

META-ANALYSIS OF GENE EXPRESSION AND SURVIVAL DATA USING THE GXA FRAMEWORK: A NEW PROGNOSIS TOOL FOR BREAST CANCER

Markus Schröder^{1,3}, Daniel Gusenleitner¹, Matthew Schwede¹, Aedín Culhane¹, Benjamin Haibe-Kains^{1,2}, Alexander Goesmann³, John Quackenbush^{1,2}

¹Computational Biology and Functional Genomics Laboratory, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA, United States; ²Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA, United States; ³Bioinformatics Research Facility, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, North-Rhine Westfalia, Germany

Introduction:

Gene expression profiling has generated unprecedented insight into our molecular understanding of cancer. In breast cancer, gene expression profiling studies have been widely employed and have not only advanced our understanding of disease, but have provided multi-gene predictive and prognostic tests including Oncotype DX, mammaprint, Veridex GGI and the Breast BioClassifier for breast cancer molecular subtypes.

In this interdisciplinary work we present a new web application that enables survival meta-analysis of single genes and gene signatures. The target user for this web application are the clinicians and biologists who will be able to assess the prognostic value of genes of interest in breast cancer, either for one gene at a time or by uploading gene lists in an intuitive and easy way.

Objective:

This application combines state-of-the-art tools for breast cancer prognostication (R packages: *survcomp*, *genefu*) with a large compendium of manually curated breast cancer microarray gene expression data within the framework of the Gene Expression Atlas (GXA). The GXA framework provides a powerful java front-end for displaying results on a website and enables standardized access to hundreds of fully annotated cancer gene expression datasets. We create an index for each patient in each dataset to allow comparative analysis of the gene expression data from different platforms even across different technologies. We have developed a pipeline for analysis of the prognostic value of genes across these datasets and use the GXA framework to display these results.

Implementation:

We present a pilot of this application focusing on currently available breast cancer datasets with survival data. Future plans involve the expansion to all available types of cancer where gene expression data with survival data is available.