expressed in group 2 ($p < 0.05$). Overall, functional activity of this node is higher in group 1. In addition, this functional node has prognostic value in our cohort.

Development of a prognostic protein signature in MIUC. 66 proteins were found to be associated with recurrence risk in MIUC (Supplementary Table 2). A recurrence signature was developed as previously described¹³. Six proteins of these 66 were included in the prognostic signature: ANXA1 (Annexin A1), BGN (Biglycan), IGFBP7 (Insulin Like Growth Factor Binding Protein 7), ISLR (Immunoglobulin Superfamily

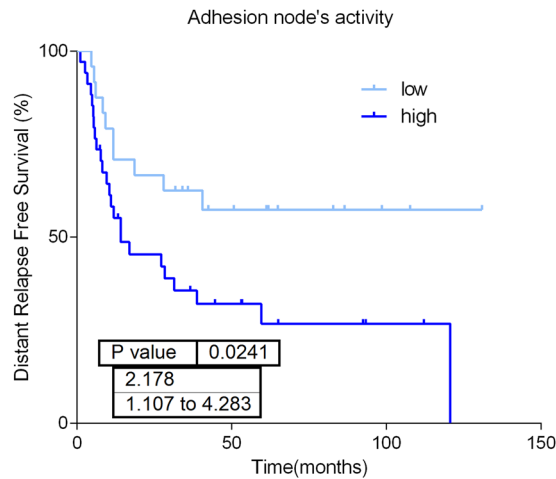


Figure 3. Focal adhesion node's activity has prognostic value (p-value = 0.0241, HR = 2.178, IC95 = 1.107 to 4.283).

Containing Leucine-Rich Repeat), MDP1 (Magnesium-Dependent Phosphatase 1) and PLS3 (Plastin 3). The recurrence protein score split our population into two risk groups with different five year distant relapse free survival: 70% vs. 20% (HR 3.53 95% CI 1.8–6.7; [$p < 0.001$]) (Fig. 4). The association between the score and DRFS was similar for patients with stage III and those with stage IV (Fig. 4). These results were verified using gene expression data from the MD Anderson cohort⁴. In this population, the 6 proteins predictor identifies two populations with different relapse risk (HR 2.10 95% CI 0.2–1.1; [$p = 0.04$]) (Supplementary Figure 3).

Functional proteomics add prognostic information to prognostic signature. Information provided by the prognostic signature is complementary to the prognostic information provided by focal adhesion functional node activity signature and both predictors combined establish four different classes into the population with different relapse risk (p-value = 0.0003) (Fig. 5). Univariate analyses of clinical (stage, tumor size and nodal involvement) and proteomics-based variables (6 protein signature and focal adhesion node activity) showed that focal adhesion node activity (p-value = 0.013), 6 proteins signature (p-value = 0.002) and tumor size (p-value = 0.033) have prognostic value in MIUC. Multivariate analyses showed that both focal adhesion (p-value = 0.011) and 6 proteins signature (p-value = 0.020) has independent prognostic value (Table 2).

Discussion

The principal aim of the study was to establish a molecular classification and survival prediction in MIUC based on proteome analysis since bladder cancer classifications have generally been based on histology features¹⁴ and genomics have yet to be implemented in the clinic. In the clinical practice, there seems to be different groups of patients beyond pathological characteristics. Some patients, even with positive lymph nodes, may never relapse after surgery. However, other subset of patients with apparent favorable features may become metastatic. Therefore it is clear that stratification using the current system is inadequate to satisfactorily differentiate prognosis. Consequently, it is necessary to characterize MIUC patients in accordance to prognostic evolution and molecular features. The proposal of new classifications and characterization of MIUC could lead to further stratification of MIUC tumors and may drive treatment selection.

Our proteomics pipeline allowed us detecting 4,405 proteins in 58 FFPE MIUC samples. We identified groups in our protein data using sparse k-means and confirmed its consistency by random forest, supporting that different molecular subgroups exist in MIUC. Sparse k-means classification is based on 34 proteins, most of them related with focal adhesion. These two molecular groups provide additional information to clinical parameters. As we demonstrated in previous works, probabilistic graphical models using protein expression data allow characterizing differences in biological processes and pathways between groups of patients^{12,15}. We were able to establish different functional nodes according to biological functions. The analysis identified 18 different functional nodes, 12 of them monitoring eleven biological processes showed differential activity between the prognostic groups previously established. These results confirm that this approach is valid to study the differential activity of biological functions between tumor groups¹².

Group 1 showed higher expression of some proteins related with focal adhesion and extracellular matrix. Specifically, some of these proteins have been related with epithelial-to-mesenchymal transition (EMT) markers such as EH domain containing 2 (EHD2), which can inhibit metastasis by regulating cadherins¹⁶ or tensin 1 (TNS1), involved in focal adhesion. Low levels of TNS1 have been associated with worsening-free survival in non-muscle invasive bladder cancer¹⁷. Additionally, filamin A, a downstream effector of mTORC2, plays an important role in motility and invasion¹⁸. Additionally, we showed an increased activity of biological processes in Group 1, such as Cytoskeleton and Focal Adhesion, Metabolism and tRNA and ribosomes. It is noteworthy that related nodes, such as Cytoskeleton and Focal Adhesion nodes, and also tRNA and ribosomes nodes showed a similar behaviour, showing consistency for obtained biological information. Metabolism A node includes proteins related with negative regulation of protein metabolic process whereas Metabolism B node included proteins

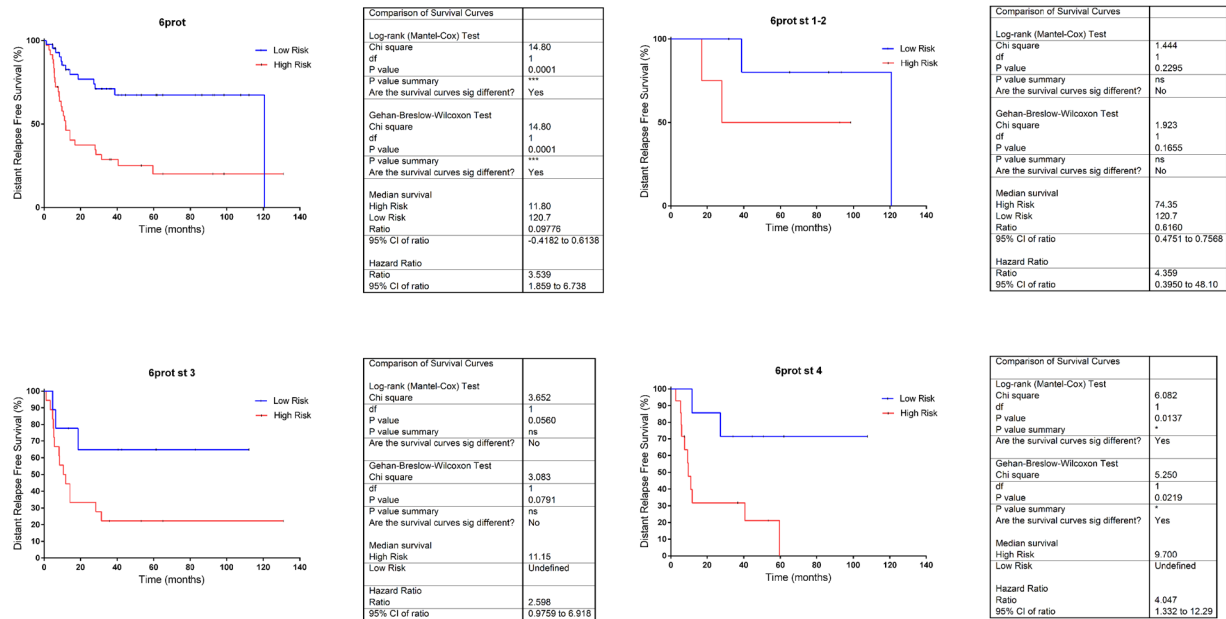


Figure 4. Prognostic signature composed by 6 proteins. A. All data. B. Stage 1–2. C. Stage 3 D. Stage 4.

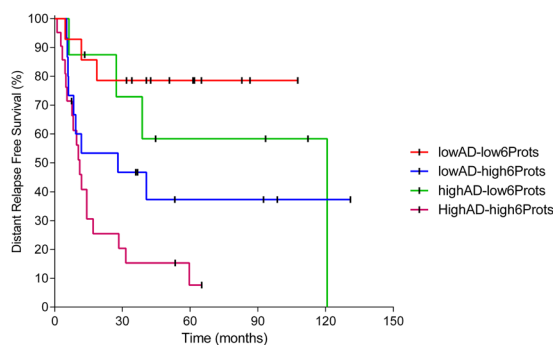


Figure 5. Kaplan–Meier curve curves showing overall survival based on 6 protein signature merged with focal adhesion node activity signature (p-value = 0.0003).

Multivariate analysis				
	Sig.	Exp(B)	95.0% IC para Exp(B)	
			Inferior	Superior
6prots	0.020	3.486	1.217	9.981
Adhesion Node	0.011	3.029	1.287	7.130
Stage	0.840	1.086	0.489	2.412
Size	0.452	0.910	0.711	1.164
N	0.747	1.150	0.492	2.687

Table 2. Multivariate analysis.

related to glycolysis and pyruvate metabolism, involved in generation of precursor metabolites and energy. All together, these results suggest that Group 1 have lower metastatic potential and specific features regarding metabolism and protein synthesis when compared with Group 2.

On the other hand, several proteins showed higher expression in group 2. Some immune proteins such as HSBP1 (heat shock factor binding protein 1) were associated with a decreased immune activity which may have therapeutic implications¹⁹. Additionally, group 2 showed increased activity in Vesicles, Transport, Proteasome, Splicing and RNA nodes. Again, we found coherence in the biological information, as long as nodes with comparable function showed similar behavior. These results suggest differences regarding intracellular trafficking, RNA processing and Proteasome activities when comparing new defined groups.

The differences in biological functions, after proper validation, could lead to develop specific treatments in concrete groups of patients. For instance, differences in metabolism could be targeted with 2-D-deoxy-glucose or metformin²⁰, which are being currently tested in clinical trials for breast cancer treatment. On the other hand, proteasome targeting drugs have demonstrated therapeutic value in multiple myeloma treatment²¹.

In this study we show that the discovery of proteins as prognostic biomarkers is feasible using FFPE samples and proteomics. Indeed, we were able to identify a six protein-signature with prognosis value independently of stage, size and lymph node status. Proteins contained in this predictor are involved in multiple processes. ANXA1, a membrane-located protein, has been related with prognosis in breast cancer²². BGN is a protein involved in inflammation processes²³. Niedworok *et al.*²⁴ suggested that biglycan is an endogenous inhibitor of bladder cancer cell proliferation and its high expression is associated with good prognosis. PLS3 was proposed as biomarker for breast cancer prognosis²⁵. To our knowledge, no previous information about MDP1, IGFBP7 and ISLR role in cancer disease has been previously reported. The prognostic value of the 6-protein signature was validated using gene expression data from another cohort.

Probabilistic graphical models allow to compare biological functions between groups but also, to build prognostic signatures. Focal adhesion functional node activity had prognostic value and split population in low and high risk of relapse. Strikingly, prognostic information provided by a traditional protein signature was complementary to information provided by focal adhesion functional node activity signature, and also to the prognostic information provided by clinical factors, as shown in the multivariate analysis. Merging these molecular features, it is feasible to establish four different risk populations. These results confirm that functional approaches could provide additional information to traditional gene/protein-centered analyses.

Our study has some limitations. Technically, proteomics provide less information when compared with genomics, thus an improvement in number of detected proteins is still necessary. On the other hand, peptide and protein identification relies in statistical parameters. Due to this, we applied strict filters for peptides/proteins selection, in order to avoid false detections, ensuring that proteins with the highest confidence in both identification and quantification are selected for analyses. Finally, although a meta-validation has been performed, these results should be validated in additional cohorts to evaluate the 6-protein signature robustness and the functional differences between new defined molecular groups. Other limitations of this study include the relatively small sample size and there may be other bias that could affect outcomes. We believe that our findings serve as important hypothesis generating findings that can be explored in future studies.

In conclusion, our approach, combining proteomics and probabilistic graphical models allow the integration of different levels of molecular information that can improve MIUC molecular characterization. We were able to differentiate two different molecular groups from our proteomics data, with different functional features that may represent new therapeutic opportunities for bladder cancer treatment. Moreover, we defined a 6 protein-signature that can predict the outcome of MIUC patients and we identified a functional node with prognosis value in MIUC, adding prognostic information to the prognostic 6-protein signature and to clinical factors.

Methods

Patient's characteristics and samples selection. Patients treated at University Hospital 12 de Octubre (Madrid, Spain) were included if they had histologically documented (TNM staging²⁶, T1-T4a and any N, M0) urothelial carcinoma (including of the renal pelvis, ureter, urinary bladder, or urethra). In total, 58 patients who underwent curative surgical resection between 2006 and 2012 were selected. FFPE samples were retrieved from I + 12 Biobank (RD09/0076/00118). Samples were reviewed by a genitourinary pathologist and included if cases had at least 50% of urothelial tumor cells and were invasive in the muscularis propria. The study was approved by independent review board and Ethical Committee of Hospital Universitario 12 de Octubre. All experiments were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants before starting treatment.

Liquid chromatography - mass spectrometry shotgun analysis. Proteins were extracted from FFPE samples as previously described²⁷. Mass spectrometry analysis was performed on a QExactive mass spectrometer coupled to a nano EasyLC 1000 (Thermo Fisher Scientific). Solvent composition at the two channels was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B. For each sample 2 μ L of peptides were loaded on a self-made column (75 μ m \times 150 mm) packed with reverse-phase C18 material (ReproSil-Pur 120 C18-AQ, 1.9 μ m, Dr. Maisch GmbH) and eluted at a flow rate of 300 nL/min by a gradient from 2 to 35% B in 80 min, 47% B in 4 min and 98% B in 4 min. Samples were acquired in a randomized order. The mass spectrometer was operated in data-dependent mode (DDA), acquiring a full-scan MS spectra (300–1700 m/z) at a resolution of 70000 at 200 m/z after accumulation to a target value of 3000000, followed by HCD (higher-energy collision dissociation) fragmentation on the twelve most intense signals per cycle. HCD spectra were acquired at a resolution of 35000 using normalized collision energy of 25 and a maximum injection time of 120 ms. The automatic gain control (AGC) was set to 50000 ions. Charge state screening was enabled and singly and unassigned charge states were rejected. Only precursors with intensity above 8300 were selected for MS/MS (2% underfill ratio). Precursor masses previously selected for MS/MS measurement were excluded from further selection for 30 s, and the exclusion window was set at 10 ppm. The samples were acquired using internal lock mass calibration on m/z 371.1010 and 445.1200.

Protein identification and label free protein quantification. The acquired raw MS data were processed by MaxQuant (version 1.5.2.8), followed by protein identification using the integrated Andromeda search engine. Spectra were searched against a forward Swiss Prot-human database, concatenated to a reversed decoyed fasta database and common protein contaminants (NCBI taxonomy ID9606, release date 2014-05-06). Carbamidomethylation of cysteine was set as fixed modification, while methionine oxidation and N-terminal

protein acetylation were set as variable. Enzyme specificity was set to trypsin/P allowing a minimal peptide length of 7 amino acids and a maximum of two missed-cleavages. Precursor and fragment tolerance was set to 10 ppm and 20 ppm, respectively for the initial search. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins. Label free quantification was enabled and a 2 minutes window for match between runs was applied. The re-quantify option was selected. For protein abundance the intensity was used, corresponding to the sum of the precursor intensities of all identified peptides for the respective protein group.

Sparse k-means classification. Sparse k-means was used to establish differential groups between samples. Classification consistency was tested using random forest. An analysis with the Consensus Clustering algorithm²⁸, applied on the data containing the variables that were selected by the sparse K-means method²⁹, has provided an optimum classification into two subtypes in previous studies³⁰.

Functional network construction. Network construction was performed using probabilistic graphical models compatible with high dimensional data using correlation as associative method as previously described¹². In order to identify functional nodes in the networks we split them in several branches and we used Gene Ontology analysis to assign a majority function to each node. Activity measurement was calculated by the mean expression of all the proteins of each node related with the assigned node function.

Gene-Ontology Analysis. Protein to Gene Symbol conversion was performed using Uniprot and DAVID³¹. Gene Ontology Analysis was also done in DAVID selecting “*Homo sapiens*” background and GOTERM-FAT, Biocarta, KEGG and Panther databases.

Protein signature construction. We computed a statistical significance level for each protein based on a univariate proportional hazards model with the aim of identifying proteins whose expression were significantly related to the distant metastasis-free survival (DMFS) as described previously¹³. Leave-one-out cross-validation was used to evaluate the predictive accuracy of the profiles. The cutoff point was established *a priori* and to test the statistical significance, the p-value of the log-rank test statistic for the risk groups was evaluated using 1000 random permutations. Analyses were performed in BRB-ArrayTools v4_2_1 and R v3.2.4³². BRB-ArrayTools has been developed by Dr. Richard Simon and BRB-ArrayTools Development Team.

Prognostic signature meta-validation. With the aim to verify the utility of 6 protein signature, gene expression data from a MD Anderson cohort was used⁴. All probes in dataset for each gene were retrieved. Probes with higher CV were selected when multiple probes were found for a single gene, then expression values of each gene were z-score transformed as previously described¹⁵. To apply protein expression based signatures to gene expression values, per-gene normalization was applied as previously described¹³.

Statistical analyses. Statistical analyses (class comparisons contingency analyses, etc.), were performed using GraphPad Prism v6 and Cytoscape³³. Univariate and multivariate Cox regression models were performed using IBM SPSS Statistics. All p-values where two-sided, and $p < 0.05$ was considered statistically significant.

References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359–386, <https://doi.org/10.1002/ijc.29210> (2015).
2. Collaboration, A. B. C. A. M.-a. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur Urol* **48**, 202–205; discussion 205–206, <https://doi.org/10.1016/j.eururo.2005.04.006> (2005).
3. Penson, D. F. Re: Effectiveness of Adjuvant Chemotherapy for Locally Advanced Bladder Cancer. *J Urol* **196**, 352–354, <https://doi.org/10.1016/j.juro.2016.05.032> (2016).
4. 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, <https://doi.org/10.1016/j.ccr.2014.01.009> (2014).
5. Kim, J. *et al.* Invasive Bladder Cancer: Genomic Insights and Therapeutic Promise. *Clin Cancer Res* **21**, 4514–4524, <https://doi.org/10.1158/1078-0432.CCR-14-1215> (2015).
6. Stadler, W. *et al.* Randomized trial of p53 targeted adjuvant therapy for patients with organ-confined node-negative urothelial bladder cancer. *Journal of Clinical Oncology* **27**, 5017–5017 (2009).
7. George, B. *et al.* p53 gene and protein status: the role of p53 alterations in predicting outcome in patients with bladder cancer. *J Clin Oncol* **25**, 5352–5358, <https://doi.org/10.1200/JCO.2006.10.4125> (2007).
8. Hanash, S. Disease proteomics. *Nature* **422**, 226–232, <https://doi.org/10.1038/nature01514> (2003).
9. Marko-Varga, G. *et al.* Personalized medicine and proteomics: lessons from non-small cell lung cancer. *J Proteome Res* **6**, 2925–2935, <https://doi.org/10.1021/pr070046s> (2007).
10. Pastwa, E., Somiari, S. B., Czyz, M. & Somiari, R. I. Proteomics in human cancer research. *Proteomics Clin Appl* **1**, 4–17, <https://doi.org/10.1002/prca.200600369> (2007).
11. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24**, 971–983, <https://doi.org/10.1038/nbt1235> (2006).
12. Gámez-Pozo, A. *et al.* Vol. 75 2243–2253 (Cancer Res, 2015).
13. Sánchez-Navarro, I. *et al.* An 8-gene qRT-PCR-based gene expression score that has prognostic value in early breast cancer. *BMC Cancer* **10**, 336, <https://doi.org/10.1186/1471-2407-10-336> (2010).
14. Humphrey, P. A., Moch, H., Cubilla, A. L., Ulbright, T. M. & Reuter, V. E. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. *Eur Urol* **70**, 106–119, <https://doi.org/10.1016/j.eururo.2016.02.028> (2016).
15. Gámez-Pozo, A. *et al.* Functional proteomics outlines the complexity of breast cancer molecular subtypes. *Scientific Reports* **7**, 10100, <https://doi.org/10.1038/s41598-017-10493-w> (2017).
16. Shi, Y. *et al.* Decreased expression and prognostic role of EHD2 in human breast carcinoma: correlation with E-cadherin. *J Mol Histol* **46**, 221–231, <https://doi.org/10.1007/s10735-015-9614-7> (2015).
17. Fahmy, M. *et al.* Relevance of the mammalian target of rapamycin pathway in the prognosis of patients with high-risk non-muscle invasive bladder cancer. *Hum Pathol* **44**, 1766–1772, <https://doi.org/10.1016/j.humpath.2012.11.026> (2013).

18. Chantaravisoort, N., Wongkongkathep, P., Loo, J. A., Mischel, P. S. & Tamanoi, F. Significance of filamin A in mTORC2 function in glioblastoma. *Mol Cancer* **14**, 127, <https://doi.org/10.1186/s12943-015-0396-z> (2015).
19. Yashin, D. V. *et al.* The heat shock-binding protein (HspBP1) protects cells against the cytotoxic action of the Tag7-Hsp70 complex. *J Biol Chem* **286**, 10258–10264, <https://doi.org/10.1074/jbc.M110.163436> (2011).
20. *ClinicalTrials.gov A service of the U.S National Institutes of Health.*
21. Palumbo, A. *et al.* Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. *N Engl J Med* **375**, 754–766, <https://doi.org/10.1056/NEJMoa1606038> (2016).
22. Bhardwaj, A. *et al.* Annexin A1 Preferentially Predicts Poor Prognosis of Basal-Like Breast Cancer Patients by Activating mTOR-S6 Signaling. *PLoS One* **10**, e0127678, <https://doi.org/10.1371/journal.pone.0127678> (2015).
23. Hsieh, L. T., Nastase, M. V., Zeng-Brouwers, J., Iozzo, R. V. & Schaefer, L. Soluble biglycan as a biomarker of inflammatory renal diseases. *Int J Biochem Cell Biol* **54**, 223–235, <https://doi.org/10.1016/j.biocel.2014.07.020> (2014).
24. Niedworok, C. *et al.* Inhibitory role of the small leucine-rich proteoglycan biglycan in bladder cancer. *PLoS One* **8**, e80084, <https://doi.org/10.1371/journal.pone.0080084> (2013).
25. Ueo, H. *et al.* Circulating tumour cell-derived plastin3 is a novel marker for predicting long-term prognosis in patients with breast cancer. *Br J Cancer* **112**, 1519–1526, <https://doi.org/10.1038/bjc.2015.132> (2015).
26. Edge, S. *et al.* *Urinary bladder*. 7th edition, New York, Springer edn, 497-505 (AJCC Cancer Staging Manual, 2010).
27. Gámez-Pozo, A. *et al.* Shotgun proteomics of archival triple-negative breast cancer samples. *Proteomics Clin Appl* **7**, 283–291, <https://doi.org/10.1002/prca.201200048> (2013).
28. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine learning* **52**, 91–118 (2003).
29. Witten, D. M. & Tibshirani, R. A framework for feature selection in clustering. *J Am Stat Assoc* **105**, 713–726, <https://doi.org/10.1198/jasa.2010.tm09415> (2010).
30. Rousseeuw, P. J. Silhouettes. A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. *Journal of Computational and Applied Mathematics* **20**, 53–65 (1987).
31. Huang, d. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
32. (Vienna, Austria. R Foundation for Statistical Computing, 2013).
33. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504, <https://doi.org/10.1101/gr.1239303> (2003).

Acknowledgements

I + 12 Biobank (RD09/0076/00118) is integrated in the Spanish Hospital Biobank Network (RetBioH; www.redbiobancos.es). LT-F is supported by Spanish Economy and Competitiveness Ministry (DI-15-07614). This work was partially supported by a grant from the FMM (2012-0085) to GdV and by FEDER cofounded MINECO (SAF2015-66015-R) and ISCIII grants (RD12/0036/0009, PIE15/00076, and CB/16/00228 to JMP).

Author Contributions

Conception and design, obtaining funding: A.G.P., J.A.F.V., D.C., G.D.V.; acquisition of data: G.R.A., I.O., F.V., J.M.S., D.C., M.U., G.D.V.; drafting of the manuscript: L.T.F., A.G.P., G.D.V.; Experimental procedures: R.L.V., A.G.P., L.T.F., A.Z.M.; revision of the manuscript: all authors; statistical analysis: L.T.F., A.G.P.; G.P.V., J.M.A., H.N., G.P.V.; supervision: A.G.P., J.A.F.V., D.C., H.N.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-15920-6>.

Competing Interests: J.A.F.V. and A.G.-P. are shareholders in Biomedica Molecular Medicine S.L. L.T.-F. is an employee of Biomedica Molecular Medicine S.L. The other authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017