Novel Time Series Analysis of a Diffuse Large B-cell Lymphoma Treatment

BOSTON UNIVERSITY

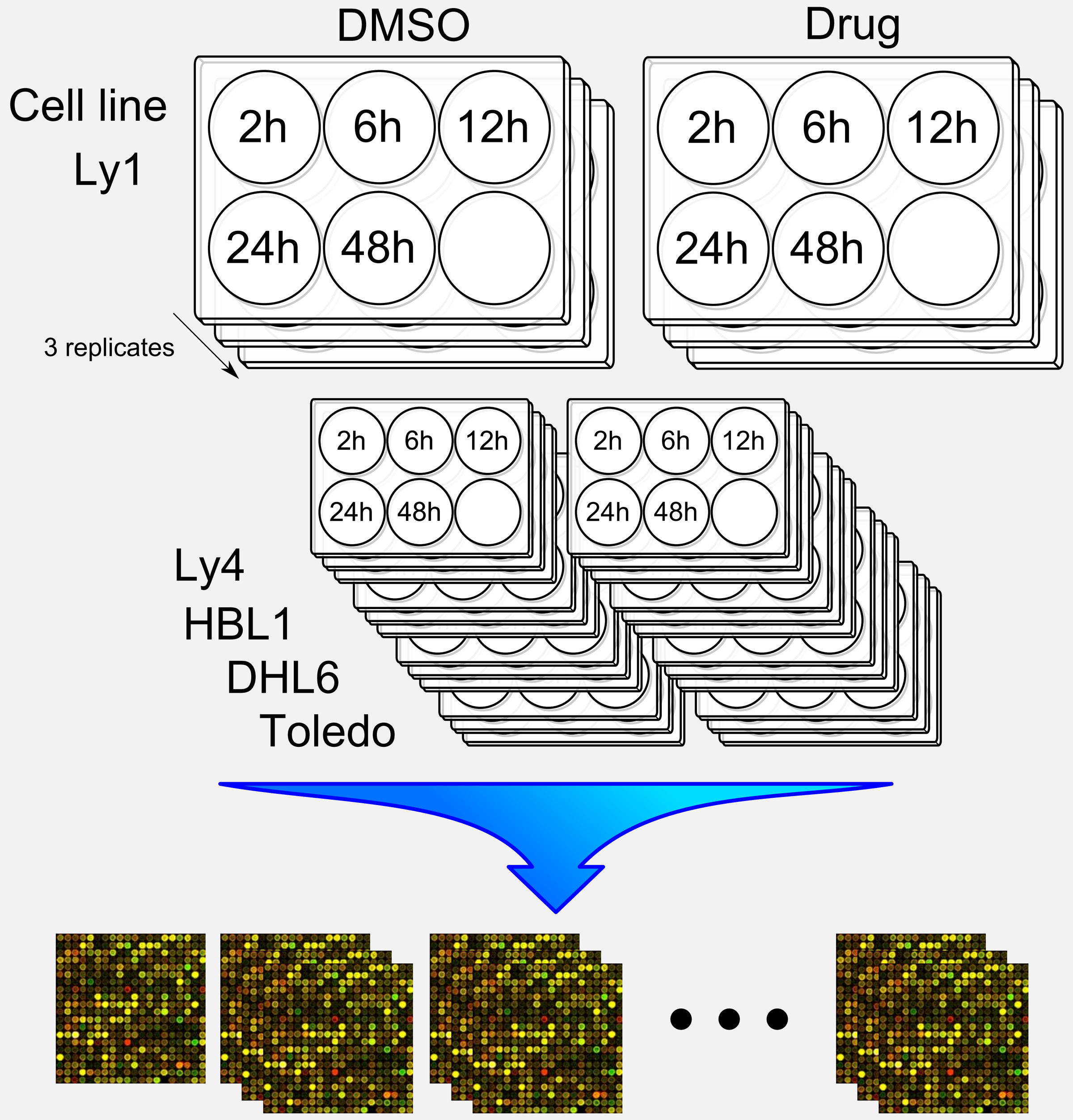
Adam Labadorf*[1], Heather Selby*[1], Arjan van der Velde*[1], Stefano Monti[1,2]

1 Bioinformatics Program, Boston University, 2 Center for Computational Biomedicine, Boston University Medical Center, *these authors contributed equally to this work

Diffuse large B-cell lymphoma (DLBCL) is the most common non-Hodgkin lymphoma in the United States. Forty percent of patients with DLBCL succumb to the disease, and new therapeutic approaches are needed. One such therapeutic is currently in clinical trials; however, the detailed biological mechanisms governing the response to this treatment in DLBCL are not well understood. Characterization of the transcriptional response to treatment is essential to understand the biological mechanisms of action of a drug. The focus of our project is the analysis of a large gene expression dataset consisting of a panel of DLBCL cell lines profiled at five time points after treatment. We developed a novel time series analysis approach to quantify the dynamic evolution of gene expression, and applied it to our dataset to carefully characterize the response to the pharmacological perturbation. The time series analysis identifies differential expression of genes, and enrichment of biologically relevant gene sets and pathways from publicly available repositories. We created a custom visualization tool to explore the various dimensions of our results at multiple levels of detail that is biologically intuitive. The combination of the time series analysis pipeline and the visualization tool identified both novel and previously known mechanisms of actions of the therapeutic treatment on DLBCL cell lines.

Timecourse Experimental Setup

- cells treated and harvested after indicated time interval
- mRNA is extracted
- samples analyzed on Affymetrix Human Gene 1.0 ST microarrays

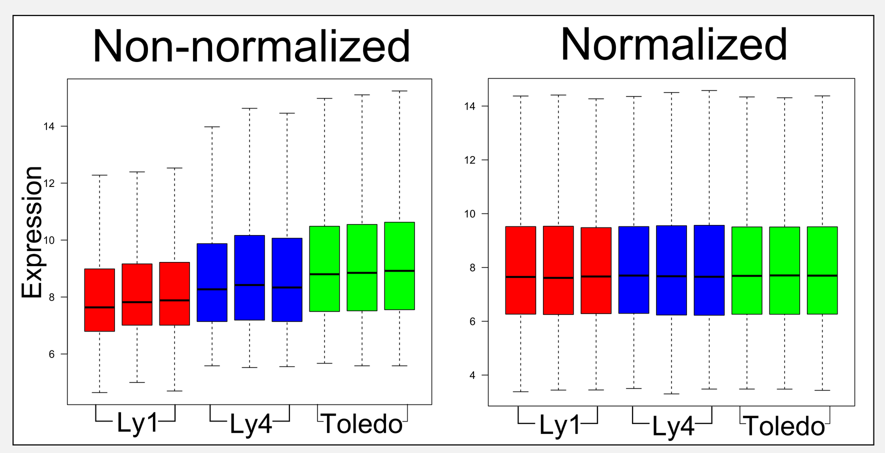


5 time points x 5 cell lines x 3 replicates x 2 conditions

150 microarrays

RMA Normalization

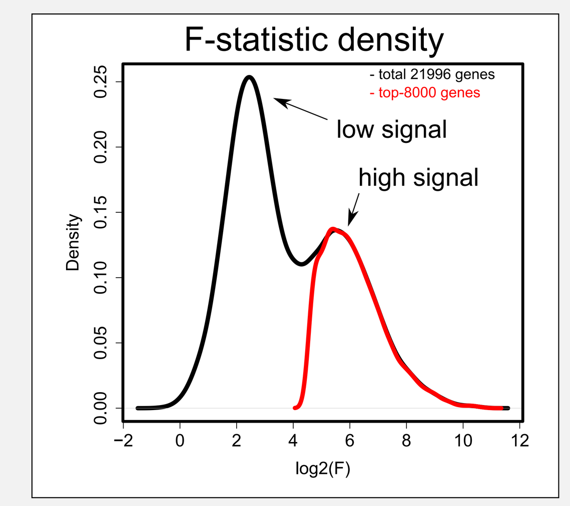
- Robust Multi-Array (RMA) Normalization
- quantile-normalization procedure for collections of microarrays
- ensures microarrays are statistically comparable
- reduces effects of biological and technical noise



Gene selection based on variance filtering

- genes with low variance across conditions cannot be differentially expressed
- many genes result in severe multiple hypothesis testing penalties
- select only 8000 genes with highest F statistic values

$$F_g = \frac{\text{variability of gene } g \text{ between conditions}}{\text{variability of gene } g \text{ between replicates}}$$

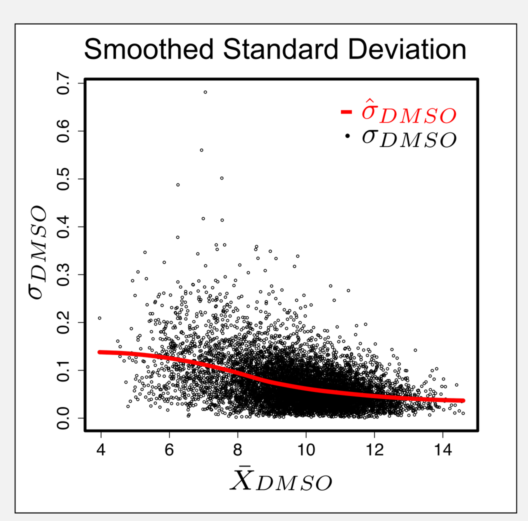


Conversion to z-score

- convert treated gene expression values to z-scores with respect to control:

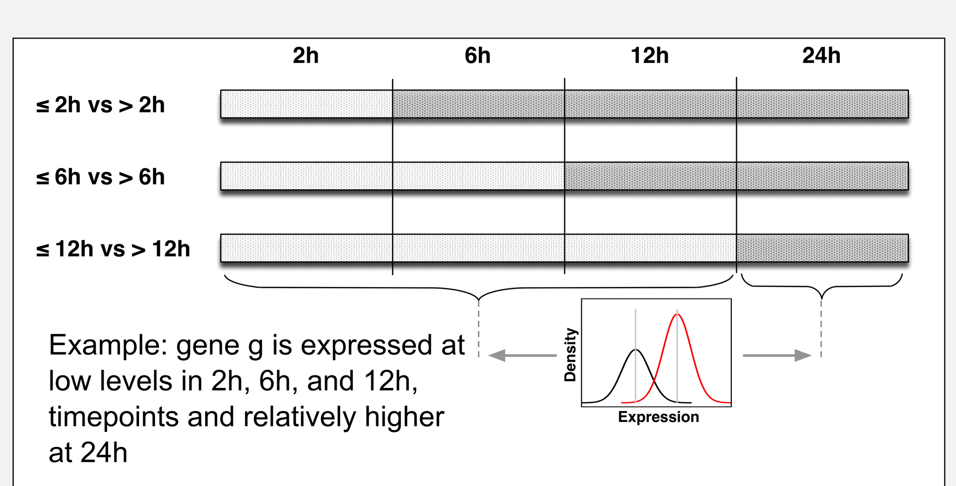
$$Z = \frac{X_{drug} - \bar{X}_{DMSO}}{\hat{\sigma}_{DMSO}}$$

- smooth $\hat{\sigma}$ using LOESS regression to avoid numerical issues due to noise



Sliding Window Time Series Analysis

- conduct differential analysis on z-scores
- for each gene, expression before and after each time point are compared
- significance is tested using an emprical distribution



Time Series Analysis

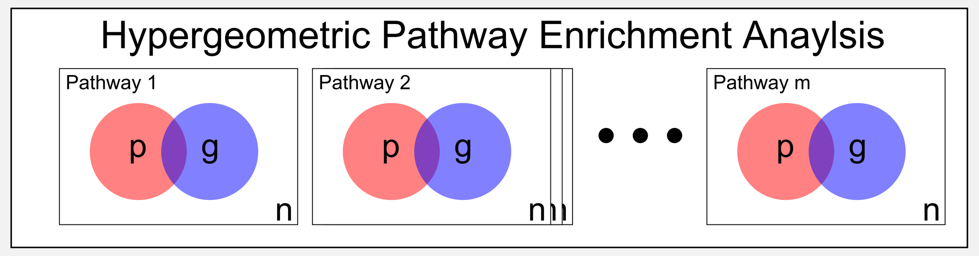
- consider each timepoint separately
- permutation-based significance testing

Linear model adjustment for cell line effects

Identify differentially expressed genes from T statistics

$$T = \frac{\text{control expression - treatment expression}}{\text{pooled standard error}}$$

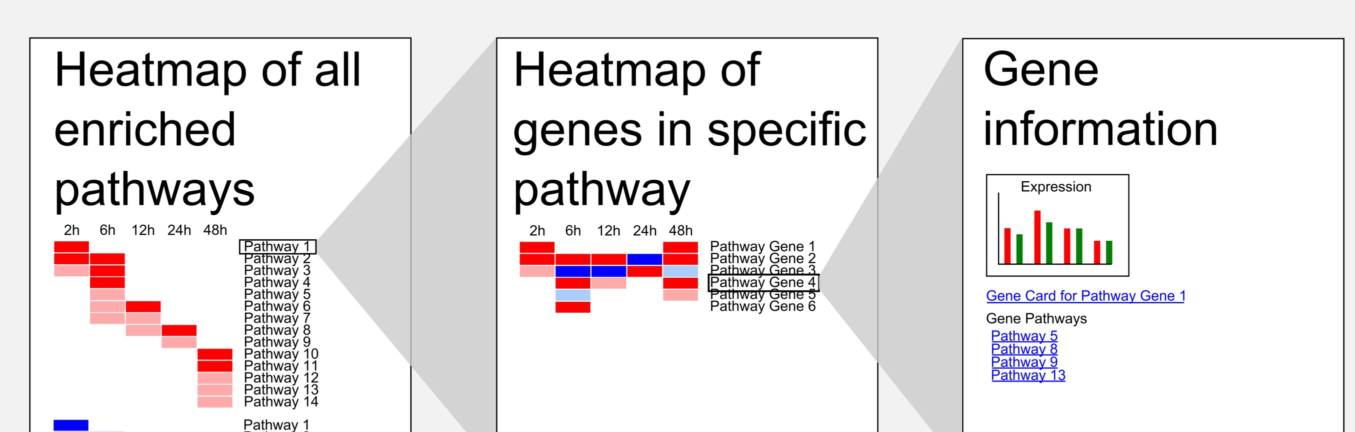
Hypergeometric Pathway Enrichment Anaylsis



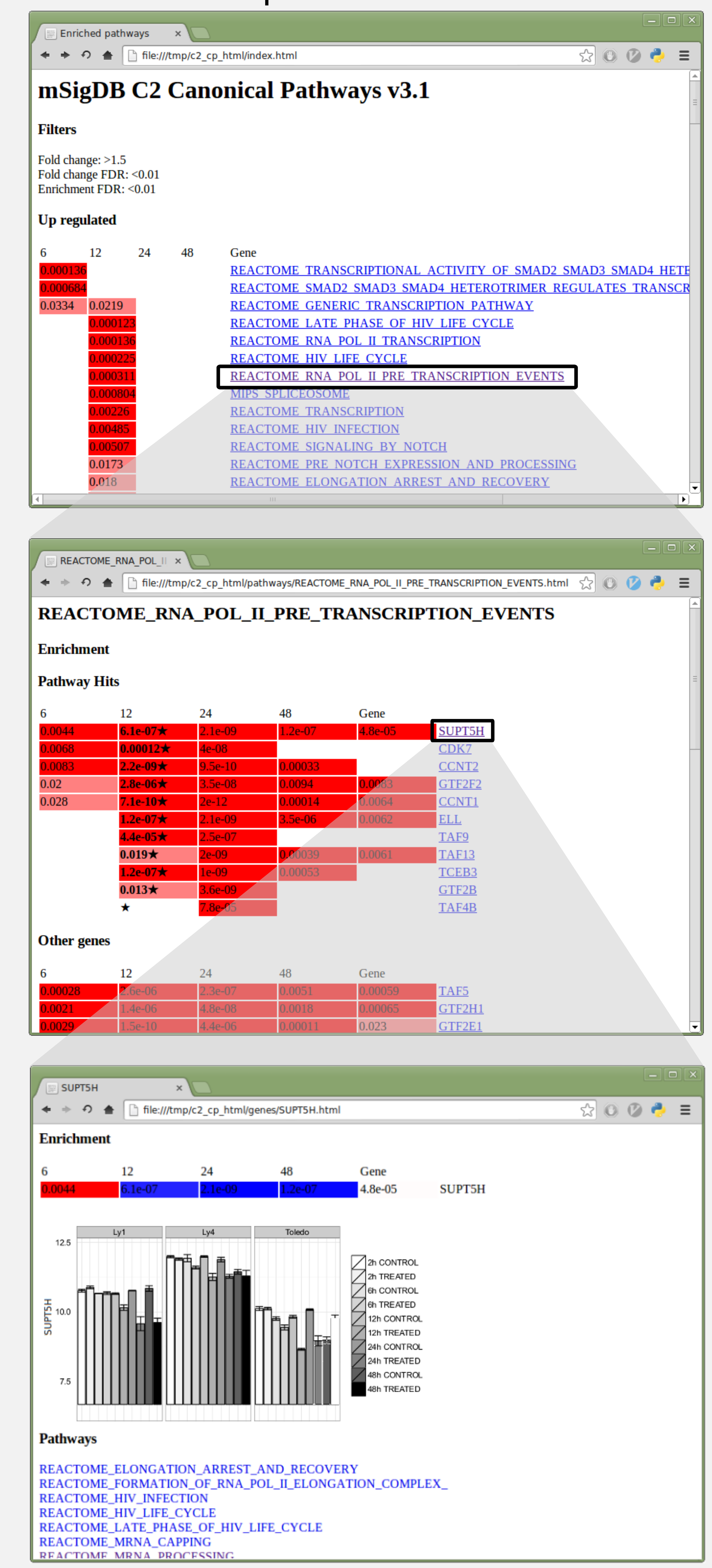
Multiple hypothesis corrected p-values

Visualization

- explore analysis at multiple levels of detail
- links to web resources e.g. Gene Card
- canned analysis produces manuscript-ready plots and heatmaps
- interactive hierarchical structure



Visualization Output



Acknowledgments

We would like to thank Dr. Stefano Monti and Dr. Gary Benson, and Dan Gusenleitner for their continued support and encouragement.

References

- Irizarry, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England), 4(2):249– 64, April 2003. 4
- Monti, et al. Integrative Analysis Reveals an Outcome-Associated and Targetable Pattern of p53 and Cell Cycle Deregulation in Diffuse Large B Cell Lymphoma. Cancer Cell, 22(3):359–72, September 2012. 2
- Subramanian, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102 (43):15545–50, October 2005. 6, 10