ST04. Development of Clinical Transcriptome (RNA-Seq) Analysis for Detection of Fusions and Gene Expression in Oncology Samples

M.B. Durso, *S. Zhong*, *A.I. Wald*, *L.M. Kelly*, *K.M. Callenberg*, *Y.E. Nikiforov*, *M. Nikiforova*
University of Pittsburgh Medical Center, Pittsburgh, PA.

Introduction: Transcriptome sequencing (RNA-Seq) serves as a discovery tool for detection of novel fusion types and quantifying gene expression levels in research laboratories. It is not routinely applied in the clinical setting due to difficulties in both wet sequencing and bioinformatics analysis. We have developed and validated an RNA-Seq approach that can be effectively used in the clinical laboratory for detection of gene fusions and gene expression in oncology samples. **Methods:** RNA-Seq analysis was performed with sequencing parameters optimized for use in the clinical laboratory and validated on 200 ng of RNA from 13 frozen tumors with known fusion types. Transcriptome libraries were generated with the TruSeq Stranded Total RNA kit (Illumina) and sequenced using rapid run on the HiSeq 2500 using on-board clustering and the HiSeq Rapid SBS Kit v2 (Illumina). A custom bioinformatics pipeline was developed that uses both existing bioinformatics tools including Tophat and Chimerascan, and also in-house developed scripts and filters for quantifying gene expression levels as well as improving the specificity of fusion detection. The pipeline was validated on 26 tumors with previously known gene fusions. **Results:** RNA-Seq analysis was optimized to sequence 5 tumor samples in a 27 hour run. An average of 192x106 total reads per sample were generated with 97% of >Q30 quality score. Between 2 and 50 reads spanning break points was detected per driver fusion. The developed bioinformatics pipeline accurately detected different types of fusions including fusions in ALK, BRAF, PPARG, RET, NTRK1, and NTRK3 genes in all samples analyzed. Custom filters narrowed down the fusion candidate list to 1-3 clinically important fusions per tumor sample that were further confirmed by evaluation of break points and expression pattern of fused genes. The cost of sequencing reagents was $832/sample and library preparation, sequencing, and analysis running time was 59 hours per sample. **Conclusions:** Whole transcriptome (RNA-Seq) analysis can be used in clinical setting for accurate detection of gene fusions and gene expression in oncology samples. Its clinical utility may be the highest for those tumors where no targetable mutations had been detected by less expensive and more rapid targeted NGS analysis.

ST24. Myxoinflammatory Fibroblastic Sarcoma: Association of t(1;10) Translocations in a Cohort of 49 Cases by Fluorescence *in Situ* Hybridization

S. Puetz, *K. Schilter*, *N. Harding-Jackson*, *S. Suster*, *A.C. Mackinnon*
Medical College of Wisconsin, Milwaukee, WI.

Introduction: Myxoinflammatory fibroblastic sarcoma (MIFS) is a rare, low-grade, soft tissue malignancy predominately occurring in young or middle-aged adults. The tumors are mainly limited to the distal extremities with a high rate of local recurrence and low metastatic potential. The lesions tend to be small and often present as slow-growing, poorly-defined, painless masses. Varying proportions of inflammatory, myxoid, and large atypical spindle and epithelioid cells make the differential diagnosis of MIFS challenging. Molecular cytogenetic assays, such as fluorescence *in situ* hybridization (FISH) are effective diagnostic aids by identifying translocations specific to soft tissue tumors. Previous studies described t(1;10) translocations in a subset of MIFS resulting in the rearrangement of two genes, *TGFBR3* and *MGEA5*. To further investigate the utility of detecting *TGFBR3* and *MGEA5* rearrangements as a potential diagnostic marker, we performed FISH on a large cohort of MIFS. **Methods**: The cohort consisted of 4 control cases demonstrating the t(1;10) translocation by conventional karyotype analysis and 49 tumors diagnosed as MIFS by histopathological review. FISH analysis was performed using a custom-designed, locus-specific probe for *MGEA5* according to the manufacturer's protocol (Empire Genomics). For each case, a total of 40 nuclei were scored by fluorescence microscopy. A case was confirmed positive for the rearrangement if ≥25% of the nuclei showed loss of the 5'-MGEA5 probe. **Results**: All 4 control cases demonstrated the t(1;10) translocation by FISH. 37 of 49 tumors could be confidently scored by FISH analysis, and none demonstrated the t(1;10) translocation. **Conclusions**: The t(1;10) translocation, previously described as a common cytogenetic alternation in MIFS, was not identified in our cohort of MIFS cases. Additional studies interrogating larger cohorts and other described aberrations, such as amplification of chromosome 3, are necessary to cytogenetically characterize this complex sarcoma.

I32. MSK-LYMPHOCLONE: Data Analysis Pipeline and Tools for Immune Repertoire Analysis

M.H. Syed[1], N. Khedoudja[1], T. Baldi[1], A. Zehir[1], D. Cheng[2], M. Ladanyi[1], M.E. Arcila[1]
[1]Memorial Sloan Kettering Cancer Center, New York, NY; [2]Illumina, Santa Clara, CA.

Introduction: Clonality assessment is an integral part of the diagnosis and follow-up of B and T cell malignancies. The emergence of NGS methods for clonality testing has opened the door to a new era in diagnosis and understanding of clonality. NGS assays allow for the full characterization of clonal populations with higher sensitivity and specificity than traditional methods but the optimal bioinformatics pipelines remain to be established. Here we describe our pipeline and tools for immune repertoire analysis for use in routine clinical testing. **Methods:** Clonality testing was performed using LymphoTrack assays (InVivoScribe) with adapters for Illumina MiSeq. We have developed a bioinformatics pipeline and data analysis tools to identify the diversity and clonality in samples of patients diagnosed with B or T cell malignancies. The pipeline starts with demultiplexing, adaptor trimming and creating FastQ files from the output of MiSeq (Illumina) sequencer. Paired-end reads are combined and fasta files are created. Identical sequences are removed and only a single sequence is kept in the final fasta file, along with the count in the header. Each unique sequence is blasted using igBlast to search local IGH and TCR databases. IGH and TCR reference data are downloaded from IMGT website. igBlast output is grouped based on different VJ combinations. The generated output summary is loaded into a MySQL database, along with hit details for each sequence. Visualization tools were created for easy pathologist review. **Results:** A total of 249 samples were analyzed. Clonality calls were made and compared to results generated by the proprietary LymphoTrack Software provided by Invivoscribe. The custom pipeline (MSK-LYMPHOCLONE) was able to analyze and assign same clonality calls in all samples. Visualization tools created for easy viewing and mining of results were made available through web interfaces for analysis during clinical signouts. A run page lists all samples in each clinical run with tabbed links to tools for review and data analysis including: 1) Interactive, zoomable, stacked histogram showing different VJ combinations in different colors; 2) Tabular clone summary interface; 3) A detailed igBlast output for each VJ combination sequence. Multiple selected sequences can be downloaded, or aligned through this interface; 4) A BLAST interface to do sequence search against sample sequences. This tool is used to track a clone sequences in minimum residual disease studies. **Conclusions:** We have developed and validated MSK-LYMPHOCLONE, a custom data analysis pipeline and suite of tools to mine and analyze next generation sequencing data for IGH and TCRG in lymphoid malignancies. This may be used independently or in conjunction with existing software solutions.