

# Validation of a 3-gene signature and development of an authentic cohort database to improve overall survival prediction and clinical treatment decision for patients with newly diagnosed prostate cancer

Zhuochun Peng<sup>1,7</sup>, Karl Andersson<sup>2,8</sup>, Johan Lindholm<sup>3</sup>, Olga Dethlefsen<sup>4</sup>, Yudi Pawitan<sup>5</sup>, Monica Nistér<sup>1,3</sup>, Sten Nilsson<sup>1,6</sup> and Chunde Li<sup>1,6,7\*</sup>

<sup>1</sup>Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden; <sup>2</sup>Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; <sup>3</sup>Clinical Pathology/Cytology, Karolinska University Hospital, Stockholm, Sweden. <sup>4</sup>Bioinformatics Bioinformatics Infrastructure for Life Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden. <sup>5</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>6</sup>Clinical Oncology, Karolinska University Hospital, Stockholm, Sweden; <sup>7</sup>Chundsell Medicals AB, Stockholm, Sweden; <sup>8</sup>Ridgeview Instruments AB, Uppsala, Sweden.

**Background:** since the majority of newly diagnosed prostate cancers are indolent and slowly progressing, a better prognostic tool is needed to improve survival prediction accuracy so as to make better treatment decisions at the time of diagnosis.

**Method:** a previously reported 3-gene signature (VGLL3, IGFBP3 and F3) was validated in improving survival time prediction at diagnosis when combined with clinical parameters such as: age at diagnosis, Gleason score, tumor stage and PSA in formalin fixed paraffin embedded prostate needle biopsy materials. A cohort-based algorithm, the classification of prostatic malignancy algorithm (CPMA), collected gene expression values of 3 genes and four clinical parameters derived from 452 historical patients with complete documented 7 to 11 year clinical follow-up. Patients were selected from a population based cohort (Stockholm area) and were diagnosed from 2004-2008. The actual overall, cancer-specific and non-cancer-specific survival time were compared with the corresponding estimations by the CPMA parameters, the NCCN risk groups and the CAPRA risk groups by using both the Weibull regression model and the CPMA algorithm.

**Results:** In Weibull regression model, AUC for overall survival prediction is significantly increased, from 0.68 for NCCN risk groups, 0.72 for CAPRA risk groups (0-2, 3-5 and 6-10) and up to 0.81 for CPMA parameters. For 234 individuals who did not treated with prostatectomy/radiation we performed survival prediction accuracy using either CAPRA risk score or CPMA algorithm. The overall survival prediction accuracy is 88% for CAPRA 0-2 group, 56 % for CAPRA 3-5 and 66% for CAPRA 6-10. Comparatively, the overall accuracy is increased to 88%, 76% and 83%, respectively, when using the CPMA parameters.

**Conclusion:** the CPMA parameters and its clinically implemented database algorithm (CPMA) significantly increased overall survival time prediction accuracy compared to the currently wide-used prognostic risk tools. It is justified to consider the use of CPMA in clinical practice.