



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

803.EMERGING TOOLS, TECHNIQUES AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

Feasibility and Accuracy of Nanopore Transcriptome Sequencing for Pediatric Acute Leukemia Lineage Classification in a Resource-Constrained Setting

Javeria Aijaz, FCPS, PhD¹, Fouzia Naseer, MSc, MBA¹, Shakir Hussain, MSc, MPhil¹, Muhammad Rafie Raza, DCH, MCPS, FCPS², Saba Jamal, MBBS³, Shamvil Ashraf, DCH, MCPS, FCPS, MRCP⁴, Larissa Furtado, MD⁵, Nickhill Bhakta, MD MPH⁶, Thomas B. Alexander, MDMPH⁷, Jeremy Wang, PhD⁸

¹ Pathology Department, Indus Hospital & Health Network, Karachi, Pakistan

² Pediatric Hematology/Oncology Department, Indus Hospital & Health Network, Karachi, Pakistan

³ Pathology Department, Indus Hospital & Health Network, Karachi, PAK

⁴ Executive Director Medical Services, Indus Hospital & Health Network, Karachi, Pakistan

⁵ Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN

⁶ Department of Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN

⁷ Department of Pediatrics, University of North Carolina, Chapel Hill, NC

⁸ Department of Genetics, University of North Carolina, Chapel Hill, NC

Introduction: Gene expression profiling (GEP) has the potential to offer alternative approaches to conventional testing methodologies (flow cytometry, immunohistochemistry, karyotyping, and FISH) for cancer diagnostics. Possible advantages of this approach include unified protocols, automated analyses obviating high-end technical expertise, and perhaps lower training, capital, and skilled human resource requirements in comparison with complex traditional testing. These factors are of particular value in resource-constrained settings, where even leukemia lineage classification is unavailable or inconsistent in many locations, with critical implications for risk stratification and treatment selection. The goals of this study were to (a) ascertain and subsequently improve the accuracy of a lineage classifier in a different context than the pre-testing stage, and (b) determine feasibility of mRNA sequencing for cancer classification in a low-resource setting. The approach utilized PCR-amplified full-length cDNA sequencing using Oxford Nanopore Technologies (ONT) PCR-cDNA barcoding kit (SQK-PCB109) and MinION Mk1b.

Methods: One hundred and ten pediatric acute leukemia specimens were collected and sequenced at Indus Hospital & Health Network (IHHN), Karachi, Pakistan and analyzed at the University of North Carolina at Chapel Hill (UNC), USA. Specimens (isolated peripheral blood and / or bone marrow mononuclear cells), frozen in liquid nitrogen or Zymo DNA/RNA shield at -80 °C, were acquired from IHHN's pediatric biorepository. Specimen-associated data included clinical history, CBC, flow cytometry, FISH and karyotyping results from routine clinical testing at the IHHN's ISO 15189-accredited clinical laboratory. The study site did not have prior experience with NGS and did not require external hands-on training for this project. 100 ng of extracted RNA from each of 48 B-ALL, 30 T-ALL, 31 AML, and 1 mixed phenotype acute leukemia specimens was used to prepare libraries. Sequencing data were transferred to UNC and analyzed with an acute leukemia lineage classification algorithm.

Results: Following delivery of reagents and equipment, a 6-week protocol optimization and self-training phase utilized open-source protocols and training videos. The next runs to sequence 110 specimens were completed in 4 months by 2 technical staff at the study site. Using a probability threshold for confident calling (>0.8), established during the development phase at UNC, 91 of 110 specimens (82.7%) were called with high confidence. Of the confident calls, 85 were correctly classified (42 B-ALL, 21 T-ALL, and 22 AML), representing an accuracy of 93.4%. The read N50 of confidently classified, concordant specimens ranged from 270 to 745. The total number of reads ranged from 10,000 to >900,000, while the number of aligned reads ranged from 5,000 to 1.5 million. On the other hand, 6 (6.6%) confidently called specimens were discordant. Of these, the mixed phenotype specimen was classified as AML, but mixed phenotype acute leukemia is not included in our training dataset. One misclassified T-ALL specimen had a read count below 10,000 - a threshold demonstrated in the development phase to be optimal for correct classification. Two additional misclassified samples had very low diversity of transcripts express, which is a quality indicator in development. Read counts of these specimens ranged from 22,000 to 311,000, while aligned reads were between 10,000 and 293,000. Further investigation of these discordant calls is ongoing. Low confidence calls

(17% of total), with a classification probability between less than 0.8, were either concordant (13) or discordant (6). Further investigation and retesting of these specimens to ascertain reproducibility of low confidence calls has not been performed yet.

Conclusions: Taken together, the approach tested - with automated analysis, limited space, training and technical staffing requirements - was feasible at the study site. It provides proof of concept for a low-cost, ONT sequencing based GEP approach for acute leukemia diagnostics whereby testing can be performed in peripheral laboratories while analysis can be centralized, even across continents. Next steps include (a) incorporating these data into a new training model with updated quality control parameters for the next validation run, and (b) developing the genomic subtype classification results.

Disclosures No relevant conflicts of interest to declare.

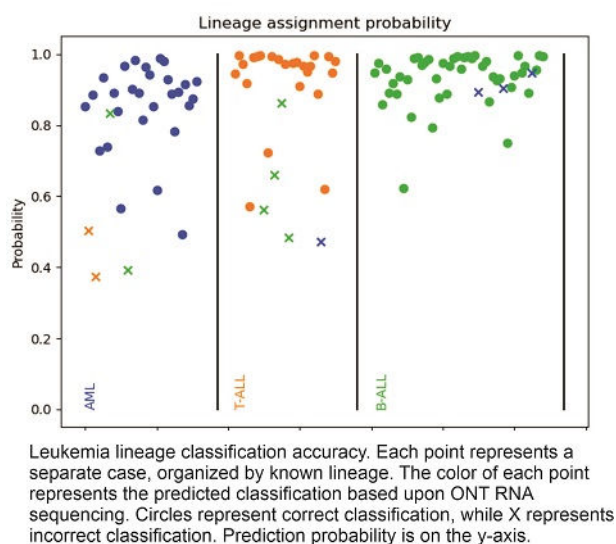


Figure 1

<https://doi.org/10.1182/blood-2023-186696>