RNA-Seq For Identifying Gene Expression Changes Associated with Relapse in Acute Lymphoblastic Leukemia (ALL)

Dustin Holloway, Steve McLaughlin, Christopher Clouser, Elizabeth Levandowsky, Tamara Gilbert, Tristen Ross, Jessica Spangler, Letha Phillips, Sue Heatley, Lei Wei, Jinghui Zhang, Clarence Lee, Heath Peckham, and Steven Mullighan

1 Presenting author 2 Life Technologies, Beverly, MA, USA 3 St. Jude Children’s Research Hospital, Memphis, TN, USA 4 Principle investigator

ABSTRACT

In this study we have performed RNA-Seq using the SOLiD™ system on samples from five children with acute lymphoblastic leukemia (ALL) with paired samples taken at the time of initial presentation with cancer (“initial” or “I”) and at the time of relapse (“relapse” or “R”). Whole transcriptome analysis was performed using several published statistical methods. These methods identified gene sets that can select cancer relapse samples with almost 80% accuracy, and have further identified genes that were found to be correlated with tumor relapse and prognosis in other studies. These results point the way to important follow up studies that may further elucidate the causes of cancer relapse.

INTRODUCTION

Although the rate of cure for ALL has been improving, a substantial minority of those afflicted with the disease still have poor outcome. Up to one fourth of children with ALL fail therapy and relapse. The biologic determinants of disease relapse are poorly understood. Previous studies have identified structural genetic alterations acquired at the time of relapse, and differences in gene expression patterns between matched samples obtained at the initial cancer and relapse conditions. However, a genome-wide analysis of sequence variation in relapsed ALL has not been performed. Moreover, next-generation sequencing approaches offer the opportunity to profile changes in gene expression patterns in great detail. Several methods including BaySeq, Weighted-voting, and hierarchical clustering were employed to select gene panels that help us to better understand cancer and leukemia.

MATERIALS AND METHODS

RNA samples were extracted using the TruSeq™ kit from bone marrow samples obtained at diagnosis and relapse from five children with ALL for the SOLiD™ system. Each sample was prepared with the SOLiD™ Total RNA-Seq kit (v. 1.0) for sequencing. Each of the 10 samples was barcoded and sequenced in 4 slides on the SOLiD™ system. Each slide contained 10 samples, with 5 samples sequenced per individual on a single 4-slide run. All samples were sequenced on the SOLiD™ system using the Bioscope™ analysis suite. Samples were analyzed using several statistical methods including bayesian component analysis (BPCA), significance analysis of microarrays (SAM), weighted-voting (WV), BaySeq, and Support Vector Machines (SVM). These methods produce sample-specific gene sets that identify relapse samples with 80% accuracy, and have further identified genes that were found to be correlated with tumor invasiveness and poor prognosis in other studies. Further details of the SNP results can be found in supplementary sequence alignments unique to the relapse state in some individuals, providing a post of variations which may include potential drivers of metastasis.

RESULTS

Figure 1. Average Exon Coverage per Sample. Aggregate aggregate gene expression and BaySeq primary expression coverage per child.

Figure 2. PCA on samples with standard-gene-wise normalization. Principle Component Analysis on gene expression values from all samples across samples. Each point in the graph corresponds to one sample. The blue points correspond to the initial state, and the red points correspond to the relapse state. The first two components capture much of the variability in the data.

Figure 3. Relapse classifier genes can be selected with a variety of methods. Both BaySeq and Weighted-voting selected a set of genes that is highly predictive of relapse. The selected genes were also consistent with previous findings.

Figure 4. BaySeq uses row read counts to separate relapse and initial samples. BaySeq shows a good job at identifying differentially expressed genes.

Figure 5. 142-classifier genes identified by SAM are differentially expressed in the initial and relapse states. Expression is shown as normalized RPKM values. SAM was conducted using the two-sample paired statistic. The 40 samples (including replicates) are listed at the bottom of each expression graph.

Figure 6. 20-classifier genes identified by at least 2 methods cluster according to initial and relapse states.

CONCLUSIONS

The results from this study provide a powerful tool for exploring gene expression and for discovering potential new candidates in whole-transcriptome analyzes such as BaySeq are able to identify novel candidate genes which may yield excellent sample separation. Combined analysis with multiple techniques is successful at selecting gene panels that help us to better understand cancer and leukemia.

REFERENCES

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

© 2011 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners.