Biomarker databases in Bladder Cancer

Garrett Dancik

(This would lead to the project: Development of an Online Tool and Bladder Cancer Gene Expression Database for Biomarker Evaluation)

Note: this literature review uses a different format for references. You must use the IEEE format as described in class

Literature Review

Cancer is a genetic disease whose biology is driven by mutated and abnormally expressed genes. Gene expression profiling, which simultaneously measures the expression levels of thousands of genes, is consequently a powerful tool for investigating cancer. In the past ~15 years, the identification of diagnostic and prognostic biomarkers from gene expression data has increased our understanding of cancer biology and has led to advances in the personalized treatment of cancer. A *diagnostic* biomarker is a molecule that is indicative of cancer diagnosis or existence, such as the stage, grade, and clinical subtype of a tumor; a *prognostic* biomarker is indicative of disease outcome. Although gene expression profiles of bladder cancer patients are publicly available, their analysis is time-consuming and requires computational resources and bioinformatics expertise often not available to biologists or clinician-researchers.

Cancer is a genetic disease (Stratton et al., 2009). A cancer cell inherits or acquires mutations that enable it to grow efficiently, replicate indefinitely, support angiogenesis, avoid apoptosis, and in some cases metastasize (Hanahan and Weinberg, 2011). In the past ~15 years, gene expression profiling of human cancers has revolutionized our understanding of cancer as a genetic disease and has expedited the identification of driver mutations and biomarkers for personalized treatment. Examples of prognostic biomarkers in routine clinical use include the Onco*type*Dx and MammaPrint gene panels, which both predict the likelihood of disease recurrence in breast cancer and provide patients and clinicians with relevant information regarding the potential benefit of chemotherapy (Knauer et al., 2009; Markopoulos et al., 2011).

In the United States, bladder cancer is the fourth most common cancer in males, the eighth most common cancer in females (Siegel et al., 2012), and one of the most expensive cancers to treat (Botteman et al., 2003). At diagnosis, approximately 20-30% of bladder cancer patients harbor muscle-invasive (MI) tumors (Jacobs et al., 2010) and these patients have a five-year survival rate of approximately 43% (Stein et al., 2001). In patients harboring non-muscle invasive (NMI) tumors, progression to MI disease occurs in ~ 20% of all patients, and in ~50% of high risk patients with high grade, recurrent tumors (Cookson et al., 1997). Despite the importance of this disease there are no prognostic or predictive biomarkers in clinical use.

Gene expression datasets are typically deposited into public databases such as the Gene Expression Omnibus (GEO; Barrett and Edgar, 2006) and ArrayExpress (Rustici et al., 2013). These databases function primarily as data repositories where data can be downloaded and analyzed using in-house bioinformatics tools. In particular, GEO includes some curated datasets with pre-computed sets of differentially expressed genes, but does not include curated datasets for bladder cancer patients. Other databases include the KM plotter (Gyorffy et al., 2010), which generates Kaplan-Meier survival curves for lung, breast, and ovarian cancer patients only; and

PrognoScan (Mizuno et al., 2009) and SurvExpress (Aguirre-Gamboa et al., 2013), which generate Kaplan-Meier curves for only a small subset of the available bladder cancer datasets, do not evaluate diagnostic biomarkers, and do not allow for patient subset selection (e.g., of patients with MI tumors only, an important clinical subgroup). However, important questions are of the form *is the gene TP53 a biomarker of disease outcome in bladder cancer patients with MI tumors?* Currently, an answer to this clinically important type of question cannot be answered directly using any of the existing databases.

Bibliography

Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, et al. (2013) SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS One* 8: e74250.

Botteman MF, Pashos CL, Redaelli A, et al. (2003) The health economics of bladder cancer: a comprehensive review of the published literature. *Pharmacoeconomics* 21: 1315-1330.

Cookson MS, Herr HW, Zhang ZF, et al. (1997) The treated natural history of high risk superficial bladder cancer: 15-year outcome. *J Urol* 158: 62-67.

Gyorffy B, Lanczky A, Eklund AC, et al. (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 123: 725-731.

Hanahan D and Weinberg RA. (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674.

Jacobs BL, Lee CT and Montie JE. (2010) Bladder cancer in 2010: how far have we come? *CA Cancer J Clin* 60: 244-272.

Knauer M, Straver M, Rutgers E, et al. (2009) The 70-gene MammaPrint signature is predictive for chemotherapy benefit in early breast cancer. *Breast* 18: S36-S37.

Markopoulos C, Xepapadakis G, Venizelos V, et al. (2011) Clinical Use of OncotypeDX Recurrence Score as an Adjuvant-Treatment Decision Tool in Early Breast Cancer Patients. *European Journal of Cancer* 47: S379-S379.

Mizuno H, Kitada K, Nakai K, et al. (2009) PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics* 2: 18.

Rustici G, Kolesnikov N, Brandizi M, et al. (2013) ArrayExpress update--trends in database growth and links to data analysis tools. *Nucleic Acids Res* 41: D987-990.

Siegel R, Naishadham D and Jemal A. (2012) Cancer statistics, 2012. *CA Cancer J Clin* 62: 10-29.

Stein JP, Lieskovsky G, Cote R, et al. (2001) Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J Clin Oncol* 19: 666-675.

Stratton MR, Campbell PJ and Futreal PA. (2009) The cancer genome. *Nature* 458: 719-724.