Identification of protein biomarkers associated with prostate cancer risk using genetic prediction models: analysis of over 140000 subjects

Hua Zhong, PhD[1*]; Jingjing Zhu, PhD[1*]; Shuai Liu, PhD[1*]; Dalia H Ghoneim, PhD[1]; Praveen Surendran, PhD[2]; Tao Liu, PhD[3]; Sarah Fahle, PhD[2]; Adam Butterworth, PhD[2,4]; Md Ashad Alam, PhD[5]; Hong-Wen Deng, PhD[5]; Herbert Yu, MD, PhD[1]; Chong Wu, PhD[6#]; Lang Wu, PhD[1#]

[*]Co-first authors; [#]Co-senior authors

[1]Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, USA; [2]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA; [4]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [5]Tulane Center for Biomedical Informatics and Genomics, Division of Biomedical Informatics and Genomics, Deming Department of Medicine, Tulane University, New Orleans, LA, USA; [6]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Prostate cancer (PCa) is the second most frequently diagnosed malignancy among males. Identification of biomarkers is critical for understanding the pathogenesis of this common cancer. A growing number of conventional observational studies report associations of multiple circulating proteins with PCa risk. However, the existing findings may be subject to incoherent biases of conventional epidemiologic studies. To better characterize their associations, herein, we evaluated associations of genetically predicted concentrations of plasma proteins with PCa risk.

Leveraging genome and plasma proteome data of 2,481 healthy European descendants included in the INTERVAL study subcohort 1, we established models using four methods (Best Linear Unbiased Predictor, Least Absolute Shrinkage and Selection Operator, elastic net, and top1) to predict protein levels based on genetic variants. We further conducted model external validation using independent INTERVAL subcohort 2 dataset (N=820), and retained models with a prediction performance (R^2) of >0.01 in cross-validation and external validation for association analyses with PCa risk.

After testing 1,308 proteins in 79,194 cases and 61,112 controls of European ancestry included in the consortia of BPC3, CAPS, CRUK, PEGASUS, and PRACTICAL, 24 proteins showed significant associations with PCa risk, including 16 previously reported proteins and eight novel proteins. Of them, 14 proteins showed negative associations and 10 showed positive associations with PCa risk. For 18 of the identified proteins, potential functional somatic changes of encoding genes were detected in PCa patients in The Cancer Genome Atlas. Genes encoding these proteins were significantly involved in cancer-related pathways. We further identified drugs targeting the identified proteins, which may serve as candidates for drug repurposing for treating PCa.

In conclusion, this study identifies novel protein biomarker candidates for PCa risk, which may provide new perspectives on the etiology of PCa and improve its therapeutic strategies.