








The expression of *YWHAZ* and *NDRG1* predicts aggressive outcome in human prostate cancer

Sofia Lage-Vickers^{1,2,6}, Juan Bizzotto ^{1,2,6}, Maria Pia Valacco^{1,2}, Pablo Sanchis^{1,2}, Sergio Nemirovsky ^{1,2}, Estefania Labanca ³, Carlos Scorticati⁴, Osvaldo Mazza⁴, Antonina Mitrofanova⁵, Nora Navone³, Elba Vazquez ^{1,2}, Javier Cotignola^{1,2,7}  & Geraldine Gueron ^{1,2,7} 

Some prostate cancers (PCas) are histo-pathologically grouped within the same Gleason Grade (GG), but can differ significantly in outcome. Herein, we aimed at identifying molecular biomarkers that could improve risk prediction in PCa. LC ESI-MS/MS was performed on human PCa and benign prostatic hyperplasia (BPH) tissues and peptide data was integrated with *omic* analyses. We identified high *YWHAZ* and *NDRG1* expression to be associated with poor PCa prognosis considering all Gleason scores (GS). *YWHAZ* and *NDRG1* defined two subpopulations of PCa patients with high and intermediate risk of death. Multivariable analyses confirmed their independence from GS. ROC analysis unveiled that *YWHAZ* outperformed GS beyond 60 months post-diagnosis. The genomic analysis of PCa patients with *YWHAZ* amplification, or increased mRNA or protein levels, revealed significant alterations in key DNA repair genes. We hereby state the relevance of *YWHAZ* in PCa, showcasing its role as an independent strong predictor of aggressiveness.

¹Laboratorio de Inflamación y Cáncer, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EGA, Argentina. ²Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), CONICET-Universidad de Buenos Aires, Buenos Aires C1428EGA, Argentina. ³Department of Genitourinary Medical Oncology and The David H. Koch Center for Applied Research of Genitourinary Cancers, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. ⁴Cátedra de Urología, Hospital de Clínicas, Buenos Aires C1120AAR, Argentina. ⁵Department of Biomedical and Health Informatics, Rutgers School of Health Professions, Rutgers Cancer Institute of New Jersey, New Jersey, NJ 07101, USA. ⁶These authors contributed equally: Sofia Lage-Vickers, Juan Bizzotto. ⁷These authors jointly supervised this work: Javier Cotignola, Geraldine Gueron. ✉email: jcotignola@qb.fcen.uba.ar; ggueron@gmail.com

When prostate cancer (PCa) is diagnosed at a local or regional stage, the 5-year survival rate approaches 100%. But PCa is often symptomless in its early stages and once it has metastasized, survival rates decrease to 30%¹. Hence, screening for aggressive cancer at early stages may be crucial to improve survival.

Widespread use of prostate-specific antigen (PSA) levels for screening has led to a large increase in the incidence of diagnosed PCa and a reduction in both, advanced disease and PCa mortality rates². However, the overtreatment of PCa is widely recognized³. The difficulty in preventing overtreatment is the current inability to distinguish men who will have an indolent disease from those who will have aggressive disease. For men with newly diagnosed PCa, the strongest predictor is the Gleason Grade (GG). GG histologically groups PCa and presupposes an outcome for each grade within certain margins. However, in many cases, the outcome does not conform to these expectations and is usually more compromised⁴. Because of tumor heterogeneity that is inadequately captured by the biopsies, between 25 and 50% of biopsies with certain GG come from men with higher-grade PCa^{4–6}. This is known as “upgrading” between biopsy and prostatectomy, indicating that patients diagnosed with indolent PCa might have higher-grade and subsequently higher-risk cancers.

Hence, efficient PCa management should encompass not only early accurate diagnosis but also the identification of prognostic factors which help in foreseeing the outcome for all individual cases. Clinical, genomic, and/or radiological biomarkers are the key to appropriate risk stratification. Genomic biomarkers are being developed for screening for lethal disease subtypes, monitoring of PCa recurrence after initial treatments, prognosis, and prediction of drug efficacy⁷. The application of translational molecular profiling in PCa may, in the near future, have the potential to enhance clinical management. In this regard, the literature reflects some interesting avenues such as the prostate cancer antigen 3 (PCA3) score⁸, the Prostate Health Index (phi)⁹, and the Oncotype Dx¹⁰. While PCA3 has a better diagnostic performance than PSA^{11–13}, it does not add predictive value for GG or tumor stage^{14,15}. In direct comparisons, the phi is a better predictor of PCa at initial biopsy, and therefore more suitable for screening^{16–18}. Oncotype Dx for Prostate is used to further stratify low to low-intermediate risk PCa by calculating a Genomic Prostate Score. However, to prove clinical utility, potentially novel prognostic molecular markers will need to provide added, independent value, in multivariable analysis, beyond PSA, pathologic stage, and GG.

In this work, we undertook an in-depth mass spectrometry approach to profile proteomes from formalin-fixed paraffin-embedded (FFPE) specimens of prostate adenocarcinomas and benign prostate hyperplasias (BPHs), with available disease stage, Gleason score (GS), and patient age. The objective of our study was to identify novel biomarkers for risk stratification of PCa with an eye toward those that could behave independently from GS and further recognize intermediate GSs that may be more likely to progress. These molecular biomarkers might improve the prediction of lethal disease and provide insight into the biological mechanisms underlying the strong relation of GS and disease progression.

Results

Proteomic analysis of FFPE PCa and BPH. To identify potential PCa biomarkers for risk stratification, we performed LC ESI-MS/MS in human PCa ($n = 10$) and BPH ($n = 10$) archival tissues. The proteomics yield averaged 540 and 536 proteins per sample in PCa and BPH tissues, respectively. To analyze these proteins, we proceeded as described in the data analysis pipeline

(Supplementary Fig. 1a). We then selected proteins enriched in PCa samples compared with BPH samples (PCa enriched protein data set, $n = 109$) (Fig. 1a). We subjected the candidate proteins to clinical validation in extended cohorts of PCa patients (32 data sets; 5974 samples) (Supplementary Fig. 1b).

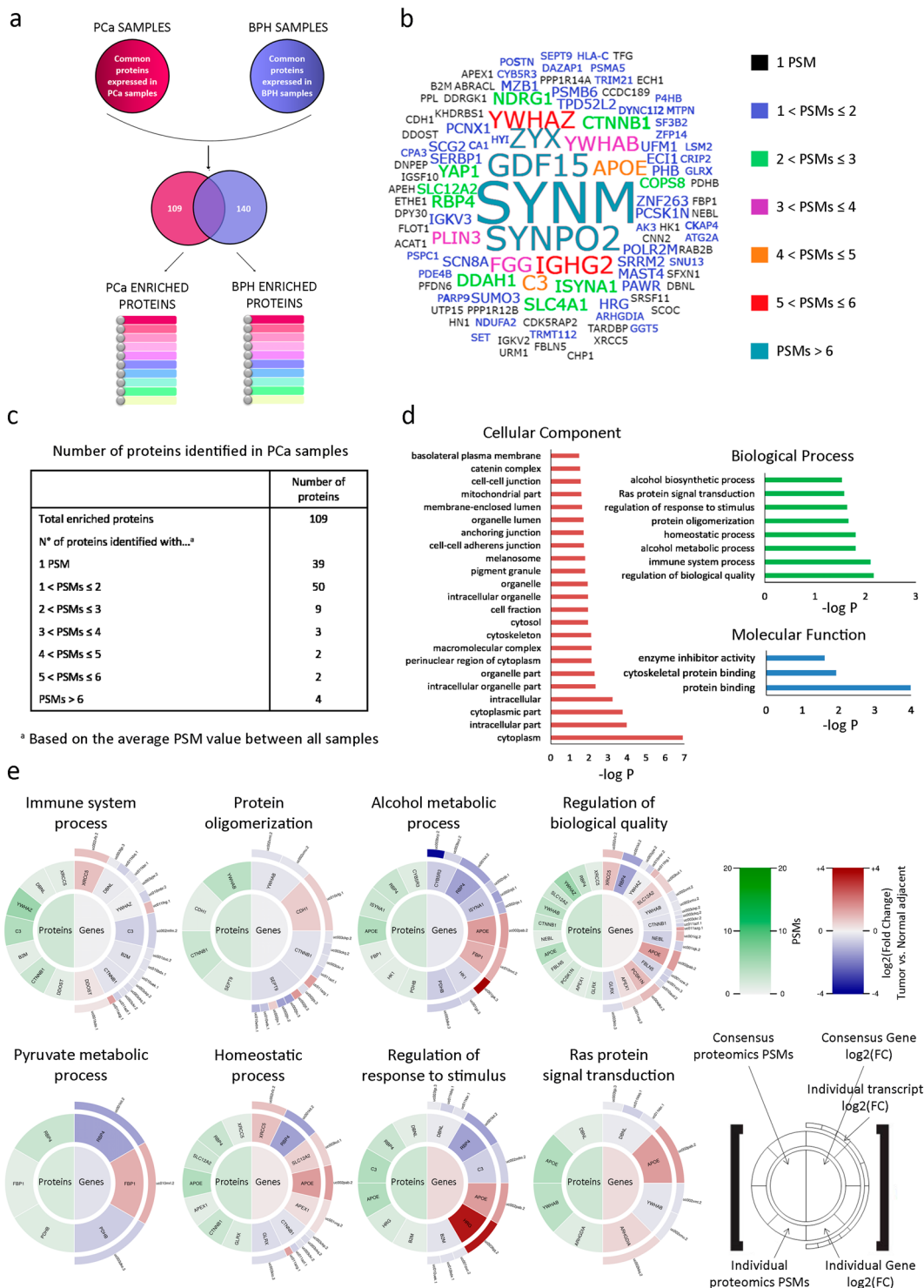
An integrated proteomics and transcriptomic atlas of PCa. To assess the clinical significance of the PCa enriched protein data set, we first generated a tag cloud network of the 109 proteins (Fig. 1b). These data are summarized in Fig. 1c, and complete lists of enriched proteins identified within each data set are presented in Supplementary Data 1 and 2.

Gene ontology (GO) classifications for the top biological processes (BP) categories of the PCa enriched protein list included RAS protein signal transduction, regulation of biological processes, and homeostasis (Fig. 1d). We then examined simultaneously the PCa proteome within the enriched GO categories, layering it with transcriptomic data from the *TCGA-PRAD* data set ($n = 499$)¹⁹. The patterns in the proteograms highlighted some accordance between proteins from the PCa enriched protein data set compared with the corresponding gene expression analysis from the *TCGA-PRAD* data set (Fig. 1e).

Analysis of the PCa proteome through multiple microarray data sets and selection of candidate biomarkers. In order to prioritize candidate biomarkers for PCa, we selected proteins within the PCa proteomes identified with more than 2 Peptide Spectral Matches (PSMs) (Fig. 1c). This filter resulted in the selection of 20 proteins (Fig. 2a). We next subjected our candidate list to bioinformatics analysis using *OncoPrint*²⁰. We selected 16 data sets ($n = 1128$) that analyze gene expression between PCa and normal prostate gland samples (Supplementary Table 1). The expression profile for most of these genes in the PCa enriched list identified with more than 2 PSMs showed significant dysregulation in PCa compared with normal gland (Fig. 2a). However, only *SLC12A2* ($P = 4.81e-4$), *DDAH1* ($P = 4.99e-4$), *NDRG1* ($P = 0.005$), *APOE* ($P = 0.011$), *YWHAZ* ($P = 0.002$), and *GDF15* ($P = 0.014$) displayed significant upregulated gene expression levels when comparing PCa vs. normal prostate gland (Fig. 2a, black boxes). The meta-analysis combining data from the independent data sets showed that the above-mentioned dysregulated genes lie within the 25% of the most consistently highly expressed genes across this comparison. Of note, *DDAH1*, *APOE*, and *YWHAZ* lie within the top 13% (Fig. 2a).

Further, in-depth analysis of *TCGA-PRAD* data revealed significantly higher gene expression profiles for *SLC12A2* ($P = 0.025$), *DDAH1* ($P < 0.0001$), *NDRG1* ($P = 0.0013$), *APOE* ($P < 0.0001$), *YWHAZ* ($P = 0.0385$), and *GDF15* ($P < 0.0001$), in PCa tumor tissue compared with non-tumoral adjacent tissue (Fig. 2b). Hence, the pattern evident in the proteogram, highlights a concordance between proteins enriched in PCa and the corresponding gene expression levels from the *TCGA-PRAD* data set (Fig. 2c).

Identification of risk predictors of PCa. We next evaluated the overall survival (OS) in PCa patients that had undergone TURP or adenoma enucleation with a high or low expression for each gene. The *Sboner* data set²¹ demonstrated that higher expression of *YWHAZ*, *NDRG1*, and *APOE* is significantly associated with poor OS in PCa ($P < 0.001$ for all genes) (Fig. 3a). Interestingly, *YWHAZ* is an androgen-responsive gene that activates proliferation, cell survival, and androgen receptor transcriptional activity²², *NDRG1* is a downstream target of *c-MYC* proto-oncogene²³, and *APOE* is associated with fat metabolism and cancer²⁴. *DDAH1* was not available in this data set (Fig. 3a).



To validate the potential of these biological markers to improve risk stratification in PCa, multivariable analyses were performed in the presence of clinico-pathological parameters previously associated with increased death risk. These parameters included GS, age group, and *TMPRSS2-ERG* fusion (Fig. 3b). High *YWHAZ* and *NDRG1* expression significantly correlated with poor prognosis. Both genes behaved independently from the patient's GS, age, or *TMPRSS2-ERG* fusion (Fig. 3b). When we further adjusted the model to include simultaneously all variables, the associations remained significant for both genes ($P < 0.001$ and $P = 0.011$, respectively) (Fig. 3b). Although high *APOE*

expression was associated with poor OS and the multivariable analyses including either age or *TMPRSS2-ERG* fusion were significant, no independence from the clinico-pathological parameters was observed when considering all variables simultaneously (Fig. 3b). These results may suggest that dysregulation of *APOE* may be either accompanying GS or maybe just a molecular/metabolic consequence of it.

Further, we considered a multivariable Cox proportional hazard model including *YWHAZ*, *NDRG1*, *APOE*, GS, age, and *TMPRSS2-ERG* fusion in treatment-naive PCa. Results showed that only *YWHAZ*, *NDRG1*, GS, and age could be independent

Fig. 1 Identification of PCa enriched proteins. **a** Data analysis was based on label-free spectral counting, obtaining an average of 540 and 536 proteins per sample in PCa and BPH tissue proteomes, respectively. Two proteomic data sets were generated (PCa and BPH), taking into account the proteins that were found in at least 50% of the samples of each type of tissue, and that were also not shared between both groups. 109 proteins were enriched in PCa samples, while 140 proteins were enriched in BPH samples. **b** Tag cloud network of the 109 PCa enriched proteins identified consistently across the proteomics analyses performed in PCa compared to BPH FFPE human tissue samples. The font size increases proportionally to the average PSMs of these proteins across all the PCa samples. The color reflects the same information. **c** Semi-quantitative analysis of proteins identified by LC ESI-MS/MS classified by the number of PSMs obtained from PCa tissue samples. Results represent the average PSMs for each group. **d** GO analysis of PCa enriched proteins using the DAVID software which includes only significant categories ($-\log P \geq 1.5$) from cellular components, biological process, and molecular functions. **e** Protein levels (PSMs) and gene differential expression (PCa vs. normal adjacent tissue, *TCGA-PRAD*) of PCa enriched proteins were grouped by GO terms from the biological process category using DAVID software and visualized in proteogram plots. These graphs are circular heat-plots depicting both protein PSMs on the left semicircle and gene expression fold change (as \log_2 (fold change)) on the right semicircle, with the center indicating the averages. For the gene expression, a half outer semicircle was added to include each gene's transcript's expression level. PSM data were locally generated while gene expression data was gathered from *TCGA-PRAD*.

predictors of death risk ($P < 0.001$, $P = 0.02$, $P < 0.001$, and $P < 0.001$, respectively) (Fig. 4a). Hence, we dropped *APOE* and categorized PCa patients based on *YWHAZ* and *NDRG1* gene expression levels. The heatmap depicts patient subgroups with (1) low *YWHAZ* and *NDRG1* expression ($n = 118$); (2) high *NDRG1* and low *YWHAZ* expression ($n = 131$); (3) high *NDRG1* and *YWHAZ* expression ($n = 24$), and (4) low *NDRG1* and high *YWHAZ* expression ($n = 8$) (Fig. 4b). Next, we performed OS analyses of these patient subgroups. Patients in groups 2, 3, and 4 had significantly decreased OS compared with patients in group 1 ($P < 0.001$ for all comparisons) (Fig. 4c). No significance was observed when comparing OS between patients in groups 3 and 4 (Fig. 4c). These results evidenced three clear and distinct subpopulations of PCa patients with low (group 1), intermediate (group 2), and high (groups 3 and 4) risk of death. Since groups 3 and 4 presented a similar OS, we inferred that *NDRG1* has no significant effect on patients' OS with high *YWHAZ* expression. Hence, *YWHAZ* seemed to be the main driver for the increased risk of death in patients with both high *NDRG1* and *YWHAZ*. In light of these results, we merged these two subgroups of patients for further analysis and named it group 3. Next, when categorizing patients based on GS, we were able to see once more how these subgroups separated into low, intermediate, and high-risk patients (Fig. 4d, e). In particular, for GS 7, we could observe high statistical significance for all comparisons (2 vs. 1 $P = 0.004$; 3 vs. 1 $P < 0.001$; 3 vs. 2 $P = 0.007$), reinforcing the power of these two biomarkers to define three subgroups of PCa risk (Fig. 4d). We further split the GS 7 patients into GS 7 (3 + 4) and GS 7 (4 + 3). KM curves evidenced the power of *YWHAZ* to predict a worse outcome for both GS: 7 (3 + 4) (3 vs. 1 $P < 0.001$) (Fig. 4f) and 7 (4 + 3) (3 vs. 1 $P < 0.001$) (Fig. 4g). *NDRG1* only defined an intermediate risk group for 7 (4 + 3) (2 vs. 1 $P = 0.016$) (Fig. 4g).

When performing the same analysis stratifying patients based on age, high *YWHAZ* expression was also associated with significantly poorer outcomes (3 vs. 1 $P < 0.001$ for both age groups) (Supplementary Fig. 2).

Since two independent PCa prognostic factors were delineated using the *Sboner* data set as our training data set, we sought to predict survival based on *YWHAZ* and *NDRG1* in three other independent non-overlapping validation data sets: *TCGA-PRAD* ($n = 499$), *Ross-Adams* ($n = 206$)²⁵, and *Jenkins* ($n = 596$)²⁶. The heatmaps depict patient subgroups with (1) low *YWHAZ* and *NDRG1* expression; (2) high *NDRG1* and low *YWHAZ* expression; (3) high *YWHAZ* expression regardless of *NDRG1* expression (Fig. 5a (*TCGA-PRAD*), c (*Ross-Adams*), f (*Jenkins*)). Similar results were obtained in which *NDRG1* defined a PCa subgroup with intermediate-risk and *YWHAZ* defined a PCa subgroup with high risk, regarding disease-specific survival (DSS) (*TCGA-PRAD* and *Jenkins*) and relapse-free survival (RFS) (*Ross-*

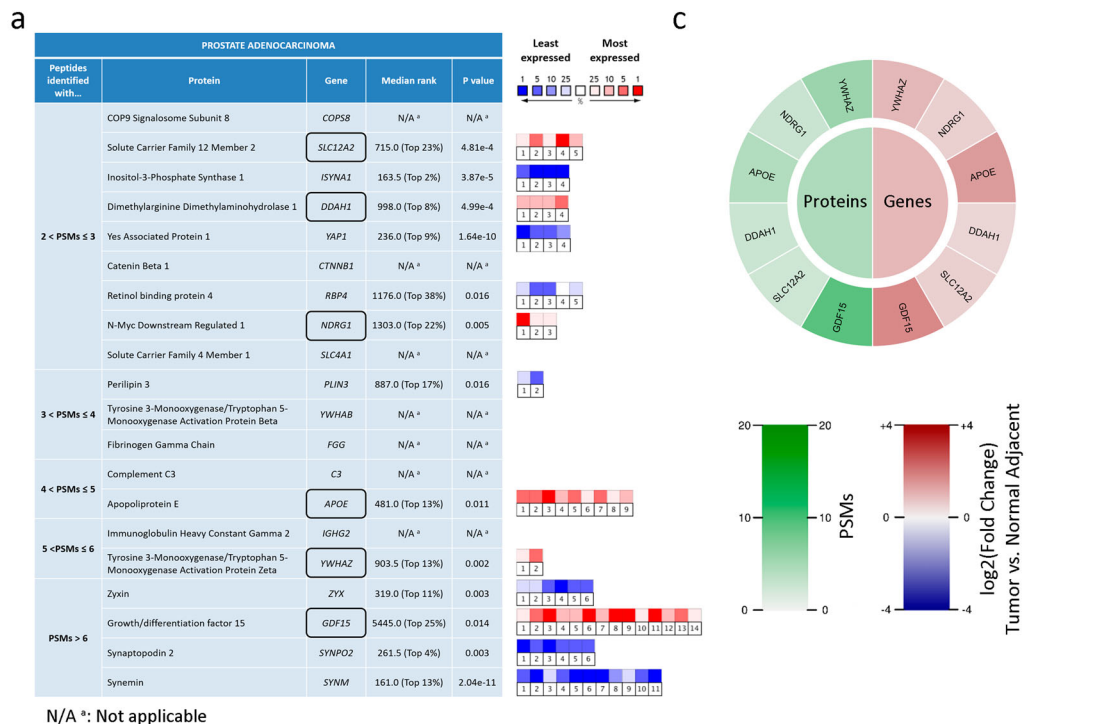
Adams) for all PCa patients (Fig. 5b, d, g). When subcategorizing by GS 7, only patients with high *YWHAZ* were associated with decreased RFS (*Ross-Adams*) (3 vs. 1 $P = 0.009$) and DSS (*Jenkins*) (3 vs. 1 $P < 0.001$) (Fig. 5e, h).

Further, we used the time-dependent univariable AUC metric to evaluate *YWHAZ* and *NDRG1* performance as predictors of death in PCa patients, using the *Jenkins* data set which contains DSS. Time-dependent AUC curves showed that, although *NDRG1* alone did not outperform GS, the model including both, *NDRG1* and GS, improved the prediction of DSS compared with GS alone beyond 120 months post-diagnosis (AUC = 0.614 for *NDRG1* + GS and AUC = 0.545 for GS at 120 months; $P = 0.0043$) (Fig. 5i). In the case of *YWHAZ*, the univariable AUC model comparing this factor with the univariable AUC for GS revealed that *YWHAZ* outperformed GS prognosis beyond 60 months (AUC = 0.718 for *YWHAZ* and AUC = 0.551 for GS at 60 months; $P = 0.037$) (Fig. 5j). The model including both, GS and *YWHAZ*, did not present significant differences compared with the predictive value of *YWHAZ* alone (Fig. 5j). Thus, *YWHAZ* rises as a potential predictor of aggressiveness in PCa.

Genomic landscape of PCa patients with *YWHAZ* alterations.

We extended the bioinformatics analysis of *YWHAZ* using *cBioPortal for Cancer Genomics*²⁷. 11 PCa data sets were selected that met our eligibility criteria ($n = 2820$) (Supplementary Table 2). The most frequent genetic alteration found was gene amplification (Supplementary Fig. 3a). Results showed significantly reduced disease-free survival in patients with *YWHAZ* gene alteration ($P = 8.141e-3$) (Supplementary Fig. 3b). The only data set available for direct correlation of exome data with RNAseq data was *TCGA-PRAD* ($n = 499$). There was a significant direct correlation between amplification for *YWHAZ* and mRNA levels ($P < 0.001$) (Supplementary Fig. 3c) and a positive correlation between 14-3-3 ζ/δ levels and *YWHAZ* expression (Spearman coefficient = 0.29, $P = 2.84e-8$) (Supplementary Fig. 3d). Interestingly, the Reverse Phase Protein Array (RPPA) from *TCGA-PRAD* showed that patients with high 14-3-3 ζ/δ levels had poorer disease-specific survival (DSS) ($P = 0.021$) (Supplementary Fig. 3e).

Our results prompted the question as to whether PCa patients with *YWHAZ* genetic alterations, high *YWHAZ* expression or high 14-3-3 ζ/δ levels, harbored a differential genomic landscape. We focused our work on key DNA repair genes that were previously described to have a high prevalence of mutations and copy number alterations in studies on localized and metastatic PCa^{28,29}. When analyzing the genetic alterations of patients with localized PCa from the *TCGA-PRAD* data set with gene amplification for *YWHAZ*, there was a significantly higher number of altered genes per patient with *YWHAZ* amplification compared with those with unaltered *YWHAZ* ($P < 0.001$) (Fig. 6a



and Supplementary Fig. 4). From these genes, nine (*MYC*, *FANCA*, *BRCA2*, *RBI*, *PALB2*, *ATM*, *TP53*, *CXCL12*, and *MLH1*) had significantly more genetic alterations when *YWHAZ* was amplified (Fig. 6b, red bars, c). Further, when analyzing patients with high vs. low *YWHAZ* mRNA and protein expression, four of those nine genes (*MYC*, *PALB2*, *TP53*, and *CXCL2*) also had increased significant genomic alterations (Fig. 6b, blue and green bars, respectively, c). Accordingly, similar results were obtained when performing the same comparison at the genomic level for *YWHAZ* in patients with metastatic PCa using the *SU2C/PCF 2019* data set ($n = 444$)³⁰ (Supplementary Fig. 5). These

results suggest a link between *YWHAZ* alterations and genomic instability and raises the question as to whether *YWHAZ* may also be a driver of progression.

Discussion

In this work, we are reporting the relevance of *YWHAZ*/14-3-3ζ/δ as an independent strong predictor of death in PCa. Furthermore, we have discovered that this factor outperforms GS prediction of DSS beyond 5 years after initial diagnosis.

Briefly, in this study, we have: (1) applied high throughput proteomics methods to discover signature protein biomarkers

Fig. 2 Correlation between proteomics and transcriptomics data for the PCa selected enriched proteins. **a** Summary table showing the gene symbol & name, median gene rank and corresponding *P*-value obtained with *OncoPrint* ($n = 1128$) for the PCa enriched proteins identified with a higher number of PSMs. The median rank for a gene is the rank for the gene across each of the analyses for the 16 data sets assessed. The *P*-value for a gene is the *P*-value for the median ranked analysis. The heatmap on the right indicates the level of expression for each gene in each study selected (blue: least expressed, red: most expressed). Each square represents each *OncoPrint* study that met our eligibility criteria and thresholds (see “Methods” section). Colors are z-score normalized to depict relative values within a row. Proteins found in PCa samples and upregulated in prostate carcinoma vs. normal prostate gland (*OncoPrint*) are boxed in black. N/A ^a: not applicable. **b** Tukey box plots showing median gene expression of *SLC12A2*, *DDAH1*, *NDRG1*, *APOE*, *YWHAZ*, and *GDF15* comparing tumor vs. Normal adjacent tissues using the *TCGA-PRAD* data set ($n = 499$). The top and bottom of each rectangular box represent the 75th and 25th percentiles, respectively, with the median indicated with a solid line inside the box. Horizontal bars extending from each box represent more extreme values defined as 1.5-times the interquartile range (25th percentile subtracted from the 75th percentile) above the 75th percentile or below 25th percentile. Circles represent outliers. Student’s *t*-test was used to ascertain statistical significance. Statistical significance was set at $P \leq 0.05$. **c** Protein levels (PSMs) and gene differential expression (PCa vs. normal adjacent tissue, *TCGA-PRAD*) of PCa selected enriched proteins were put together in a proteogram plot.

that may be useful when analyzing PCa, and (2) identified and validated risk stratification markers within PCa patient data sets. We resumed our results in 20-candidate signature proteins enriched in PCa compared with BPH. Of those, 55% (11/20) correspond to a subset of proteins directly associated with PCa, and ~85% (17/20) are known cancer markers or have been previously associated with cancer in general^{31–35}. These results clearly highlight the robustness of our working strategy and methodology. When analyzing these 20 candidates, the following gene transcripts appeared to be increased in PCa compared with the normal prostate gland: *SLC12A2*, *DDAH1*, *GDF15*, *APOE*, *NDRG1*, and *YWHAZ*. Although we did not validate the MS results on the protein level, there are previous reports that show increased protein expression of *DDAH1*, *GDF15*, *APOE*, *NDRG1*, and 14-3-3 ζ/δ in PCa tissues compared with benign prostate tissues^{22,36–39}.

In this work, we further merged our FFPE human PCa tissue proteomes data with large PCa transcriptome data sets, OS, RFS, and DSS data, and refined markers that could predict PCa survival. The univariable analyses performed on a well-characterized cohort (the Swedish Watchful Waiting Cohort after TURP or adenoma enucleation), revealed a positive association of increased expression of *APOE*, *YWHAZ*, and *NDRG1* with poor OS. When performing multivariable analyses, only *YWHAZ* and *NDRG1* showed independence from GS, age at diagnosis, and *TMPRSS2-ERG* fusion in PCa patients. These two genes were able to categorize PCa patients in low, intermediate, and high risk of death. Moreover, this pattern was also observed with statistical significance for GS 7. Further, we split the GS 7 patients in GS 7 (3 + 4) and GS 7 (4 + 3) in line with the current classification GG 2 and GG 3, and once more, the KM curves still evidenced the power of *YWHAZ* to predict a worse outcome for both groups.

Since TURP might trigger tumor cell dissemination^{40–43}, we might speculate that this could be one of the reasons for the short OS in this cohort. For this reason, we furthered the analysis using the *Jenkins* data set that contains DSS. The ROC analysis showed that, although *NDRG1* did not outperform GS, the model including GS + *NDRG1* improved the predictive value of aggressive disease compared with GS alone beyond 120 months post-diagnosis. High *NDRG1* expression relates to increased cell differentiation signals in various cancer cell lines and the suppression of tumor metastasis²³. In particular, its role in PCa is controversial since there are both, reports suggesting *NDRG1* as a tumor suppressor⁴⁴, and as an oncogene^{39,45}.

In the case of *YWHAZ*, this factor outperformed GS beyond 60 months post-diagnosis. Interestingly, the ROC curves did not showcase improved predictive value when adding GS to the *YWHAZ* model, further ascertaining this factor as a potential prognostic tool in the clinic. 14-3-3 ζ/δ is an adapter protein

implicated in the regulation of a large spectrum of both general and specialized signaling pathways. 14-3-3 ζ/δ belongs to the 14-3-3 family of proteins that mediate signal transduction by binding to phosphoserine-containing proteins and is encoded by the *YWHAZ* gene. Increased expression of *YWHAZ* relates to tumor cell proliferation and malignant outcome of gastric carcinoma⁴⁶. In localized PCa, Ruenuver et al. associate *YWHAZ* with PCa⁴⁷, however, these authors fail to demonstrate its relevance as a prognostic factor independent from the common PCa clinicopathological parameters.

We hereby state the relevance of *YWHAZ*/14-3-3 ζ/δ in PCa showcasing its role as an independent strong predictor of death that outperforms GS. With the identification of the mutational landscape of organ-confined and advanced-stage disease, a major contribution has been made to the development of molecular biomarker profiling in addition to serum PSA. It is convenient that clinico-pathological parameters, imaging, and molecular markers are integrated together to better predict tumor behavior. *YWHAZ*/14-3-3 ζ/δ could be a promising tool when taking into consideration the difficulties that PCa presents at the time of decision making.

It is worth mentioning that when evaluating the association of *YWHAZ* with survival, its genetic alteration was significantly related to poor prognosis. These observations made us speculate as to whether patients with *YWHAZ* genetic alterations, high *YWHAZ* gene expression, or high 14-3-3 ζ/δ protein levels, harbored a differential genomic profile compared with PCa patients with no *YWHAZ* alterations. We centered our attention on key DNA repair genes associated with localized and metastatic PCa^{28,29}. Interestingly, PCa patients with *YWHAZ* amplification harbored significantly more genetic alterations. Of note, *MYC*, *PALB2*, *TP53*, and *CXCL2*, also had increased genetic alterations when comparing patients with high vs. low *YWHAZ* mRNA and protein expression. In the case of *MYC*, almost 92% of PCa patients with amplified *YWHAZ* presented *MYC* alterations as opposed to only 1.8% of patients with no *YWHAZ* alteration in *TCGA-PRAD*. Although co-occurrence of *MYC* and *YWHAZ* alterations could be explained by previously reported 8q gains associated with tumor progression and poor prognosis in PCa²⁹, it is worth noticing that *YWHAZ* amplification, or increased *YWHAZ* mRNA and protein levels, significantly correlated with genetic alterations in other genes that are located at different chromosomes. These results reflect that a subpopulation of PCa patients with high *YWHAZ*/14-3-3 ζ/δ shows greater number of genetic alterations in key DNA repair genes. Future work should address whether *YWHAZ* may also be a driver of progression.

We acknowledge the limitation of validating the initial proteomics screening in PCa and BPH samples with transcriptomics data. However, since we were processing archival FFPE samples,

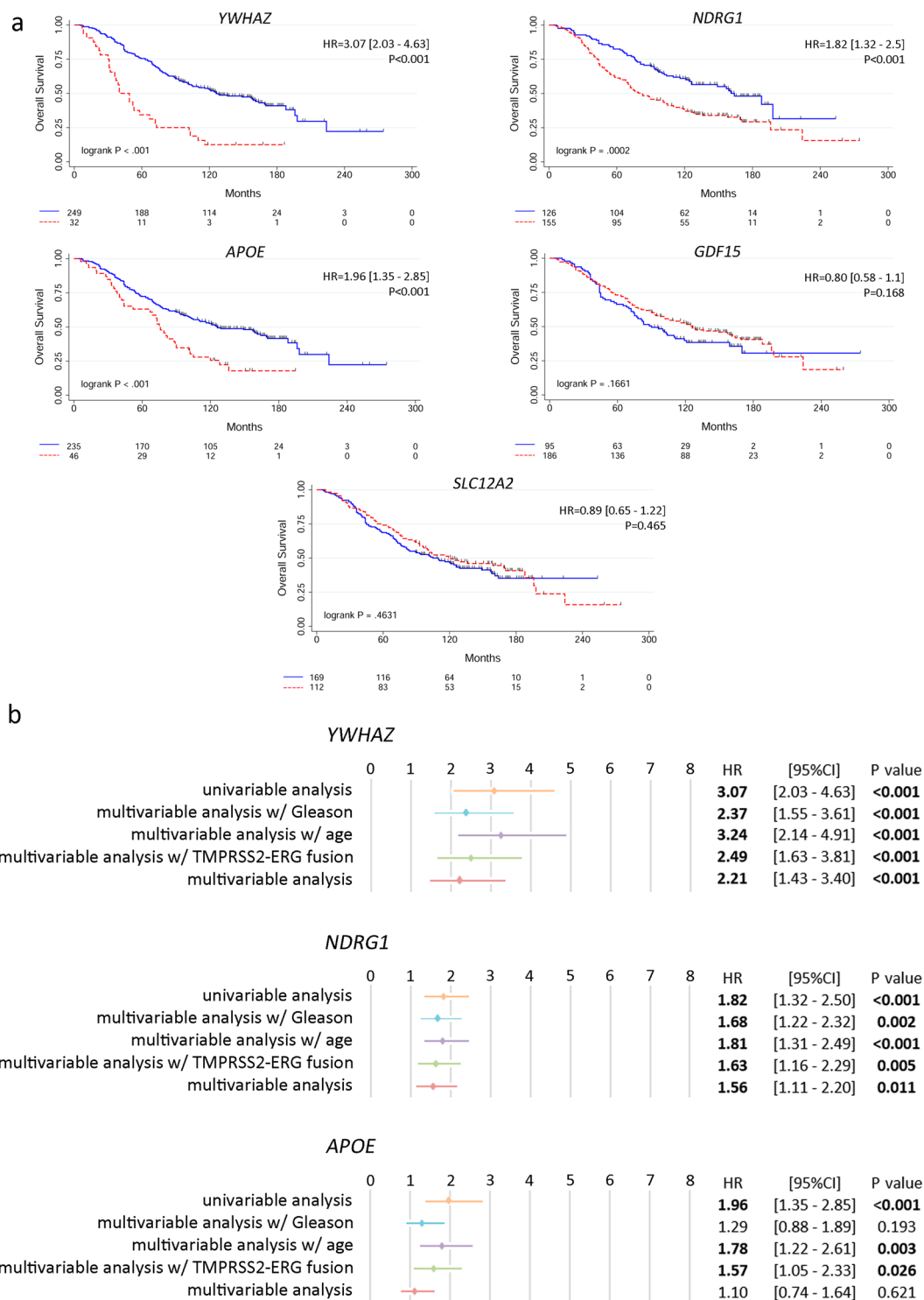


Fig. 3 Overall survival (OS) of PCa patients naive of treatment based on YWHAZ, NDRG1, APOE, GDF15, and SLC12A2 expression (Sboner data set, GSE16560, n = 281). **a** Kaplan-Meier curves for OS for PCa patients segregated based on the gene expression levels for YWHAZ, NDRG1, APOE, GDF15 and SLC12A2. OS of patients with high (red dotted-lines) vs. low (blue full-lines) expression for each gene. **b** Multivariable analyses presented by forest plots including each gene with GS, age or TMPRSS2-ERG fusion or all the variables together. Statistical significant associations are bolded. Multivariable analysis w/Gleason (light blue) = adjusted for GS (6; 7 (3 + 4); 7 (4 + 3); 8-10). Multivariable analysis w/age (purple) = adjusted for age at diagnosis (age groups: 50-70; 70-80; 90-100). Multivariable analysis w/TMPRSS2-ERG fusion (green) = adjusted for TMPRSS2-ERG fusion. Multivariable analysis (red) = adjusted for GS, age at diagnosis and TMPRSS2-ERG fusion. GS = Gleason score. HR = hazard ratios [95% confidence interval] for the univariable analysis. All comparisons consider low expression patients as the reference group. P = Cox proportional hazard model P-value. Statistical significance was set at P ≤ 0.05.

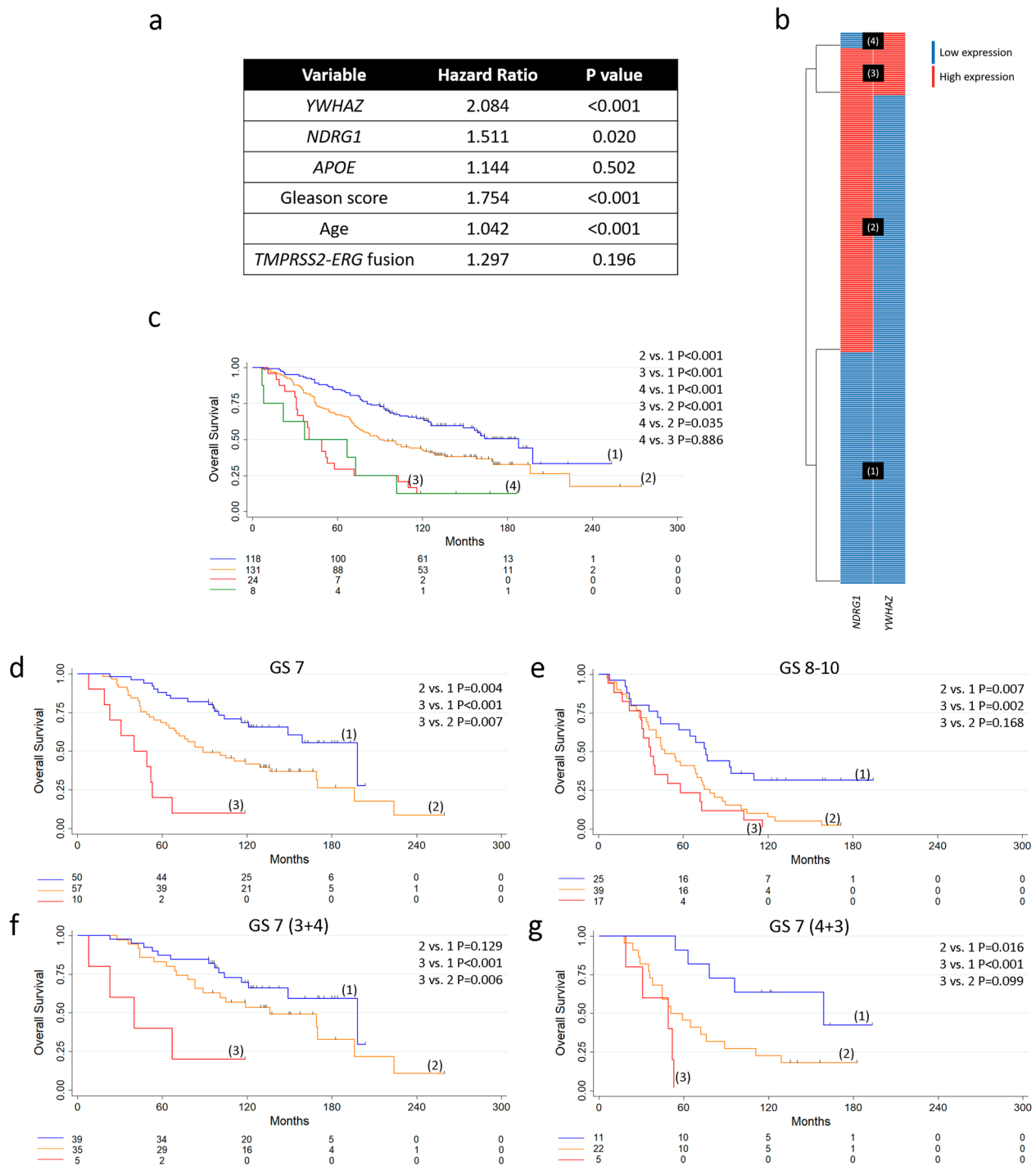


Fig. 4 *YWHAZ* and *NDRG1* as risk stratification genes in PCa patients naive of treatment (*Sboner* data set, GSE16560, $n = 281$). **a** Multivariable analyses based on gene expression of *YWHAZ*, *NDRG1*, *APOE*, GS, age and *TMPRSS2-ERG* fusion for patients with PCa. $P =$ Cox proportional hazard model P -value. **b** Heatmap depicting low (blue) or high (red) *NDRG1* and *YWHAZ* gene expression for patients with PCa. Patient subgroups are presented in black boxes. **c–g** OS of patients with low *YWHAZ* and *NDRG1* gene expression (1), high *NDRG1* gene expression (2), high *YWHAZ* gene expression (4), and high *YWHAZ* and *NDRG1* gene expression (3). Kaplan–Meier curves for OS for PCa patients segregated based on gene expression levels for *YWHAZ* and *NDRG1* for GS 7 (**d**) and GS 8–10 (**e**), and gene expression levels for *YWHAZ* and *NDRG1* for GS 7 (3 + 4) (**f**) and GS 7 (4 + 3) (**g**). GS = Gleason score. All comparisons consider low expression patients as the reference group. $P =$ pairwise log-rank P -values. Statistical significance was set at $P \leq 0.05$.

we had to prioritize tissue for LC ESI–MS/MS. Of note, the *Sboner* data set (GSE16560) included TURP and adenoma enucleation specimens, which might introduce short overall survival bias.

Methods

Experimental design. The study aimed to identify potential biomarkers for PCa risk stratification. An in-depth proteomics analysis (LC ESI–MS/MS) was done on human PCa and BPH tissues (previously confirmed by histological analyses performed by pathologists at the Hospital de Clinicas), since normal prostate gland

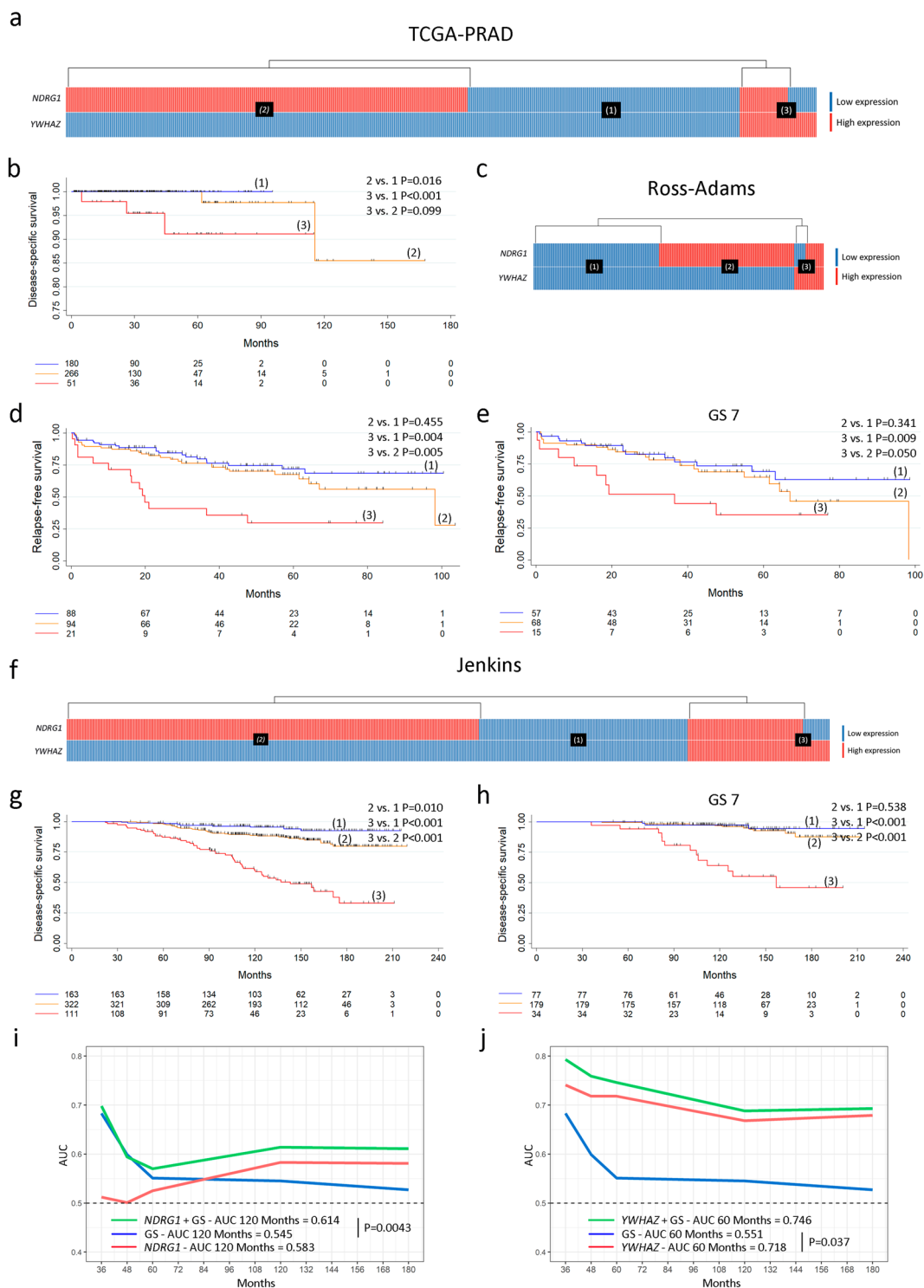


Fig. 5 Validation of *YWHAZ* and *NDRG1* as risk stratification genes in PCa patients (TCGA-PRAD, $n = 499$; Ross-Adams, GSE70770, $n = 206$; and Jenkins, GSE10645, $n = 596$). **a, c, f** Heatmaps depicting low (blue) or high (red) *NDRG1* and *YWHAZ* gene expression for patients with PCa according to the TCGA-PRAD (**a**), Ross-Adams (**c**), and Jenkins (**f**) data sets. Patient subgroups are presented in black boxes: low *YWHAZ* and *NDRG1* expression (1), high *NDRG1* expression (2), and high *YWHAZ* expression (3). **b, d, g** Kaplan-Meier curves for DSS (TCGA-PRAD) (**b**), RFS (Ross-Adams) (**d**), and DSS (Jenkins) (**g**), for PCa patients segregated based on gene expression levels for *YWHAZ* and *NDRG1*. **e, h** Kaplan-Meier curves for RFS (Ross-Adams) (**e**) and DSS (Jenkins) (**h**), for PCa patients segregated based on gene expression levels for *YWHAZ* and *NDRG1* in GS 7. $P =$ pairwise log-rank P -values. **i, j** Time-dependent AUC curves measured from 36 to 180 months reflecting the performance of GS (blue), *NDRG1* (**i**) or *YWHAZ* (**j**) (red), and *NDRG1* + GS (**i**) or *YWHAZ* + GS (**j**) (green) in PCa patients. AUC = area under the ROC curve. $P = P$ -value for DeLong’s test for two ROC curves. GS = Gleason score. All comparisons consider low expression patients as the reference group. Statistical significance was set at $P \leq 0.05$.

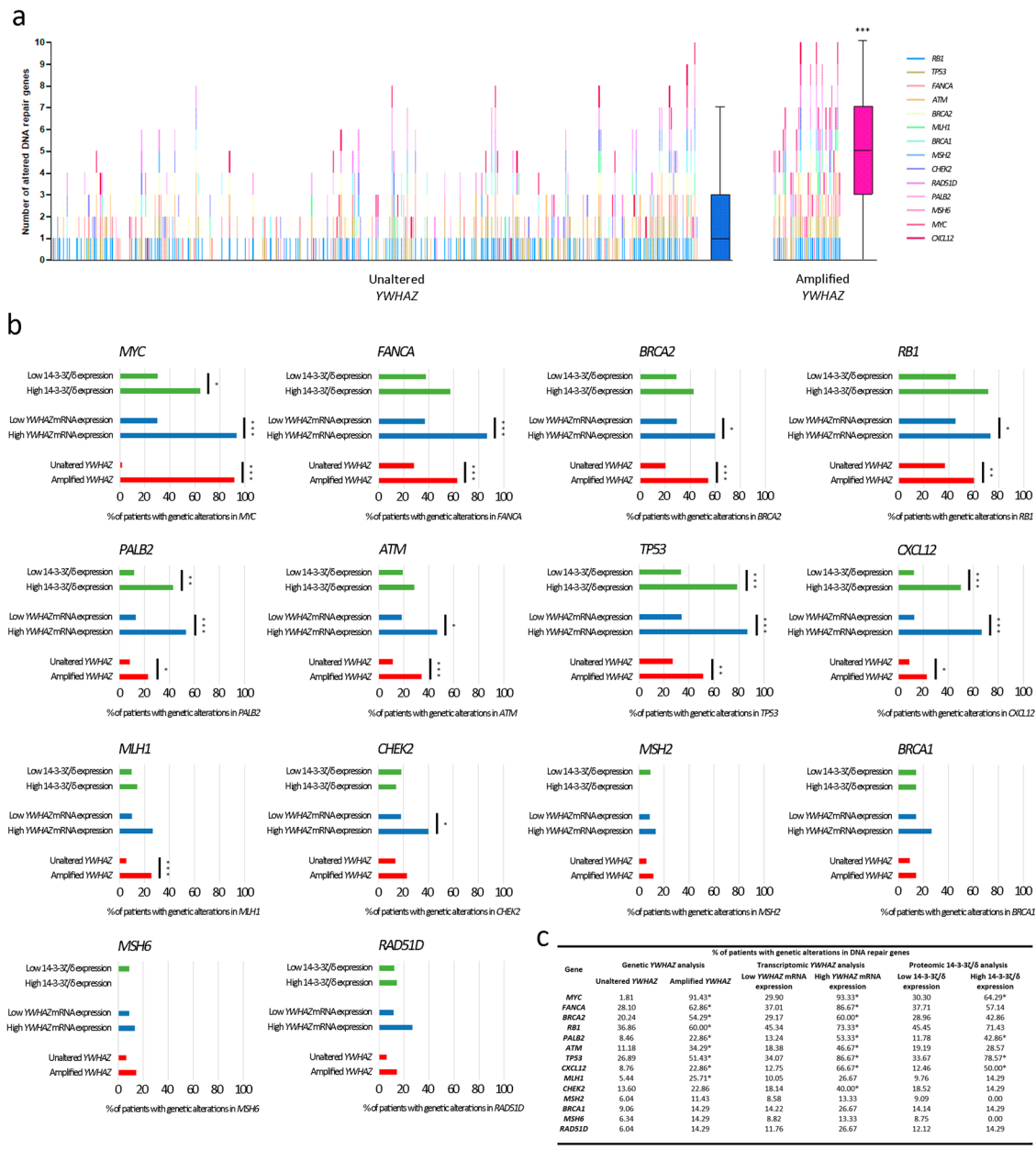


Fig. 6 Genetic alterations (amplification, gain, shallow deletion, deep deletion, and point mutations) in DNA repair genes in PCa patients with YWHAZ amplifications and high YWHAZ mRNA and protein levels (TCGA-PRAD data set, $n = 499$). **a** Mutational landscape analysis (amplification, gain, shallow deletion, deep deletion, and point mutations) of DNA repair genes in patients with no alterations in YWHAZ ($n = 331$) and with amplifications in YWHAZ ($n = 35$). Each vertical line is a patient and the different colors represent alterations in a different gene, as specified in the references. The graph includes box plots showing the median number of altered DNA repair genes. The top and bottom of each rectangular box represent the 75th and 25th percentiles respectively, with the median indicated with a solid line inside the box. Horizontal bars extending from each box represent more extreme values defined as 1.5-times the interquartile range (25th percentile subtracted from the 75th percentile) above the 75th percentile or below 25th percentile. Student's *t*-test was used to ascertain statistical significance. **b** Bar plots representing the percentage of PCa patients that present genetic alterations in each DNA repair gene based on whether they have amplifications in YWHAZ (red bars), high or low YWHAZ mRNA levels (blue bars), and high or low 14-3-3ζ/δ protein levels (green bars). **c** Comparative table of the percentage of patients that present genetic alterations in each DNA repair gene based on whether they have amplifications in YWHAZ, high or low YWHAZ mRNA levels, and high or low 14-3-3ζ/δ protein levels. Fisher's exact test was used to test the statistical significance of contingency tables of genetic alterations. Statistical significance was set at $P \leq 0.05$. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P < 0.001$.

samples were not available. Formalin-fixed and paraffin-embedded (FFPE) section tissues from 20 patients (ten radical prostatectomy specimens of treatment-naive PCa patients and ten BPH patients) were analyzed. Three FFPE sections per patient were used for protein extraction. Disease- and patient-associated data including pathologic and clinical stage and patient age were also obtained (Supplementary Table 3). Samples with less than 250 peptides were excluded from the subsequent analysis. Enriched PCa and BPH protein lists were formulated considering peptides that were found at least in $\geq 50\%$ of tissue samples and were not common to both

groups (PCa and BPH enriched protein data sets, respectively). Subsequently, extended cohorts of PCa patients (32 data sets; 5974 samples) were used for validation of proteins of interest.

FFPE processing. PCa and BPH FFPE tissue sections mounted on microscope slides were processed as previously described in Wakabayashi et al.⁴⁸ PCa and BPH FFPE tissue sections were deparaffinized and rehydrated by successive washes

in 100% *n*-octane (1 × 1 h), 100% (2 × 6 min), 96% (2 × 6 min), 70% (2 × 6 min), 50% ethanol (2 × 6 min), and water (3 × 6 min). After air-drying, the tissue sections were percolated with CellLytic MT Mammalian Tissue Lysis/Extraction Reagent (Sigma) for 90 min and harvested with a scalpel blade. The collected tissues were incubated on a heating block at 99 °C for 60 min and then sonicated for five cycles of 30 s ON, 30 s OFF. The debris was pelleted by spinning at 1500 × g at 4 °C for 20 min. The recovered proteins in the supernatant were quantified with a Pierce BCA Protein Assay kit.

LC ESI-MS/MS analysis. The digests were analyzed by nanoLC-MS/MS in a ThermoScientific Q-Exactive Mass Spectrometer coupled to a nanoHPLC EASY-nLC 1000 (ThermoScientific). For the LC ESI-MS/MS analysis, ~1 μg of peptides was loaded onto the column and eluted for 120 min using a reverse-phase column (C18, 2 μm, 100 Å, 50 μm × 150 mm), Easy-Spray Column PepMap RSLC (P/N ES801) suitable for separating protein complexes with a high degree of resolution. The flow rate used for the nano column was 300 nl min⁻¹ and the solvent gradient range was 7% B (for 5 min) to 35% B in 120 min. Solvent A was 0.1% formic acid in water whereas B was 0.1% formic acid in acetonitrile. The injection volume was 2 μl. The MS equipment has a high collision dissociation cell (HCD) for fragmentation and an Orbitrap analyzer (Q-Exactive-ThermoScientific Germany). A voltage of 3.5 kV was used for Electro Spray Ionization (ThermoScientific, EASY-SPRAY).

XCalibur 3.0.63 (ThermoScientific) software was used for data acquisition with a configuration that allows peptide identification at the same time as their chromatographic separation. A Data dependant method was used: Full-scan mass spectra were acquired in the Orbitrap analyzer. The scanned mass range was 400–1800 *m/z*, at a resolution of 70000 at 400 *m/z* and the twelve most intense ions in each cycle were sequentially isolated, fragmented by HCD, and measured in the Orbitrap analyzer. Peptides with a charge of +1 or with an unassigned charge state were excluded from fragmentation for MS2.

Analysis of LC ESI-MS/MS data. Raw data generated with Xcalibur software was processed and analyzed with Proteome discoverer 2.1.1.21 with SEQUEST Search engine. Spectrum Selector node with default parameter settings was used to generate peak lists. Minimum and maximum precursor masses were set at 350 and 5000 with an S/N of 1.5. Data were searched against Uniprot Homo sapiens UP000005640 database, October 2017, with trypsin specificity (full cleavage) and a maximum of two missed cleavages per peptide. Carbamidomethylation of cysteine residues was set as a fixed modification and oxidation of methionine was set as variable modification. Proteome Discoverer searches were performed with a precursor mass tolerance of 10 ppm and product ion tolerance to 0.05 Da. Proteome Discoverer default settings were used: Target FDR = 0.01; Z = 1 High confidence XCorr 1.5; Z = 2 High confidence XCorr 2; Z = 3 High confidence XCorr 2.5; z ≥ 4 High confidence XCorr 3. Protein hits were filtered for high confidence peptide matches with a maximum protein and peptide false discovery rate of 1% calculated by employing a reverse database strategy.

Proteograms. Proteograms were constructed as an adaptation from the web-based analytic tool Metabogram⁴⁹. Protein levels (Peptide Spectrum Matches, PSMs) and differential gene expression (PCa vs. normal adjacent tissue, *TCGA-PRAD*)¹⁹ of PCa enriched proteins were grouped by GO Terms from the biological process category and visualized in proteogram plots. These are circular heat-plots depicting both protein PSMs on the left semicircle and gene expression fold change (as log₂ (fold change)) on the right semicircle, with the center indicating the averages. For the gene expression, a half outer semicircle was added to include the expression level for each gene transcript. PSM data was locally generated while gene expression data was gathered from the publicly available *TCGA-PRAD* data set.

Gene ontology (GO) analysis. GO analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.7. (Leidos Biomedical Research, Inc., Bethesda, MD, USA)⁵⁰.

Bioinformatics analysis

Information source and eligibility criteria (The Cancer Genome Atlas (TCGA)). We used the data set from the Prostate Adenocarcinoma Project of The Cancer Genome Atlas (*TCGA-PRAD*)¹⁹ that has gene expression data from 499 prostate tumor and normal adjacent tissue samples (last access: December 2019), measured by massively parallel sequencing (IlluminaHiSeq). A descriptive table regarding patient characteristics at baseline (start of the follow-up for analyses) is depicted in Supplementary Table 4.

Information source and eligibility criteria (Oncomine) (n = 1128). We searched the public cancer microarray database *Oncomine*²⁰ (715 data sets and 68 tumor types; last access: December 2019) to identify expression microarray data sets that compared the expression of prostate adenocarcinoma vs. normal prostate gland. To be included in our study, a data set was required to (1) be generated from human prostate tumors, and (2) compare prostate adenocarcinoma vs. normal prostate gland. Differential genes were considered when: (1) they presented a *P*-value < 0.05

and (2) had an increase or decrease in expression ≥ 1.5 times and/or had a gene rank within the top 10%. Although the *P*-value criteria was strict for the data set selection, some genes were considered even if the fold change or the gene rank was < 1.5% or > 10%, respectively, when the gene showed a significant over or under expression. Genes were ranked by their *P*-value for every analysis scoring a gene rank. Median rank is the median *P*-value rank across data sets, for each gene assessed.

Search criteria: We performed a search for each gene using its HGNC gene symbol as the search term. The resulting studies were analyzed on the basis of healthy prostate gland vs. prostate adenocarcinoma. Cited literature was reviewed to confirm that the analysis was as documented in the *Oncomine* database.

Information source and eligibility criteria (GEO: Gene Expression Omnibus). Gene expression data sets for primary PCa samples: To study the impact of the expression of the selected genes on the survival of patients, three data sets were selected according to the following criteria: (1) the study includes metadata for each patient, with ≥ 5 years of follow-up survival, (2) the study consists of ≥ 200 samples, and (3) the study is published and available on GEO (Gene Expression Omnibus).

Search/study selection:

- Sboner 2010* (GSE16560)²¹: a PCa patient's cohort that had undergone transurethral resection of prostate (TURP) or adenoma enucleation taken at the time of the initial diagnosis. It is comprised of 281 tumor tissue samples from men with PCa from the Swedish Watchful Waiting Cohort, with up to 30 years of clinical follow-up, with complete Illumina GPL5474 Human 6k Transcriptionally Informative Gene Panel data. A descriptive table regarding patient characteristics at baseline (start of the follow-up for survival analyses) is depicted in Supplementary Table 5.
- Ross-Adams 2015* (GSE70770) GPL10558 series²⁵: a PCa patient's cohort with 206 tumor tissue samples from men with PCa who had undergone radical prostatectomy and clinical follow-up of 9 years, including relapse information (biochemical relapse). Biochemical relapse was defined according to European Guidelines as a persistent rise of PSA above 0.2 ng/ml. Tumor sample expression of 31,000 transcripts was measured by 47,000 probes using the Illumina HumanHT-12 V4.0 platform. A descriptive table regarding patient characteristics at baseline (start of the follow-up for survival analyses) is depicted in Supplementary Table 6.
- Jenkins* (GSE10645) GPL5873 DASL Custom Prostate Panel²⁶: a PCa patient's cohort with 596 tumor tissue samples from men with PCa undergoing radical prostatectomy, with PCa specific death information and a mean follow-up of 20 years. A descriptive table regarding patient characteristics at baseline (start of the follow-up for survival analyses) is depicted in Supplementary Table 7.

Information source and eligibility criteria (cBioPortal). We searched the cBio Cancer Genomics Portal²⁷, an open-source cancer genomics data platform created by Memorial Sloan-Kettering Cancer Center (MSKCC), to analyze the selected genes' most common mutations, copy number alterations, and gene expression in PCa (11 data sets, n = 2820 samples, last access: December 2019). The criteria for inclusion of the data sets in our analysis were as follows: (1) type of cancer: prostate adenocarcinoma or metastasis, (2) the study must be published, and (3) the study must consist of a sample number > 50.

Information source and eligibility criteria (SU2C/PCF Dream Team 2019 data set (SU2C))³⁰. We used the data set from the SU2C-PCF Dream Team: Precision Therapy for Advanced Prostate Cancer that has whole-exome sequencing of 444 castrate-resistant prostate cancer tumor/normal pairs. A descriptive table regarding patient characteristics at baseline (start of the follow-up for analyses) is depicted in Supplementary Table 8.

Time-dependent ROC curves. TimeROC⁵¹ package was used for the estimation of time-dependent receiver operating characteristic (ROC) curve and area under time-dependent ROC curve (AUC) in the presence of censored data. Confidence intervals (CI) were computed using the iid-representation tool within the package. The confidence level was set to 0.95.

Statistics and reproducibility. We used the web-based bioinformatic tool Nexus⁵² for the analysis of differentially expressed genes based on the comparison of prostate tumor and normal adjacent tissue samples (*TCGA-PRAD*) (expressed as log₂ fold change).

GraphPad Prism software (La Jolla, CA, USA) was used to calculate student's *t*-test for testing differences in gene expression across tissue samples, and percentage of the genome altered in patients with high or low *YWHAZ* or 14-3-3ζ/δ expression. Statistical significance was set at *P* ≤ 0.05.

Stata software (StataCorp LLC, Texas, USA) was used to explore the patients' survival and to generate Kaplan–Meier curves. To find the cutoff value to stratify patients into two groups based on the expression levels of each gene, we used the Cutoff Finder tool⁵³. For univariable and multivariable analyses of prognostic

factors, log-rank test and Cox proportional hazard model regression were employed.

Fisher's exact test was used to test the statistical significance of contingency tables of genetic alterations. Statistical significance was set at $P \leq 0.05$.

One way ANOVA followed by a Tukey's test was performed to assess significant differences when comparing gene expression, and percentage of the genome altered across samples with copy number alterations.

Ethics statement. Written informed consent and institutional review board approval from the teaching hospital "Hospital de Clínicas José de San Martín" in Buenos Aires, Argentina, were acquired.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data sets generated during the current study are available in the ProteomeXchange (<http://www.proteomexchange.org>) repository, via the identifier PXD014291. Source data underlying plots shown in figures are provided in Supplementary Data 3. All other relevant data are available from the authors upon request.

Received: 11 June 2020; Accepted: 16 December 2020;

Published online: 22 January 2021

References

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Stark, J. R., Mucci, L., Rothman, K. J. & Adami, H.-O. Screening for prostate cancer remains controversial. *BMJ* **339**, b3601 (2009).
- Schroder, F. H. et al. Screening and prostate-cancer mortality in a randomized European study. *N. Engl. J. Med.* **360**, 1320–1328 (2009).
- Epstein, J. I., Feng, Z., Trock, B. J. & Piorozazio, P. M. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. *Eur. Urol.* **61**, 1019–1024 (2012).
- Davies, J. D. et al. Prostate size as a predictor of Gleason score upgrading in patients with low risk prostate cancer. *J. Urol.* **186**, 2221–2227 (2011).
- Kvåle, R. et al. Concordance between Gleason scores of needle biopsies and radical prostatectomy specimens: a population-based study. *BJU Int.* **103**, 1647–1654 (2009).
- Sanhueza, C. & Kohli, M. Clinical and novel biomarkers in the management of prostate cancer. *Curr. Treat. Options Oncol.* **19**, 8 (2018).
- Filella, X. et al. PCA3 in the detection and management of early prostate cancer. *Tumour Biol.* **34**, 1337–1347 (2013).
- Loeb, S. Biomarkers for prostate biopsy and risk stratification of newly diagnosed prostate cancer patients. *Urol. Pract.* **4**, 315–321 (2017).
- Knezevic, D. et al. Analytical validation of the oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genomics* **14**, 690 (2013).
- Ferro, M. et al. Prostate Health Index (Phi) and Prostate Cancer Antigen 3 (PCA3) significantly improve prostate cancer detection at initial biopsy in a total PSA range of 2–10 ng/ml. *PLoS ONE* **8**, e67687 (2013).
- Leyten, G. H. J. M. et al. Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *Eur. Urol.* **65**, 534–542 (2014).
- Kent, C. K. et al. Urinary PCA3 as a predictor of prostate cancer in a cohort of 3,073 men undergoing initial prostate biopsy. *J. Urol.* **191**, 1743–1748 (2014).
- Auprich, M. et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. *Eur. Urol.* **59**, 96–105 (2011).
- Hessels, D. et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. *Prostate* **70**, 10–16 (2010).
- Seisen, T. et al. Accuracy of the prostate health index versus the urinary prostate cancer antigen 3 score to predict overall and significant prostate cancer at initial biopsy. *Prostate* **75**, 103–111 (2015).
- Cantiello, F. et al. Prognostic accuracy of Prostate Health Index and urinary Prostate Cancer Antigen 3 in predicting pathologic features after radical prostatectomy. *Urol. Oncol. Semin. Orig. Invest.* **33**, 163.e15–163.e23 (2015).
- Cantiello, F. et al. PHI and PCA3 improve the prognostic performance of PRIAS and Epstein criteria in predicting insignificant prostate cancer in men eligible for active surveillance. *World J. Urol.* **34**, 485–493 (2016).
- The Cancer Genome Atlas. *Definitions*. <https://doi.org/10.32388/e1plqh> (2020).
- Rhodes, D. R. et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
- Sboner, A. et al. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genomics* **3**, 8 (2010).
- Murata, T. et al. 14-3-3zeta, a novel androgen-responsive gene, is upregulated in prostate cancer and promotes prostate cancer cell proliferation and survival. *Clin. Cancer Res.* **18**, 5617–5627 (2012).
- Ellen, T. P., Ke, Q., Zhang, P. & Costa, M. NDRG1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states. *Carcinogenesis* **29**, 2–8 (2008).
- Liu, H., Shui, I. M., Platz, E. A., Mucci, L. A. & Giovannucci, E. L. No association of ApoE genotype with risk of prostate cancer: a nested case-control study. *Cancer Epidemiol. Biomark. Prev.* **24**, 1632–1634 (2015).
- Ross-Adams, H. et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. *EBioMedicine* **2**, 1133–1144 (2015).
- Nakagawa, T. et al. A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS ONE* **3**, e2318 (2008).
- Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
- Mateo, J. et al. DNA repair in prostate cancer: biology and clinical implications. *Eur. Urol.* **71**, 417–425 (2017).
- El Gammal, A. T. et al. Chromosome 8p deletions and 8q gains are associated with tumor progression and poor prognosis in prostate cancer. *Clin. Cancer Res.* **16**, 56 LP–56 64 (2010).
- Abida, W. et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **116**, 11428–11436 (2019).
- Crone, S. G. et al. microRNA-146a inhibits G protein-coupled receptor-mediated activation of NF-kappaB by targeting CARD10 and COPS8 in gastric cancer. *Mol. Cancer* **11**, 71 (2012).
- Kypta, R. M. & Waxman, J. Wnt/beta-catenin signalling in prostate cancer. *Nat. Rev. Urol.* **9**, 418–428 (2012).
- Uehara, H., Takahashi, T. & Izumi, K. Induction of retinol-binding protein 4 and placenta-specific 8 expression in human prostate cancer cells remaining in bone following osteolytic tumor growth inhibition by osteoprotegerin. *Int. J. Oncol.* **43**, 365–374 (2013).
- Vocka, M. et al. Growth/differentiation factor 15 (GDF-15) as new potential serum marker in patients with metastatic colorectal cancer. *Cancer Biomark.* **21**, 869–874 (2018).
- De Ganck, A. et al. Down-regulation of myopodin expression reduces invasion and motility of PC-3 prostate cancer cells. *Int. J. Oncol.* **34**, 1403–1409 (2009).
- Ummanni, R. et al. Identification of clinically relevant protein targets in prostate cancer with 2D-DIGE coupled mass spectrometry and systems biology network platform. *PLoS ONE* **6**, e16833–e16833 (2011).
- Chen, S.-J. et al. Prostate-derived factor as a paracrine and autocrine factor for the proliferation of androgen receptor-positive human prostate cancer cells. *Prostate* **67**, 557–571 (2007).
- Venanzoni, M. C. et al. Apolipoprotein E expression in localized prostate cancers. *Int. J. Oncol.* **22**, 779–786 (2003).
- Symes, A. J. et al. Quantitative analysis of BTF3, HINT1, NDRG1 and ODC1 protein over-expression in human prostate cancer tissue. *PLoS ONE* **8**, e84295 (2013).
- Pilepich, M. V. et al. Correlation of radiotherapeutic parameters and treatment related morbidity—analysis of RTOG Study 77-06. *Int. J. Radiat. Oncol. Biol. Phys.* **13**, 1007–1012 (1987).
- McGowan, D. G. The adverse influence of prior transurethral resection on prognosis in carcinoma of prostate treated by radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **6**, 1121–1126 (1980).
- Forman, J. D. et al. The correlation of pretreatment transurethral resection of prostatic cancer with tumor dissemination and disease-free survival. A univariate and multivariate analysis. *Cancer* **58**, 1770–1778 (1986).
- Hanks, G. E., Leibel, S. & Kramer, S. The dissemination of cancer by transurethral resection of locally advanced prostate cancer. *J. Urol.* **129**, 309–311 (1983).
- Sharma, A. et al. The prostate metastasis suppressor gene NDRG1 differentially regulates cell motility and invasion. *Mol. Oncol.* **11**, 655–669 (2017).
- Song, Y. et al. N-myc downstream regulated gene-1/Cap43 may play an important role in malignant progression of prostate cancer, in its close association with E-cadherin. *Hum. Pathol.* **41**, 214–222 (2010).
- Nishimura, Y. et al. Overexpression of YWHAZ relates to tumor cell proliferation and malignant outcome of gastric carcinoma. *Br. J. Cancer* **108**, 1324–1331 (2013).
- Ruenaiver, K. et al. Prognostic significance of YWHAZ expression in localized prostate cancer. *Prostate Cancer Prostatic Dis.* **17**, 310–314 (2014).

48. Wakabayashi, M. et al. Phosphoproteome analysis of formalin-fixed and paraffin-embedded tissue sections mounted on microscope slides. *J. Proteome Res.* **13**, 915–924 (2014).
49. Hakimi, A. A. et al. An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **29**, 104–116 (2016).
50. Huang, D. W. et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183–R183 (2007).
51. Blanche, P., Dartigues, J.-F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).
52. Li, J.-R. et al. Cancer RNA-Seq Nexus: a database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res.* **44**, D944–D951 (2016).
53. Budczies, J. et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS ONE* **7**, e51862 (2012).

Acknowledgements

This work was supported by grants from AGENCIA-PICT 2016-1366 (ARGENTINA), AGENCIA-PICT 2016-0056 (ARGENTINA), UBACyT 2018-20020170100585BA (Argentina), and AGENCIA-PICT-RAICES-2018-02639 (Argentina).

Author contributions

Conception/design: G.G., J.C., S.L.-V., and J.B. Provision of reagents, and facilities: P.V., N.N., J.C., E.V., and G.G. Acquisition of data: S.L.-V., J.B., P.V., P.S., C.S., and O.M. Analysis/interpretation of data: S.L.-V., J.B., P.V., P.S., A.M., J.C., E.V., and G.G. Writing of the manuscript: S.L.-V., J.B., E.V., J.C., and G.G. Review of the manuscript: S.L.-V., J.B., P.V., P.S., EL, S.N., C.S., O.M., A.M., E.L., N.N., E.V., J.C., and G.G. Study supervision: E.V., J.C., and G.G.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-01645-2>.

Correspondence and requests for materials should be addressed to J.C. or G.G.

Reprints and permission information is available at <http://www.nature.com/reprints>

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) 2021