Unmasking Neuroendocrine Prostate Cancer with a Machine Learning-Driven 7-Gene Stemness Signature that Predicts Progression

Agustina Sabater[1,2,3], Pablo Sanchis[1,2,3], Rocio Seniuk[1,2], Gaston Pascual[1,2], Nicolas Anselmino[4], Daniel Alonso[5], Federico Cayol[6], Elba Vazquez[1,2], Marcelo Marti[1,2], Javier Cotignola[1,2], Ayelen Toro[1,2], Estefania Labanca[d], Juan Bizzotto[1,2,3], Geraldine Gueron[1,2]

[1] Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Química Biológica, Buenos Aires, Argentina. [2] CONICET-Universidad de Buenos Aires. Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Buenos Aires, Argentina. [3] Universidad Argentina de la Empresa (UADE), Instituto de Tecnología (INTEC), Buenos Aires C1073AAO, Argentina. [4] Department 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. [5] Centro de Oncología Molecular y Traslacional y Plataforma de Servicios Biotecnológicos, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal B1876BXD, Argentina. [6] Sector de Oncología Clínica, Hospital Italiano de Buenos Aires, Buenos Aires, C1199ABB, Argentina.

BACKGROUND: Prostate cancer (PCa) is a leading cause of cancer-related morbidity and mortality, mainly due to its progression into aggressive forms such as neuroendocrine prostate cancer (NEPC). Identifying reliable biomarkers that can predict disease progression and guide therapeutic options remains a significant clinical challenge. There are reports of enriched PCa stem-like cells upon neuroendocrine transdifferentiation. Thus, we sought to develop and validate a stemness-associated gene signature with prognostic potential and to evaluate if this signature can distinguish NEPC tumors, including the identification of large-cell neuroendocrine carcinoma, a very rare, aggressive and under-diagnosed subtype.

METHODS: We first analyzed differential gene expression of 144 stemness-associated genes across PCa comparisons, spanning normal/benign prostate, primary PCa, metastatic and castration-resistant samples (n=1,259) and performed survival analysis including disease progression, disease free-time, biochemical relapse, metastasis and death as endpoints (n=1,229). Further, we employed an ensemble and feature selection-based machine learning approach to analyze large-scale transcriptomic data from multiple cohorts and Lasso regression was used to model coefficients for a stemness-associated signature with prognostic value. We validated the resulting gene signature across independent datasets (n=501) by uni and multivariable survival analyses. We performed unsupervised clustering and Principal Component Analyses of PCa Patient-Derived Xenografts (PDXs) from MD Anderson Cancer Center (n=44) to establish associations between the 7-gene signature and clinicopathological features. The signature NEPC classifier performance was evaluated through ROC curves analysis. We then analyzed a PCa dataset (n=49) with NEPC histological classification to assess its classifier performance of Large-Cell NEPC samples by ROC curve analysis.

RESULTS: We observed dysregulated expression of 139 stemness-associated genes across multiple PCa comparisons (adjusted $p<0.05$) and a generalized association with survival-events (CoxP<0.05). Through machine learning feature selection and Lasso regression, we established a 7-gene stemness-associated signature (*KMT5C*, *MEN1*, *TYMS*, *IRF5*, *DNMT3B*, *CDC25B* and *DPP4*) which demonstrated a high prognostic value for progression-free, relapse-free, metastasis-free and overall survival across PCa datasets. This signature was further validated in additional datasets, and multivariable survival analyses showed the signature predicted survival events independently of other clinico-pathological variables (CoxP<0.05). In the MDA PCa PDX series, this stemness-associated signature was able to distinguish NEPC from other PCa subtypes, confirming its classifier potential (AUC=0.92). Accordingly, we observed high signature levels in NEPC PDXs and Large-Cell NEPC patient samples ($p<0.0001$ and $p<0.01$, respectively). Importantly, this signature particularly identified the aggressive large-cell NEPC subtype in a PCa patients' cohort (AUC=0.99).

CONCLUSIONS: The herein presented stemness-associated gene signature offers a clinically valuable tool for both prognosis and classification of PCa, particularly in identifying challenging NEPC subtypes like large-cell NEPC. Its consistent association with the risk of progression and robust validation across datasets and PDX models underscore its potential to guide personalized treatment strategies and improve clinical outcomes in patients with aggressive PCa.

FUNDING ACKNOWLEDGMENTS: This work was supported by Agencia Nacional de Promoción de la Investigación el Desarrollo Tecnológico y la Innovación (ANPCyT) PICT-RAICES-2021-III-A-00080; David H. Koch Center for Applied Research in Genitourinary Cancers at MD Anderson (Houston, TX); and NIH/NCI U01 CA224044.

CONFLICTS OF INTEREST DISCLOSURE STATEMENT: Authors declare no conflicts of interests.