Towards the integration, annotation and association of historical microarray experiments with RNA-seq

Shweta S. Chavan, Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences
Michael A. Bauer, Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences
Erich A. Peterson, Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences
Donald J. Johann, Jr., Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences

Introduction

Transcriptome analysis by microarrays has produced important advances in biomedicine. For instance in multiple myeloma (MM), microarray approaches led to the development of an effective disease subtyping via cluster assignment, and a 70 gene risk score. Both enabled an improved molecular understanding of MM, and have provided prognostic information for the purposes of clinical management. Many researchers are now transitioning to Next Generation Sequencing (NGS) approaches and RNA-seq in particular, due to its discovery-based nature, improved sensitivity, and dynamic range. Additionally, RNA-seq allows for the analysis of gene isoforms, splice variants, novel gene fusions, etc. There is now a need to associate and integrate microarray and NGS data via advanced bioinformatic approaches.

Methods

Custom software was developed following a model-view-controller (MVC) approach to integrate Affymetix probe set-IDs, and gene annotation information from a variety of sources. The tool/approach employees a variety of strategies to integrate, annotate, and associate microarray and RNA-seq datasets.

Results

Output from the Cufflinks tool (from the Tuxedo suite) can be directly integrated, and/or associated with Affymetrix probe set data, as well as necessary gene identifiers and/or symbols from a variety of sources. Strategies are employed to maximize the integration/annotation process. Novel gene sets (eg, MM 70 risk score) can be specified, and the tool can be directly interfaced to the RNA-seq pipeline.

Conclusion

A novel bioinformatic approach to aid in the facilitation of both annotation and association of historic microarray data in conjunction with richer RNA-seq data is now assisting with the study of MM cancer biology.