# A191 · Robust dynamics of somatic short tandem repeat expansions using donor-specific assembly

**Presenter:** Suhas Rao — Harvard Medical School, Department of Biomedical Informatics

**Authors:** Suhas Rao, Peter J. Park

**Session:** Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Short tandem repeat (STR) expansions are associated with many neurodegenerative disorders such as Huntington's disease. However, difficulties in assembling and aligning to repetitive regions in the genome, have severely limited the study of somatic STR expansions. Using long-reads and recent techniques for building personalized/donor-specific assembly (DSA), we can study the true dynamics of these expansions in normal tissue.

We refined multiple previously curated pathogenic and nonpathogenic STR catalogs (based on hg38 + short-reads), filtering each by >20% into a final set of ~200,000 loci with DSA-resolved coordinates. We then benchmarked how using DSA alignments resolves notable error modes generated by existing long-read STR genotyping tools (TRGT, Medaka, etc.) - 1) phasing repeat expansions to the appropriate allele, 2) missing true somatic STR expansions due to upstream misalignment and poor reference genome resolution, and 3) including false positive STR expansions by mistakenly incorporating nearby low-complexity regions.

We've identified multiple cases of extreme somatic STR expansions in brain tissues within normal tissues, up to >5000bp (100x) larger than the corresponding reference allele. We've documented dozens of cases where standard SNP/indel-based phasing are unable to separate biologically distinct repeat alleles (where one allele is stable / reference size, and the other is >50bp and >100% relatively larger) and demonstrated how DSA/local realignments can accurately phase these. We've also demonstrated that using DSA methods allows existing STR genotyping algorithms like TRGT to rescue up to 10% additional reads near certain pathogenic STR loci such as NOTCH2NLC (which falls within collapsed paralogs in hg38).