Master Thesis Proposal

Meta-analysis of gene expression and survival data using the GXA framework: a new prognosis tool for breast cancer

Markus Schröder, B.Sc.
Genome-Based Systems Biology
University of Bielefeld

mschroed@cebitec.uni-bielefeld.de
mschroed@jimmy.harvard.edu

March 10, 2011

First Supervisor: Aedin Culhane, PhD
Second Supervisor: Benjamin Haibe-Kains, PhD
Third Supervisor: Dr. Alexander Goesmann

Contents

1 Introduction

This interdisciplinary work concerns the development of a web application that enables survival meta-analysis of single genes and gene signatures. In particular, we are creating an index for each patient in each dataset that allows us to compare the gene expression data from different platforms even across different technologies. The target users 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.

Cancer is one of today's major public health issues. The number of reported cases has significantly increased over the last years. In particular, the most frequently diagnosed malignancy in women in the Western world is breast cancer [1]. In order to reduce the mortality rate of affected cancer patients, several early diagnosis protocols for different cancers have been introduced to the public and are being used worldwide [2]-[5]. Guidelines for quality assurance in breast cancer screening and diagnosis in Europe have been introduced [6]. As a result of this extensive effort in the fight against cancer, the mortality rate of affected patients is decreasing and the survival time is increasing due to targeted therapy or surgery [7], all part of our current efforts in *personalized medicine* [8]-[11].

Of the many new biotechnologies that have been introduced over the last decade, one of the most influential is the microarray-based technology of genome-wide gene expression profiling. This new tool enables the study of diseases at the molecular level in order to get new insights into cancer biology for improving current cancer management.

Two of the main issues in breast cancer are development of prognosis and prediction of response/resistance to therapy. The first issue, prognostication, is the prediction of patient survival or the risk of the development of metastases independently of treatment. Accurate prognoses provide insights into the need for a treatment that might have adverse side effects. If a patient has a good prognosis, clinicians may consider not giving any treatment or a non-aggressive one. The second issue, prediction, is the ability to anticipate the response or resistance of a patient to an anti-cancer treatment. Accurate predictions of survival time and chemotherapy resistance give clinicians the ability to select the most suitable treatment available for a patient.

The objective of this thesis is to develop a web application that enables survival meta-analysis of single genes and gene signatures using state-of-the-art prognostic and predictive tools using molecular data from breast cancer patients generated by high throughput technologies.

2 Breast Cancer

Breast cancer is a global public health issue. It is one of the most frequently diagnosed malignancy in women in the Western world and together with lung cancer the commonest cause of cancer death in European and American women. According to estimates in 2002, there were 1,151,298 new cases of breast cancer diagnosed, 410,712

deaths caused by breast cancer, and more than 44 million women living with breast cancer worldwide [12]. In Europe, one out of eight to ten women, depending on the country, will develop breast cancer during her lifetime [13].

Thanks to the routine use of screening mammograms in developed countries, more and more women diagnosed with breast cancer are detected at an early stage (early breast cancer, small tumors and absence of lymph node invasion). Surgery is the primary treatment in the majority of cases, alone or in combination with radiotherapy. Despite early detection, up to half of these women will develop *distant metastasis*, i.e. development of new tumors in different organs. Metastatic breast cancer is unfortunately incurable. As a result, since the mid 1980s, randomized trials of adjuvant systemic therapy have been conducted in an effort to reduce the rate of recurrence and to prolong the survival of patients with operable disease [14].

Due to the importance of breast cancer for public health, this field has been the subject of intense research for decades. Moreover, new high throughput technologies, such as gene expression profiling, became readily available at the end of the 1990s, providing powerful tools to study and fight this disease.

3 Gene Expression Profiling

Messenger RNA (mRNA) is a molecule of RNA that is transcribed from DNA and then decoded by the ribosome into a specific amino acid chain or polypeptide which later folds into an active protein. Gene expression profiling is the indirect measurement of the concentration of that mRNA in a biological sample, typically tissues or cells.

Once a genome was sequenced, the sequence assembled and coding regions identified, the knowledge of what the cell could possibly do is available. More interesting, however, is the knowledge of what the cell is doing at a specific point in time, at a specific stress situation or how its gene expression differs from infected or malfunctioning cells. Gene expression profiling can provide some of that knowledge. It measures the activity of thousands of genes simultaneously in order to give an overview of the current functionality of the cell. In the field of breast cancer prognostication, these measurements can allow researchers to, for example, identify malignant vs. non-malignant types of tumor or if a patient has a low- or high-risk type of cancer. Gene expression profiling has been widely applied in breast cancer; gene profiles of breast tumor tissue are available in the Gene Expression Omnibus database (GEO) (89 datasets, [15]) and Array Express ([16]). A previous study of breast cancer from Perou et al. [17] identified 4 molecular subtypes in breast cancer, ER+/luminal-like, basal-like, Erb-B2+ and normal breast, which enable more specific and successful treatment for patients strongly associated to a single molecular subtype through their gene expression profile.

4 State-of-the-Art

A the beginning of this thesis, very few tools or websites addressing the same issue are available. In particular none of them are controlling for the presence of multiple

molecular subtypes, which are well-established in breast cancer research [18]-[20]. The following two websites are representative of the current tools available.

4.1 PrognoScan

PrognoScan: a new database for meta-analysis of the prognostic value of genes was published in April 2009 by Mizuno et al. from the Pharmaceutical Technology Department, Kamakura Research Laboratories in Kamakura, Japan [21].

The database (http://gibk21.bse.kyutech.ac.jp/PrognoScan/index.html) is publicly accessible and provides a powerful platform for evaluating potential tumor markers and therapeutic targets and would accelerate cancer research.

PrognoScan uses a large collection of publicly available cancer microarray datasets with clinical annotation; it is a tool for assessing the biological relationship between gene expression and prognosis. PrognoScan employs the minimum P-value approach for grouping patients for survival analysis. This approach finds the optimal cutpoint in continuous gene expression measurement without prior biological knowledge or assumption and, as a result, enables systematic meta-analysis of multiple datasets.

Pros: 55 datasets (19 for breast cancer), 12 different cancers (Bladder, Blood, Breast, Colorectal, Esophagus, Glioma, Head and Neck, Lung, Ovarian, Prostate, Renal cell carcinoma and Skin cancer).

Cons: This pioneer work was the first web application published to perform survival analysis in multiple gene expression dataset. However it lacks of a proper meta-analytical framework to efficiently combine statistics computed in different datasets. From a breast cancer perspective: it does not control for the presence of diverse molecular subtypes that could potentially confound the analysis results. PrognoScan requires an arbitrary binarization of the gene expression, in each dataset separately, making strong assumptions about the population of breast cancer patients used to generate those datasets.

4.2 KMPlot

Kaplan-Meier Plotter (KMPlot, http://kmplot.com) is an online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using Affymetrix microarray data of 1,809 patients. It was published in December 2009 from Győrffy et al. from the Joint Research Laboratory of the Hungarian Academy of Sciences and the Semmelweis University, Budapest, Hungary [22].

Their aim was to develop an online tool to draw survival plots, which can be used to assess the relevance of the expression levels of various genes on the clinical outcome of both untreated and treated breast cancer patients.

A background database was established using gene expression data and survival information of 1,809 patients downloaded from GEO (HGU133A and HGU133plus2 microarrays from Affymetrix). The median relapse-free survival is 6.43 years, 968/1,231 patients are estrogen receptor (ER) positive, and 190/1,369 are lymph-node positive. After quality control and normalization only probes present on both Affymetrix platforms were retained (n = 22,277).

In order to analyze the prognostic value of a particular gene, the cohorts are divided into two groups according to the median (or upper/lower quartile) expression of the gene. The two groups can be compared in terms of relapse free survival, overall survival, and distant metastasis free survival. A survival curve is displayed, and the hazard ratio with 95% confidence intervals and logrank P value are calculated and displayed.

Additionally, three subgroups of patients can be assessed: systematically untreated patients, endocrine-treated ER positive patients, and patients with a distribution of clinical characteristics representative of those seen in general clinical practice in the US. The tool is highly valuable for the preliminary assessment of biomarkers, especially for research groups with limited bioinformatics resources.

Pros: 14 breast cancer datasets. KM survival curves are easy to interpret.

Cons: Again, this tool does not control for the presence of breast cancer molecular subtypes. Although it reports an overall statistics about the significance of the association between a gene and patients survival, it provides only one effect-size estimate (hazard ratio), ignoring the numerous performance criteria used in state-of-the-art survival analysis. KMPlot requires an arbitrary binarization of the gene expression, in each dataset separately, making strong assumptions about the population of breast cancer patients used to generate those datasets. Lastly, KMPlot is limited to analysis of datasets generated from Affymetrix technology, which represents only a portion of what is publicly available to date.

5 A novel approach for breast cancer prognostication: OncoSurf

OncoSurf is going to be a web application combining the current state-of-the-art tools for breast cancer prognostication and currently available breast cancer microarray gene expression data with the use of the R packages *survcomp* [23] and *genefu* [24], and the framework of the Gene Expression Atlas (GXA) [25].

- R packages: The combination of the survcomp and genefu packages, developed by Benjamin Haibe-Kains, research fellow in John Quackenbush's lab, enables survival analysis of breast cancer gene expression data in a meta-analytical framework. In particular, the survcomp package contains functions to compute numerous performance criteria introduced recently to estimate the accuracy of survival models (also called risk prediction model in breast cancer prognostication), and to statistically compare the accuracy of such models. The genefu package is built upon survcomp to actually build risk prediction models from gene expression data and is focused on breast cancer, allowing the robust identification of molecular subtypes and the computation of several published gene signatures.

- GXA: The Gene Expression Atlas, maintained by the European Bioinformatics Institute, is an added-value database providing information about gene expression in different cell types, organism parts, developmental stages, disease states, sample treatments and other biological/experimental conditions. The content of

this database derives from curation, re-annotation and statistical analysis of selected data from the ArrayExpress Archive of Functional Genomics Data. The structure of GXA has three layers: R Analytics, Database, Front End. It includes a pipeline to port data from ArrayExpress/GEO to GXA. Each dataset is annotated with a standard experimental factor ontology (EFO). GXA is open source and programmatic access is available through their REST (Representational State Transfer) API. This API allows advanced users to search and retrieve complete information on any gene or experiment from the Atlas, including all gene and sample attributes, details of experimental design, meta-analysis statistics and gene expression values.

For the purpose of this thesis we concentrate on the field of breast cancer and selected datasets from this field that contain over 1400 patients in total. The datasets are called *mainz*[26], *transbig*[27], *upp*[28], *unt*[29], *vdx*[30][31] and *nki*[32][33]. However, OncoSurf is not limited to the field of breast cancer. Future plans involve the integration of all available types of cancer where gene expression data with survival data is available.

The results of the analysis pipeline behind the web application include several estimation values and plots.

- *The concordance index* [34] computes the probability that, for a pair of randomly chosen comparable patients, the patient with the higher risk prediction will experience an event before the lower risk patient. A score above 0.5 predicts low risk for the patient, a score below 0.5 indicates that the patient is in the high risk group.

- *The time-dependent ROC curve* [36] is a standard technique for assessing the performance of a continuous variable for binary classification. A ROC curve is a plot of sensitivity versus 1-specificity for all the possible cutoff values of the continuous variable. In survival analysis, the continuous variable is the risk score and the binary class to predict is the event occurrence. As the event occurrence is time-dependent, time-dependent ROC curves are more appropriate than conventional ones.

- *The Brier score* [35] is defined as the squared difference between an event occurrence and its predicted probabilities at time t. Probabilities of event can be derived from Cox's proportional hazards model fitted with the risk score or risk group predictions. Intuitively, if a patient experiences no event at time t, the event predicted probability should be close to zero. Symmetrically, if the patient experiences an event the probability should be close to one.

- *The hazard ratio* is a summary of the risk difference between several survival curves estimated by Cox's proportional hazards model [37]. Cox's model assumes that the relative risk of event between groups is constant at each interval of time.

- *The D index* [38] for a risk prediction is, i.e. an estimate of the log hazard ratio comparing two equal-sized prognostic groups. This is a natural measure of separation between two independent survival distributions under the proportional

hazards assumption. The D index is computed using the Cox model fitted on the scaled rankits of the risk scores instead of the risk scores themselves. The scaled rankits are the expected standard Normal order statistics scaled by 'kappa = sqrt(8/pi)'.

- *The forest plot* [39] is used to represent the concordance indices, the hazard ratios and D.indices for all the risk prediction methods.

The main task of this thesis is to create the web application OncoSurf using the GXA framework. We adapt the three layers from this framework mentioned above. In detail, this includes the R analytical part: the survival meta-analysis pipeline using state-of-the-art prognostication and prediction tools using the underlying datasets; the data storage: including standardized formats, expression data, annotation and patient data; and the web front-end: allowing clinicians and biologists to assess the prognostic value of single genes or gene signatures in an easy way and simultaneously receive several estimation values that manifest the results.

The work in this thesis is separated into different parts. The first step is to create an R analytical pipeline for survival analysis. This includes optimizing the concordance index (cindex) function in the survcomp package. Computing the cindex for a test set of 20 patients in a dataset containing 588 patients (*nki* and *unt* combined) takes around 20 minutes on an Apple Macbook Pro. This is prohibitively slow because the necessary genome-wide computation of the cindex for each patient would consume days; this is caused by poor management of nested for-loops in R. The ideal solution for this problem would be the relocation of those nested for-loops to a programming language that is more efficient. We made use of the R extensions support for calling C code, and re-wrote those loops within the cindex function in that language. This led to an average computation time improvement of factor 1250, enabling us to compute the cindex for over 1400 patients genome-wide in under 2 minutes.

The second step is to transform the currently available breast cancer datasets into a standard format for expression data, in order to make future integrations of other datasets as simple as possible. To accomplish this task we adopted the ExpressionSet object from the Biobase package from Bioconductor [40]. This leads to a uniform way of accessing the data. Although over 30 breast cancer datasets with survival data are available, the pilot of OncoSurf will initially include six breast cancer datasets.

The third step is, as mentioned above, integrating the R analytical results into a web application that incorporate the GXA framework. The OncoSurf web front end will utilize and extend exisiting java code within GXA. The web application is going to be integrated into the pool of currently available tools from the EBI, accessible through their main website, including further maintainance from the EBI.

The deliverables of this work are going to be the OncoSurf web application available through the EBI website. A publication in form of an applications note in the Bioinformatics Journal describing the functionality of the survcomp package, and a publication describing the OncoSurf web application. As part of the upgrade to the survcomp package, we will publish it in Bioconductor with detailed descriptions and usecases. The six datasets used in the OncoSurf pilot are also going to be published in Bioconductor.

References

[1] Peter Boyle and Bernard Levin **World Cancer Report 2008** http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/index.php

[2] Smith, R. A., V. Cokkinides, and O. W. Brawley. **Cancer Screening in the United States, 2008: A Review of Current American Cancer Society Guidelines and Cancer Screening Issues.** CA: A Cancer Journal for Clinicians 58, no. 3 (5, 2008): 161-179

[3] Whitlock, Evelyn P., Jennifer S. Lin, Elizabeth Liles, Tracy L. Beil, and Rongwei Fu. **Screening for Colorectal Cancer: A Targeted, Updated Systematic Review for the U.S. Preventive Services Task Force.** Annals of Internal Medicine 149, no. 9 (November 4, 2008): 638 -658

[4] Winawer, Sidney J. **Colorectal cancer screening.** Best Practice & Research Clinical Gastroenterology 21, no. 6 (December 2007): 1031-1048

[5] Mulshine, James L., and Rob J. van Klaveren. **Lung cancer screening: What is the benefit and what do we do about it?.** Lung Cancer 71, no. 3 (March 2011): 247-248

[6] Perry, N., M. Broeders, C. de Wolf, S. Törnberg, R. Holland, and L. von Karsa. **European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition-summary document.** Annals of Oncology 19, no. 4 (April 1, 2008): 614 -622

[7] J. Ferlay, D.M. Parkin, and E. Steliarova-Foucher, **Estimates of cancer incidence and mortality in Europe in 2008.** European Journal of Cancer 46, no. 4 (March 2010): 765-781

[8] Ng, Pauline C., Sarah S. Murray, Samuel Levy, and J. Craig Venter. **An agenda for personalized medicine.** Nature 461, no. 7265 (October 8, 2009): 724-726

[9] Schilsky, Richard L. **Personalized medicine in oncology: the future is now.** Nat Rev Drug Discov 9, no. 5 (May 2010): 363-366

[10] Weston, Andrea D., and Leroy Hood. **Systems Biology, Proteomics, and the Future of Health Care: Toward Predictive, Preventative, and Personalized Medicine.** Journal of Proteome Research 3, no. 2 (April 1, 2004): 179-196

[11] Hamburg, Margaret A., and Francis S. Collins. **The Path to Personalized Medicine.** New England Journal of Medicine 363, no. 4 (7, 2010): 301-304

[12] Veronesi, Umberto, Peter Boyle, Aron Goldhirsch, Roberto Orecchia, and Giuseppe Viale. **Breast cancer.** The Lancet 365, no. 9472 (May 14, 2005): 1727-1741

[13] Parkin, D. Maxwell, Freddie Bray, Jacques Ferlay, and Paola Pisani. **Estimating the world cancer burden: Globocan 2000.** International Journal of Cancer 94, no. 2 (10, 2001): 153-156

[14] EBCTG. **Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.** The Lancet 365, no. 9472 (May 14, 2005): 1687-1717

[15] Barrett, T., D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, et al. **NCBI GEO: archive for functional genomics data sets–10 years on.** Nucleic Acids Research 39, no. Database (11, 2010): D1005-D1010.

[16] Parkinson, Helen, Ugis Sarkans, Nikolay Kolesnikov, Niran Abeygunawardena, Tony Burdett, Miroslaw Dylag, Ibrahim Emam, et al. **ArrayExpress update–an archive of microarray and high-throughput sequencing-based functional genomics experiments.**Nucleic Acids Research 39, no. Database issue (January 2011): D1002-1004.

[17] Perou, Charles M., Therese Sorlie, Michael B. Eisen, Matt van de Rijn, Stefanie S. Jeffrey, Christian A. Rees, Jonathan R. Pollack, et al. **Molecular portraits of human breast tumours.** Nature 406, no. 6797 (2000): 747-752

[18] Sorlie, Therese, Charles M. Perou, Robert Tibshirani, Turid Aas, Stephanie Geisler, Hilde Johnsen, Trevor Hastie, et al. **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** Proceedings of the National Academy of Sciences of the United States of America 98, no. 19 (2001): 10869 -10874

[19] van 't Veer, Laura J., Hongyue Dai, Marc J. van de Vijver, Yudong D. He, Augustinus A. M. Hart, Mao Mao, Hans L. Peterse, et al. **Gene expression profiling predicts clinical outcome of breast cancer.** Nature 415, no. 6871 (January 31, 2002): 530-536

[20] Sorlie, Therese, Robert Tibshirani, Joel Parker, Trevor Hastie, J. S. Marron, Andrew Nobel, Shibing Deng, et al. **Repeated observation of breast tumor subtypes in independent gene expression data sets.** Proceedings of the National Academy of Sciences of the United States of America 100, no. 14 (July 8, 2003): 8418 -8423

[21] Mizuno, Hideaki, Kunio Kitada, Kenta Nakai, and Akinori Sarai.**PrognoScan: a new database for meta-analysis of the prognostic value of genes.** BMC Medical Genomics 2, no. 1 (2009): 18

[22] Győrffy, Balazs, Andras Lanczky, Aron C. Eklund, Carsten Denkert, Jan Budczies, Qiyuan Li, and Zoltan Szallasi. **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 Research and Treatment 123, no. 3 (12, 2009): 725-731

[23] **survcomp: Performance Assessment and Comparison for Survival Analysis** http://cran.r-project.org/web/packages/survcomp/

[24] **genefu: Relevant Functions for Gene Expression Analysis, Especially in Breast Cancer** http://cran.r-project.org/web/packages/genefu/

[25] Kapushesky, M., I. Emam, E. Holloway, P. Kurnosov, A. Zorin, J. Malone, G. Rustici, E. Williams, H. Parkinson, and A. Brazma. **Gene Expression Atlas at the European Bioinformatics Institute.** Nucleic Acids Research 38, no. Database (11, 2009): D690-D698

[26] Schmidt, Marcus, Daniel BÃ¶hm, Christian von TÃ¶rne, Eric Steiner, Alexander Puhl, Henryk Pilch, Hans-Anton Lehr, Jan G. Hengstler, Heinz KÃ¶lbl, and Mathias Gehrmann. **The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast Cancer.** Cancer Research 68, no. 13 (July 1, 2008): 5405 -5413

[27] Desmedt, Christine, Fanny Piette, Sherene Loi, Yixin Wang, FranÃ§oise Lallemand, Benjamin Haibe-Kains, Giuseppe Viale, et al. **Strong Time Dependence of the 76-Gene Prognostic Signature for Node-Negative Breast Cancer Patients in the TRANSBIG Multicenter Independent Validation Series.** Clinical Cancer Research 13, no. 11 (June 1, 2007): 3207 -3214

[28] Miller, Lance D., Johanna Smeds, Joshy George, Vinsensius B. Vega, Liza Vergara, Alexander Ploner, Yudi Pawitan, et al. **An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.** Proceedings of the National Academy of Sciences of the United States of America 102, no. 38 (2005): 13550 -13555

[29] Sotiriou, Christos, Pratyaksha Wirapati, Sherene Loi, Adrian Harris, Steve Fox, Johanna Smeds, Hans Nordgren, et al. **Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis.** Journal of the National Cancer Institute 98, no. 4 (February 15, 2006): 262 -272

[30] Wang, Yixin, Jan GM Klijn, Yi Zhang, Anieta M Sieuwerts, Maxime P Look, Fei Yang, Dmitri Talantov, et al. **Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.** The Lancet 365, no. 9460 (February 19, 2005): 671-679

[31] Minn, Andy J., Gaorav P. Gupta, David Padua, Paula Bos, Don X. Nguyen, Dimitry Nuyten, Bas Kreike, et al. **Lung metastasis genes couple breast tumor size and metastatic spread.** Proceedings of the National Academy of Sciences 104, no. 16 (April 17, 2007): 6740 -6745

[32] van 't Veer, Laura J., Hongyue Dai, Marc J. van de Vijver, Yudong D. He, Augustinus A. M. Hart, Mao Mao, Hans L. Peterse, et al. **Gene expression profiling predicts clinical outcome of breast cancer.** Nature 415, no. 6871 (January 31, 2002): 530-536

[33] van de Vijver, Marc J, Yudong D He, Laura J van't Veer, Hongyue Dai, Augustinus A M Hart, Dorien W Voskuil, George J Schreiber, et al. **A gene-expression signature as a predictor of survival in breast cancer.** The New England Journal of Medicine 347, no. 25 (December 19, 2002): 1999-2009

[34] Harrel Jr, F. E., Lee, K. L., Mark, D. B., et al. **Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error.** Statistics in Medicine 15, 361-387

[35] Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M. **Assessment and comparison of prognostic classification schemes for survival data.** Statistics in Medicine 18, 2529-2545

[36] Swets, J A. **Measuring the accuracy of diagnostic systems.** Science (New York, N.Y.) 240, no. 4857 (June 3, 1988): 1285-1293

[37] Therneau, Terry M., and Patricia M. Grambsch. **Modeling survival data: extending the Cox model.** Springer, 2000

[38] Royston, Patrick, and Willi Sauerbrei. **A new measure of prognostic separation in survival data.** Statistics in Medicine 23, no. 5 (March 15, 2004): 723-748

[39] Steff Lewis and Mike Clarke. **Forest plots: trying to see the wood and the trees.** BMJ 322, no. 7300 (June 16, 2001): 1479 -1480.

[40] Gentleman, Robert C, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, et al. **Bioconductor: open software development for computational biology and bioinformatics.** Genome Biology 5, no. 10 (2004): R80