RNASeq 2.0 @ Broad Genomics

Carrie Cibulskis*, Megan Shand*, Maura Costello*, Caroline Peterson, Kylee Bergin, Alyssa Macbeth, Gina Vicente, Yossi Farjoun, Kathleen Tibbetts, Niall Lennon

Abstract
RNA sequencing has the potential for broad reaching applications in clinical research and clinical practice. Approaches to the generation and analysis of RNA sequencing data has numerous potential endpoints including, but not limited to, the detection of variants, gene fusions, alternative splicing, and expression profiling.

The ability to generate high quality RNA sequencing data and perform sensitive and efficient analyses is critical to realizing the full value of transcriptome data in clinical research and practice. Broad Genomics and our collaborating research teams have a long history of generating and analyzing RNA sequencing data for both cancer and germline applications. Analytic tools currently available for RNA sequencing analysis are many and varied in terms of application, practicality and efficiency. To better enable our research teams to gain optimal utility from RNA based studies, we completely reassessed the data and analyses that we are delivering.

We are in progress of launching an updated RNASeq 2.0 pipeline that includes optimized read alignment, fusion, expression, and variant detection. Broad RNASeq 2.0 analysis pipelines will be made available in our cloud-based analysis workbench (FireCloud) and will facilitate scalable and rapid analysis runs, RNA/DNA analysis integration and ease of data sharing.

Goal of RNASeq 2.0 @ Broad Genomics
Our Goal is to make RNA sequencing and analysis available for translational research studies. To support this aim requires a pipeline with the following characteristics:

1. Flexible input material types (Blood, Tissue, FFPE Tissue, Extracted RNA material)
2. Rapid, scalable and auditable workflows
3. Ability to process and analyze RNA & DNA material in parallel
4. High quality analysis pipelines
5. Easy & Efficient data sharing capabilities

Current RNA Sequencing Products

| Product Name | Standard RNA Seq | Strand-Specific RNA Seq | Transcriptional Capture
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FFPE TP_1</td>
<td>FFPE TP_2</td>
<td>FFPE TP_3</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_4</td>
<td>FFPE TP_5</td>
<td>FFPE TP_6</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_7</td>
<td>FFPE TP_8</td>
<td>FFPE TP_9</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_10</td>
<td>FFPE TP_11</td>
<td>FFPE TP_12</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_13</td>
<td>FFPE TP_14</td>
<td>FFPE TP_15</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_16</td>
<td>FFPE TP_17</td>
<td>FFPE TP_18</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_19</td>
<td>FFPE TP_20</td>
<td>FFPE TP_21</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_22</td>
<td>FFPE TP_23</td>
<td>FFPE TP_24</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_25</td>
<td>FFPE TP_26</td>
<td>FFPE TP_27</td>
<td></td>
</tr>
<tr>
<td>FFPE TP_28</td>
<td>FFPE TP_29</td>
<td>FFPE TP_30</td>
<td></td>
</tr>
</tbody>
</table>

Adding Flexibility

Enabling FFPE: Transcriptome Capture Quality

Cancer applications often require high quality transcriptome analysis from FFPE tissue specimens. Our transcriptome capture process has been recently optimized to customize sample input amounts and results in lower % duplication and increased unique coverage.

Enabling RNASeq from Blood: Globin Depletion

In RNA extracted from whole blood, globin transcripts make up over 70% of the total mRNA and are of little interest to most researchers. We have established a production process using the Thermo-GLOBINclear kit. This sample prep will be available for our standard and strand-specific RNASeq products.

Improving Production Alignment Pipeline

STAR Alignment

The RNA sequencing deliverable from all Broad genomics products is an aligned BAM file ready for downstream analysis. We have updated our pipeline to use STAR (Spliced Transcripts Alignment to a Reference) instead of the previous workflow algorithm - TopHat.

Quality Control & Integrated Platforms

Adoption of RNA sequencing by translational researchers as companion data source to clinical-grade whole exome, requires upgrade of product development activities and sample tracking systems to ISO quality level.

Broad Genomics aims to upgrade all RNA sequencing processes to support bundled clinically relevant workflows for whole exome, RNA and targeted platforms.

Delivering Data to the Cloud, Downstream Analysis and Data Sharing

Enabling Collaboration: Genome Analysis Workbench

BAM Files will be delivered to Cloud Based Genome Analysis Workbench

Best Practices Downstream Analysis Methods available in Methods Repository

Pipelines Prioritized for Release in Methods Repository

- Gene Expression
- Variant Calling
- Variant Validation DNA/RNA
- Fusion Detection
- STAR Alignment

Technical variation in gene expression measurements across sample prep, sequencing and analysis runs can confound the ability to discover real biological signal. Large collaborative projects are extra sensitive to such potential variation and as such require methods to normalize measurements across samples & sequencing centers. To better support these projects we are validating automated methods to incorporate ERCC spike in controls for accurate measurement of experiment to experiment variability and improve downstream gene expression analysis.

Processes to support 96 spike ins will be implemented in both standard and capture RNA sequencing products.