A133 · Identifying pathogenic tandem repeat expansions at novel loci in short-read and long-read rare disease datasets

Presenter: Ben Weisburd — Broad Institute

Authors: Ben Weisburd, Matt Danzi, Isaac R. L. Xu, Stephan Züchner, Anne O'Donnell Luria, Heidi L. Rehm

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Background: Short-read and long-read tandem repeat (TR) genotyping tools are now routinely used to detect pathogenic expansions at many of the ~70 known disease-associated TR loci. However, there is not yet consensus on how to effectively use these tools to detect pathogenic expansions at novel TR loci genome-wide.

Material and Methods: We present a TR analysis pipeline that we have developed and applied to discover two novel candidate TR loci. This pipeline starts with the latest version of the genome-wide TR catalog described in [Weisburd, Dolzhenko et al. 2025, PMID:41279208] and runs TRGT as well as TRGT-LPS on this catalog for long-read datasets, while using a modified version of ExpansionHunter along with ExpansionHunter Denovo for short-read datasets. It then annotates the TR genotypes using population data and other publicly-available information provided by our trexplorer.broadinstitute.org portal. Finally, it loads the data into a novel analysis tool with a graphical user interface that enables users to filter and prioritize the annotated TR calls.

Results: The two candidates identified to date include a CAG repeat in the coding region of the EP400 gene identified in two unrelated families with cerebellar ataxia. The second candidate is a TGC repeat in the intron of the CNTN4 gene where expansions may explain a development delay and autism phenotype in a single family.

Conclusion: We anticipate that these publicly available tools and pipeline will enable discovery of additional novel TR candidate loci in cohorts ranging in size from a single genome to over 100,000 samples.