ApplicationForum

Convenient and Robust Genome-Wide Targeted Expression Analysis for Biomarker Discovery

Introduction

Drugs with complementary companion diagnostics that identify treatable patient populations require smaller clinical trial cohorts and have a higher likelihood of successful approval. Particularly in oncology, identification of suitable biomarkers for new potential drugs has almost become a requirement. As a result, screening and discovery groups need effective and efficient tools to readily find and validate potential biomarkers.

While proteins often provide good biomarkers, broad-based proteomic analyses remain very complex, expensive and/or inconsistent (e.g., mass spectroscopy, antibody arrays). RNA transcripts, however, provide a class of biomolecules whose expression levels can be reliably assessed in a genome-wide manner, and gene expression profiles can be used both directly as a biomarker and to indicate which proteins might be useful for biomarker assessment.

Current RNA-Seq or microarray technologies are not ideal methodologies to carry out whole transcriptome analysis for biomarker discovery. Poor design flexibility, low sensitivity, and protocol complexity restrict their utility. To address these limitations, we developed the DriverMap™ Expression Profiling Assay—a targeted RNA profiling method that greatly simplifies sample processing while providing more a quantitative and comprehensive analysis of expressed protein-coding genes than RNA-seq.

Validated One-Tube Multiplex Reaction to Profile All Human Genes

Development of the DriverMap Assay required an intensive program to identify a set of primers that function together in a single multiplex PCR reaction to amplify representative segments of each of the 19,000 protein-coding human gene transcripts. Further, the targeted transcript amplicons needed to be the correct and size and sequence composition to run on an Illumina® sequencing instrument.

Multiplex PCR presents many challenges, including high levels of non-specific amplification, mis-priming of homologous and highly abundant templates (rRNA, mitochondrial, globin RNAs), and complex primer-primer interactions. In addition, variabilities in primer performance, sample characteristics, and secondary structures complicate assay robustness and standardization. As a result, most developed multiplex PCR assays are limited to, at best, several hundred target mRNAs.

However, we were able to overcome the challenges of developing a single complex multiplex PCR reaction for 19,000 human protein coding genes by developing a pipeline that combined bioinformatics with high-throughput multiplex experimental validation of thousands of PCR-primer combinations. The combination of thorough sequence curation and reaction scale-up enabled us to develop a simple, one-tube reaction for highly reproducible (R^2=0.997) genome-wide, multiplex PCR assay of all 19,000 protein-coding genes.

Advantages of DriverMap Targeted Expression Profiling Assay over Standard RNA-Seq

An obvious advantage of the DriverMap assay's targeted amplification and sequencing approach is that there is no problem starting the assay with total RNA. This saves a lot of work by eliminating the need
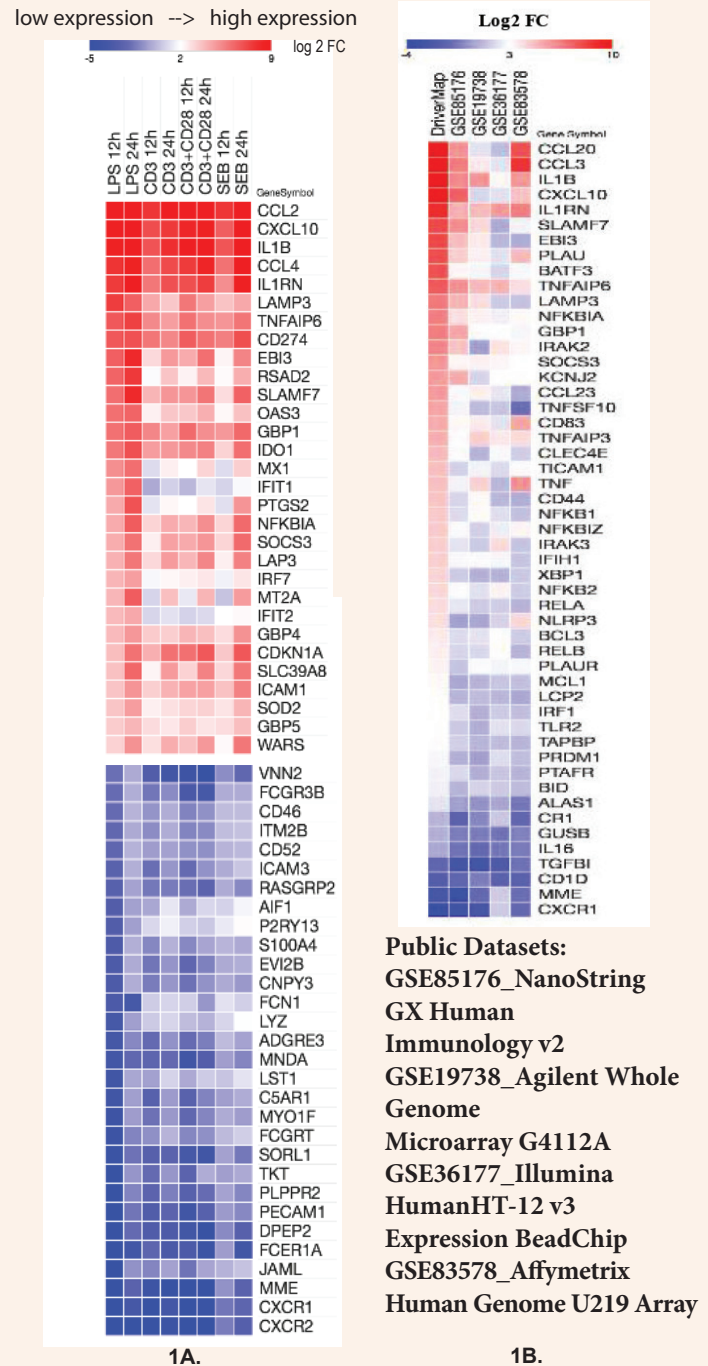
1A.

Figure 1A: Whole blood activation data. Whole blood in TruCulture(R) test tubes was activated by LPS, CD3, CD3-CD28 or Enterotoxin SEB for 12 hours or 24 hours and changes in RNA (PAXgene(R)) expression level were analyze using the DriverMap 8K assay.

1B.

Public Datasets:
GSE85176_NanoString GX Human Immunology v2
GSE19738_Agilent Whole Genome Microarray G4112A
GSE36177_Illumina HumanHT-12 v3 Expression BeadChip
GSE83578_Affymetrix Human Genome U219 Array

Figure 1B: Sensitivity Comparison. DriverMap shows significantly greater sensitivity in detection of differentially expressed genes than other data sets generated utilizing comparable technologies.

139

to remove ribosomal RNA, deplete globin sequences, or otherwise enrich for mRNA. Direct measurements made in whole blood (WB) have the advantages of minimizing sample handling and provide robust in vitro assay systems that preserve cellular interactions. To perform ex vivo stimulation while maintaining physiological cellular interactions, we utilized TruCulture tubes to activate immune cells present in WB, single- step RNA extraction and targeted RNA-Seq readout systems to define the transcriptional response to clinically relevant cytokines (Figure 1A). Data generated from whole blood samples shows significantly greater sensitivity in detection of differentially expressed genes when comparing data generated on other technology platforms with the DriverMap method (Figure 1B). In this way, researchers can go right from total RNA isolation to a one-tube protocol that starts with cDNA synthesis and ends with libraries ready to load onto an Illumina flowcell for sequencing.

In addition to the simplified one-tube protocol, the DriverMap assay's amplified cDNA products have very low complexity since there is just one
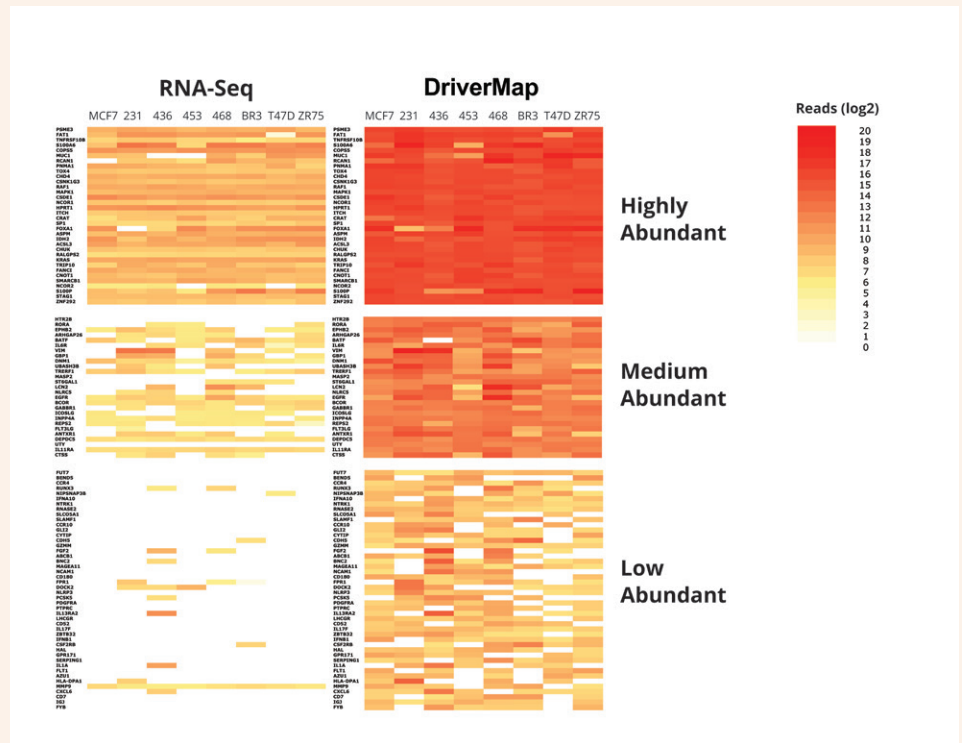

Figure 2: Comparison of sensitivity of Targeted RNA-Seq vs DriverMap assay: NGS read levels detected RNA-Seq and Driver-Map for selected high-abundant (10K-100K copies per sample), medium-abundant (1,000-10,000 copies per sample), and low-abundant transcripts (100-1,000 copies per sample) in 50ng of total RNA from seven common cancer cell lines.

defined transcript for each gene target as compared with the highly complex mix of cDNA fragments produced by standard RNA-Seq approaches. Sequencing just 19,000 defined cDNA amplicons, rather than the whole fragmented transcriptome, enables more sensitive detection of the targeted transcripts with much fewer sequencing reads. Only 5 million sequencing reads are required to accurately measure expression levels of 10-30% more transcripts than RNA-Seq (Figure 2). Further, overall counts for most transcripts, especially those on the low end, are higher and more reliable than with RNA-Seq.

Also, the sensitivity difference between DriverMap assay and standard RNA-seq improves immensely when input RNA sample sizes are reduced. DriverMap assay can reliably detect expression level differences with as little as 10 pg of input RNA.

Finally, DriverMap's targeted transcript approach also greatly simplifies analysis of the sequencing run since each sequence read matches a predefined amplicon for each transcript. Analysis, then, just requires matching and counting each sequencing read to its target in a 1-to-1 manner. Simple normalization of the individual read counts relative to the total number of reads for each sample run allows direct comparisons of expression level changes across samples with procedural and biological variations between samples addressed using standard housekeeping gene normalization.

Workflow and more available at www.cellecta.com/drivermap