

Making cancer transcriptome sequencing assays practical for the research and clinical scientist

Todd M Smith^{1*}, N Eric Olson¹, David Smith²

From Beyond the Genome: The true gene count, human evolution and disease genomics
Boston, MA, USA. 11-13 October 2010

Next generation DNA sequencing (NGS) technologies are increasing in their appeal for studying cancer genomics. High-throughput data and a growing repertoire of applications that quantitatively measure gene expression, splicing, noncoding RNAs, and genomic variation are revealing that cancer is a more complex and heterogeneous disease than previously imagined. Fully characterizing the ~10,000 types and subtypes of cancer that exist to develop biomarkers that can be used to clinically define tumors and target specific treatments requires large studies that examine specific tumors in thousands of patients. This goal will fail without significantly reducing both data production and analysis costs, so that most cancer biologists and clinicians can conduct NGS assays and analyze their data in routine ways.

Currently, most cancer biology NGS papers are published either by genome centers or through collaborations with instrument vendors. However, this is going to change rapidly with efforts like the Cancer Genome Anatomy project. In any case, large teams of bioinformaticians are involved in analyzing data through labor-intensive processes. With refinements offered by the Illumina HiSeq 2000, or Life Technologies SOLiD 4, the cost of collecting data for transcriptome analysis and mate-pair genome sequencing is sufficiently inexpensive for small groups and individuals, beyond genome centers, to conduct the required studies. However, current data analysis methods need to be automated with established tools in scalable and adaptable systems that provide standard reports to make results available to enable interactive exploration by biologists and clinicians.

In our presentation, we will examine the time and costs required to analyze data that will be collected in future cancer studies. Using data from existing matched tumor and normal transcriptome studies from random

oral cancer samples, and samples grouped by drinking and smoking behavior (as a tool to define data analysis requirements), we will compare the costs of conducting large studies using current data analysis approaches with those using integrated software systems to demonstrate how automation reduces costs, while providing comparable results for identifying transcript isoforms, mutations and novel translocations. Geospiza's GeneSifter distributed cloud-based software architecture, including open source tools, like BioHDF, will be described to share insights into high performance computing requirements for scalable data processing.

Author details

¹Geospiza Inc. Seattle WA 98119, USA. ²Mayo Clinic, Rochester MN 55905, USA.

Published: 11 October 2010

doi:10.1186/gb-2010-11-S1-P39

Cite this article as: Smith *et al.*: Making cancer transcriptome sequencing assays practical for the research and clinical scientist. *Genome Biology* 2010 **11**(Suppl 1):P39.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



¹Geospiza Inc. Seattle WA 98119, USA

Full list of author information is available at the end of the article