



**Karolinska
Institutet**

Institutionen för onkologi-patologi

**Identification, validation and clinical
application of a three-gene signature for
accurate prognosis prediction and
treatment selection of newly diagnosed
prostate cancer**

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska
Institutet offentligen försvaras i lecture hall, CCK, R8:00

Onsdagen den 14 Okt, 2015, kl 09.30

av

Zhuochun Peng

Huvudhandledare:

MD, PhD, Chunde Li
Karolinska Institutet
Institutionen för onkologi-patologi

Bihandledare:

Professor Sten Nilsson
Karolinska Institutet
Institutionen för onkologi-patologi

Professor Monica Nistér
Karolinska Institutet
Institutionen för onkologi-patologi

Fakultetsopponent:

Professor Jack A. Schalken
Radboud University Nijmegen
Institutionen för Medical Center

Betygsnämnd:

Professor Annika Lindblom
Karolinska Institutet
Institutionen för molekylär medicin och
kirurgi

Docent Weng-Onn Lui
Karolinska Institutet
Institutionen för onkologi-patologi

Docent Annika Håkansson
Linköping University
Institutionen för klinisk och experimentell
medicin (IKE), kinisk onkologi

Stockholm 2015

ABSTRACT

This thesis presents a complete, comprehensive and stepwise approach of translational molecular research. Starting from whole-genome bioinformatics analyses based on the embryonic stem (ES) cell hypothesis, a three-gene signature was identified, validated with the goal of clinical application in order to optimize treatment decision based on improvement of overall survival estimation.

In Paper I, we hypothesized that gene signatures of embryonic stem (ES) cells may have prominent importance to determine the tumor subtypes and may be associated with the prognosis of various cancers including prostate cancer (PCa). Using published microarray datasets, 641 embryonic stem cell gene predictors (ESCGPs) were identified. Using gene expression patterns of these 641 ESCGPs tumor subtypes of different cancers can be stratified, particularly for prostate cancer. We further analyzed the gene expression levels of selected ESCGP genes in fresh-frozen fine needle aspiration biopsy samples taken from a Swedish cohort of 189 prostate cancer patients. The registry follow-up period for these patients was up to 18 years, where 97.9% patients had overall and cancer-specific survival data. As a result, a three-gene signature (VGLL3, IGFBP3 and F3) was identified sufficient to categorize the patients into high-risk, intermediate-risk and low-risk subtypes directly correlated with the overall and cancer-specific survival.

Currently, formalin-fixed paraffin embedded (FFPE) prostate core needle biopsy material is the most common sample material available in clinical practice, on which Gleason grading for prostate cancer diagnosis is usually conducted. Since each patient typically has multiple biopsy samples, and since Gleason grading is an operator dependent procedure known to be difficult, the impact of the operator's choice of biopsy was evaluated in paper II. We analyzed expression levels of the three-gene signature identified in paper I, using a four multiplex one-step RT-qPCR kit specially designed and optimized for measuring the three-gene expression signature in 127 FFPE prostate core needle biopsy samples taken from 43 patients. Our results show that the assessment of expression levels of two highly expressed genes (IGFBP3 and F3) in prostate cancer tissue is independent of Gleason patterns. These findings indicate that the impact of operator's choice of biopsy is low.

In paper III, we carried out a new cohort study including 241 prostate cancer patients with 6-9 years of registry follow up in order to verify the prognostic value of the three-gene expression signature in FFPE prostate core needle biopsy tissue samples. The cohort consisted of four patient groups with different survival times and cause of death (COD). We observed that supplementing readily available clinical data with gene expression levels of IGFBP3 and F3 in FFPE PCa biopsy tissues could improve survival prediction for PCa patients at time of diagnosis.

Based on the above work, a so-called Prostatype test system has been industrially designed and developed for clinical application. It integrates a robust multiplex RT-qPCR kit to measure expression levels of the three-gene signature and, a database of reference patients with accurate clinical documentation using a kNN-algorithm called CPMA (Classification of Prostatic Malignancy Algorithm). The survival prediction in relation to different treatment modalities can greatly assist both clinicians and patients to make an individualized treatment decision.

The flowchart in Figure 1 summarizes the present thesis.

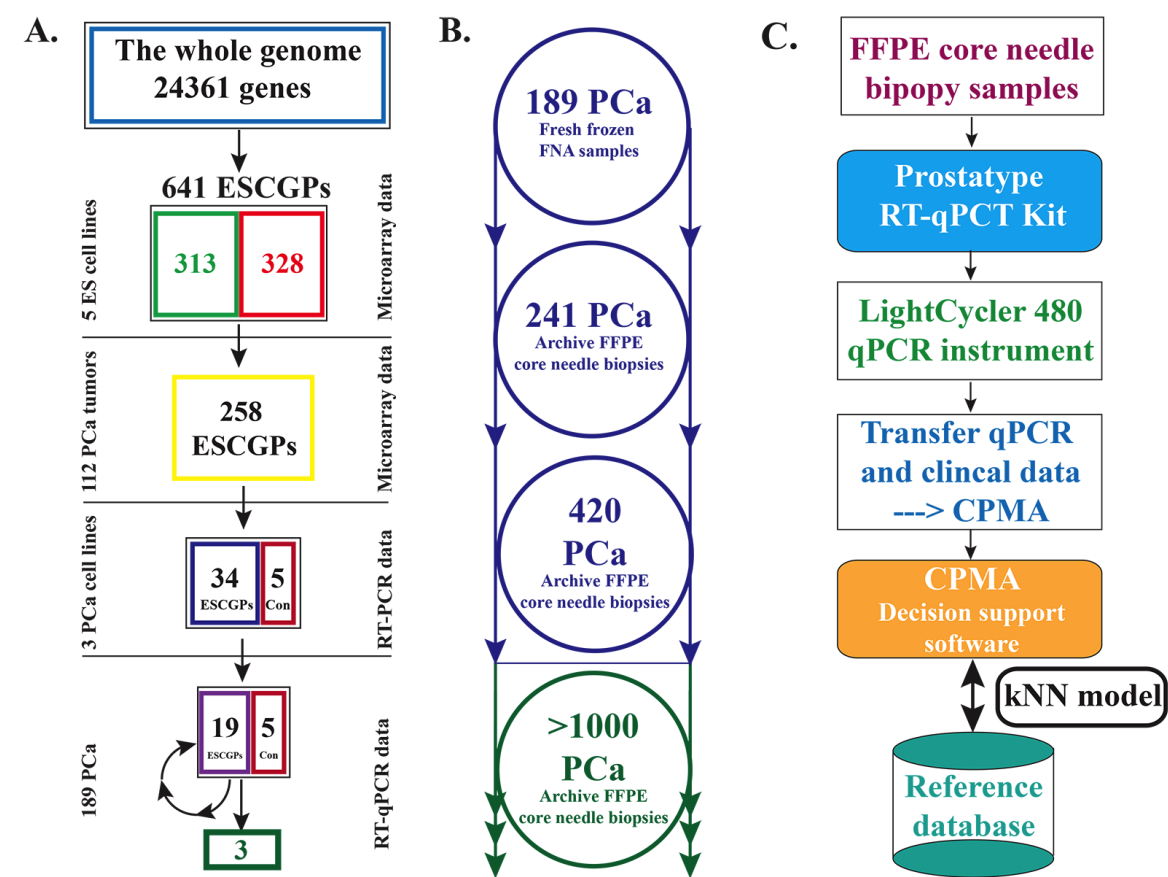


Figure 1. Graphic summary of the present thesis. A. A step-wise gene selection process, starting from bioinformatics analyses of the whole genome expression data of 24361 genes derived from 5 human embryonic stem cell lines, identified 641 Embryonic stem cell gene predictors (ESCGPs), until the identification of a three-gene signature. (Right vertical description: methodologies used in each step; left vertical description: materials used in each step of study). B. Cohorts represented in the order in which they were used for the pilot study (Paper I), the validation study (Paper II, III) and studies for filling CPMA reference database. The blue cycles are ready completed and ongoing cohort studies; the green cycle shows cohort studies planned in the near future. C. An industrially developed Prostatype Test System works according to the presented workflow. The system is composed of two parts: Prostatype RT-qPCR kit and CPMA software. It can provide prognostic statements assisting individualized/personalized estimation of overall survival time and decision of treatment for newly diagnosed prostate cancer patients.