I44. Multi-Institutional FASTQ Proficiency Testing Reveals High Concordance in Reporting Clinically Significant Single-Nucleotide Variants but Discrepancies in Reporting Insertions/Deletions

M.S. Farooqi^{1,7}, K.D. Davies², M. Gruidl³, C.E. Hill⁴, J.W. Hirschhorn⁵, H. Jones⁴, K.L. Jones², A. Magliocco³, R. O'Rourke², N.M. Patel⁶, D. Qin³, E. Ramos⁷, M.R. Rossi⁴, T.M. Schneider⁴, G.H. Smith⁴, J.Y. Park¹, D.L. Aisner²

¹University of Texas Southwestern Medical Center, Dallas, TX; ²University of Colorado – Anschutz Medical Campus, Aurora, CO; ³H. Lee Moffitt Cancer Center, Tampa, FL; ⁴Emory University, School of Medicine, Atlanta, GA; ⁵Medical University of South Carolina, Charleston, SC; ⁶University of North Carolina School of Medicine, Chapel Hill, NC; 7Children's Medical Center, Dallas, TX. Introduction: Proficiency testing (PT) is a key component of laboratory quality management. Clinical next-generation sequencing (NGS) entails multiple steps which require PT: analytic, bioinformatics (BI), and interpretation. PT design can encompass individual or all phases of testing. Here, we employ an alternative PT method to assess BI by exchanging raw data files (FASTQ (FQ)) among multiple institutions all using Illumina's TruSight Tumor (TST) gene panel kit and MiSeg instrument. Methods: Six sites that had validated the TST assay participated in the study. Each site submitted FQ files from 4 previously tested specimens. The 24 FQ datasets (including 3 commercial calibrators) were de-identified and redistributed to 5 of the centers via secure ftp server (one site could not reanalyze datasets). Concordance in reporting of variants and variant allele frequencies (VAF) were determined for the 5 centers. Results: Various alignment and variant calling tools were used across sites including BWA-MEM, GATK, GSNAP, FreeBayes, Clinical Genomicist Workstation, and NextGENe. Excluding calibrator samples, 48 variant calls (16 clinically significant variants (csV) and 32 variants of unknown significance (VUS)) were reported across 21 samples. Overall, 68.8% of total variants (33 of 48) were concordant between all sites; this included csV (11 of 16) and VUS (22 of 32). There was 100% concordance in reporting single-nucleotide csV (cs-SNV). Separately, three FQ calibrators with 11 SNV at VAF ≥5% (range 5% to 33.5%) as determined by digital droplet PCR were also analyzed. All reported values from the calibrators had VAFs within 10% of the target value; 78% of calls were within 5%. No sites reported VAF <5% in the calibrator samples, consistent with each site's reportable range. In contrast to SNV reporting, the reporting of csV insertions/deletions (indels) only had 28.6% concordance (2 of 7); discordant indels included two EGFR exon (ex) 19 deletions (del), one EGFR ex20 insertion (ins), one KIT ex11 ins, and one ERBB2 ex20 ins. Three FQs datasets had an identical EGFR ex19 del; 2 were reported with 100% concordance; the third was only identified by 2 of 5 sites at VAFs of 14 and 18%. The same 2 concordant sites also reported the KIT ins not detected by the other 3 sites. Conclusions: This report highlights the need for BI PT for clinical laboratories. The concordance for reporting cs-SNV was acceptable for the first implementation of this PT program; however, there was low concordance in reporting csV indels. The source of these discrepancies requires further evaluation to determine how best to implement corrective action. This alternative PT program assessing BI identifies areas for laboratory quality improvement in NGS.

G18. Clarifying Diagnoses: Application of Exome Sequencing for Neuromuscular Disorders

G.T. Haskell¹, M.C. Adams¹, L. Zhou¹, Z. Fan¹, N. Chahin², B. Powell¹,

K.C. Wilhelmsen¹, K.E. Weck¹, J.P. Evans¹, J.S. Berg¹

¹University of North Carolina at Chapel Hill, Chapel Hill, NC; ²University of Connecticut / Hartford Hospital, Hartford, CT.

Introduction: Neuromuscular disorders (NMD) are genetically heterogeneous conditions that can be difficult to diagnose based solely on clinical presentation. Whole exome sequencing (WES) may be particularly helpful for arriving at a molecular diagnosis in NMD cases, as many genes can be evaluated simultaneously. However, the optimal workflow for diagnostic labs performing WES for NMD is not clear. **Methods:** We evaluated the WES diagnostic yield for 106 NMD patients using either a focused list of established diagnostic genes related to the patient's phenotype (myopathy only or neuropathy only) versus a broad list of genes associated with myopathic and neuropathic NMD phenotypes. Clinical data was also collected on each patient. **Results:** NMD patients referred for WES had often undergone extensive previous clinical testing, including single gene tests, panels, muscle biopsies, and nerve conduction studies. Most cases had negative family history. A small subset of cases were thought to have conditions not amenable to diagnosis by WES, indicating that opportunities exist for educating ordering physicians. WES identified a known or likely pathogenic variant in 13% of neuropathy patients

and 17% of myopathy patients, as well as many variants of uncertain significance. For well-phenotyped individuals, the focused list of genes related to the suspected clinical diagnosis yielded an answer, Critically, for NMD cases with an unclear clinical diagnosis, the broad list identified likely pathogenic variants in genes not originally suspected as possible explanations for the patient's phenotype. One individual considered to have symptoms consistent with a nonspecific myopathy or connective tissue disorder carried a frameshifting variant in COL9A3, consistent with autosomal dominant MED3 (MIM 120270). In another case thought to have a CMT-like neuropathy, the broad NMD list identified a previously reported pathogenic variant in COL6A1, consistent with Bethlem myopathy (MIM 158810). For a patient previously thought to have hereditary spastic paraplegia, WES identified a nonsense variant in GCH1, indicating dopa-responsive dystonia (MIM 128230). This patient demonstrates the dramatic utility that can stem from making a correct diagnosis, as she was started on dopa therapy and regained the ability to walk without assistance. Conclusions: Molecular labs can establish efficient WES workflows to clarify the diagnosis in a significant fraction of NMD cases. Detailed requisition forms will help ensure WES variants are filtered by the most appropriate gene lists. Although a permissive gene filter is needed in some cases to account for ambiguity in phenotyping or phenotypic overlap, evaluating lists in a staged process can provide efficiency in the analysis.

ST02. Diagnostic Yield and Utility of RNA Sequencing in the Detection of Pathogenic Fusions in Childhood Sarcomas of Uncertain Diagnoses

J. Reuther, V. Kumar, K.R. Covington, H. Doddapaneni, O.A. Hampton, D.M. Muzny, S. White, N.R. Patel, P. Sumazin, D.A. Wheeler, S.E. Plo, D.H. Lopez-Terrada, D. Parsons, A. Roy Baylor College of Medicine, Houston, TX.

Introduction: Fusion genes are well-recognized oncogenic drivers in cancers. We assessed the feasibility and clinical utility of whole transcriptome sequencing (RNA-seq) for detecting diagnostic and targetable fusion genes in childhood sarcomas of uncertain diagnosis. Methods: A RNA-seg pipeline for fusion gene detection was trained using a test set of 9 soft tissue sarcomas with clinically diagnosed fusion genes. A validation cohort of an additional 28 sarcomas or mesenchymal tumors of uncertain diagnosis were analyzed for gene fusions. Total RNA from fresh-frozen tumors was used to prepare poly-A+ stranded Illumina libraries, generating ~89 x 106 (2 x 100bp) reads/sample. Fusion genes were detected using deFuse (v.0.6.1). Predicted fusions were ranked on clinical utility using disease-specific clinical guidelines (WHO), COSMIC database, as well as published literature and classified as category 1 (pathognomonic fusions of established clinical utility), category 2 (fusions of potential utility; e.g., targetable fusions), category 3 (fusions in cancer genes), and category 4 (other fusions). Category 1-3 fusions were confirmed by PCR. Results: A total of 7304 fusions (1452 interchromosomal, 5852 intrachromosomal including 4075 read-through events) were identified, including 233 fusions retaining an open reading frame (ORF). Analysis of the 233 ORF-retaining fusions revealed a median of 6 fusions per tumor (range 0-17). In the test set (n=9), RNA-seq correctly identified 8 of 9 expected fusions (EWSR1-WT1, 3 EWSR1-FLI1, EWSR1-ERG, PAX3-FOX01, ASPSCR1-TFE3, 1 BCOR-CCNB3), failing to call a BCOR-CCNB3 fusion that was detected but filtered. In 4 of 28 validation cases with one partner gene known to be rearranged by FISH (EWSR1-, CIC-, EWSR1-, CIC-), RNA-seq identified the corresponding partner genes (ETV1, FOXO4, FLI1, DUX4, respectively). Six category 1 or 2 fusions were identified in the remainder of the validation cases (n=24), including two activating fusions in unexpected tumor types that are potentially targetable (KIAA1549-BRAF: chest wall sarcoma; FGFR1-ADAM32: embryonal rhabdomyosarcoma), a pathognomonic fusion in a novel tumor type (MLL-MLLT10: round cell sarcoma), and in two cases, diagnostic fusions (FUS-CREB3L2: low-grade fibromyxoid sarcoma, and EWSR1-ERG: atypical Ewing sarcoma). In total, therefore, fusion genes of established or potential clinical utility were identified in 10 of 28 (35.7%) validation cases. Conclusions: RNA-seg is a powerful tool for detecting fusions of clinical utility in sarcomas, and may be especially relevant for unbiased detection of therapeutic targets, as evidenced by the unexpected identification of targetable fusions in the analyzed sarcoma cohort.

H30. Systematic Analysis of Common Copy Number Variations in Diffuse Large B-cell Lymphomas

D. Xia¹, P. Michaels¹, A. Guttapalli², R. Joyce¹, J. Houldsworth², R. Dewar¹ ¹Beth Israel Deaconess Medical Center, Boston, MA; ²Cancer Genetics Inc., Rutherford, NJ. Introduction: Diffuse large B-cell lymphoma (DLBCL) is the most common high grade lymphoma in the Western Hemisphere. The disease exhibits a wide spectrum of clinical aggressiveness, and is characterized by heterogeneity at the histologic, cytogenetic, and molecular levels. One goal of this study is to identify novel molecular prognostic biomarkers in DLBCL patients, and to integrate these with existing prognostic biomarkers. In a previous study using publicly available DLBCL datasets, we defined a set of 50 common copy number variations (CNVs) involving 36 genomic loci, for which robust scoring criteria were established (Dias L, Thodima V, Friedman J, Guttapalli A, Mendiratta G, Syrbu SI, and Houldsworth J. December, 2014. Robust Assessment of Genomic Imbalance in Diffuse Large B-Cell Lymphoma Confirms Inferior Outcome Is Associated with Genomic Complexity and Identifies Potential Therapeutic Pathway Targets. Poster presented at: American Society of Hematology Annual Meeting 2014). Here, we present the initial results of an attempt to systematically validate these CNVs using formalin-fixed, paraffin-embedded tissue blocks from an additional group of DLBCL patients. Methods: Patients diagnosed with de novo DLBCL and treated with R-CHOP from 2003 to 2009 at the Beth Israel Deaconess Medical Center (BIDMC) were identified through retrospective review of electronic medical records (EMR). Sections of one formalin-fixed, paraffin-embedded block from each patient were submitted for custom array comparative genomic hybridization (aCGH) at Cancer Genetics Inc., and each sample was scored according to the established criteria for the presence/absence of each of the 50 CNVs. Patient survival data was obtained through the BIDMC EMR. Kaplan-Meier survival analysis was performed using the R statistical software. Bonferroni correction was employed to adjust for multiple testing (cutoff p-value < 0.001). Results: To date, the 50 CNVs were successfully assayed in samples from 45 DLBCL patients with survival data. In univariate analyses, the presence of four of these CNVs were associated with poor survival at a p-value of <0.05. One of the four CNVs (loss at 1p13.1) was significantly associated with poor outcome, after adjusting for multiple testing ($p = 1.43 \times 10^{-5}$). Conclusions: Our ongoing systematic analysis of common copy number variations in DLBCLs identified several biomarkers potentially informative about patient survival. Further validation with additional cases and correlation with immunostains for cell of origin studies (ABC versus GCB subtypes) are in progress. The utility of each potentially useful novel biomarker will be assessed via multivariate analyses along with prognostic clinical variables.