TANDEM REPEAT ANALYSIS

Unprecedented detection and discovery of repeat expansions

Tandem repeats play a key role in human health and disease

Tandem repeats (TR) are repeating sequences of two or more base pairs that are adjacent to one another and are abundant thoughout the genome. Because of their repetitive nature, they are hypermutable, and they play a key role in human health and disease.[1] Expansions in repeat length in certain ranges — typically longer repeats — can become pathogenic. More than 50 diseases are known to be caused by TR expansions,[1] and further study could reveal associations with more rare diseases that are currently unexplained.

Together with HiFi sequencing, TRGT enables researchers to use a single library prep to:

- **Accurately quantify repeat count**
- **Identify interrupting sequences**
- **Determine allele phasing**
- **Determine 5mC profile**

Pathogenic repeat expansions are often longer than the read lengths of short-read platforms, and their repetitive nature makes them difficult or impossible to map. HiFi reads are highly accurate, over 100 times longer than short reads, and offer simultaneous 5-Methylcytosine (5mC) detection.

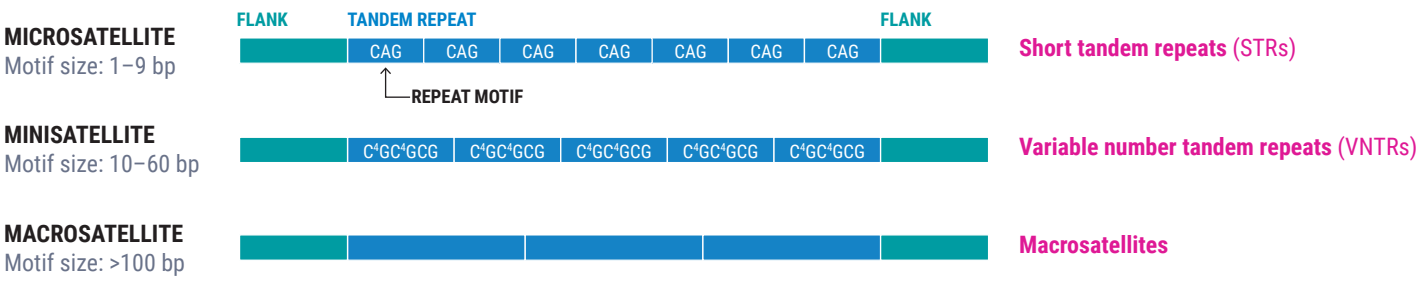
HiFi reads can span most repeats, making it possible to comprehensively profile variation in repeat regions and to measure any changes in methylation. PacBio has developed the *Tandem repeat genotyping tool* (TRGT) and *Tandem repeat visualizer* (TRVZ) as bioinformatics tools to accurately profile short tandem repeats, variable number tandem repeats, and macrosatellites.

Most abundant class of variation in the human genome[1]

- >1M tandem repeats in the human genome
- >10 higher mutation rate than any other variant class

Known to cause disease

- >50 repeat expansion disorders caused by STRs[1]
- Several VNTRs linked to diseases like Alzheimer's, autism, epilepsy, and ALS[2, 3]



PacBio

TRGT is comprehensive, powerful, and practical

PCR and Southern blot assays, which are often used for repeat expansion detection, are limited to analyzing one or a few loci at a time. Short-read whole genome sequencing (WGS) is more comprehensive but offers limited repeat expansion resolution due to read-length limitation.

WGS with HiFi sequencing offers researchers unmatched information density from their samples, by enabling interrogation of genome-wide TRs, structural variants, and 5mC modifications from a single library prep.

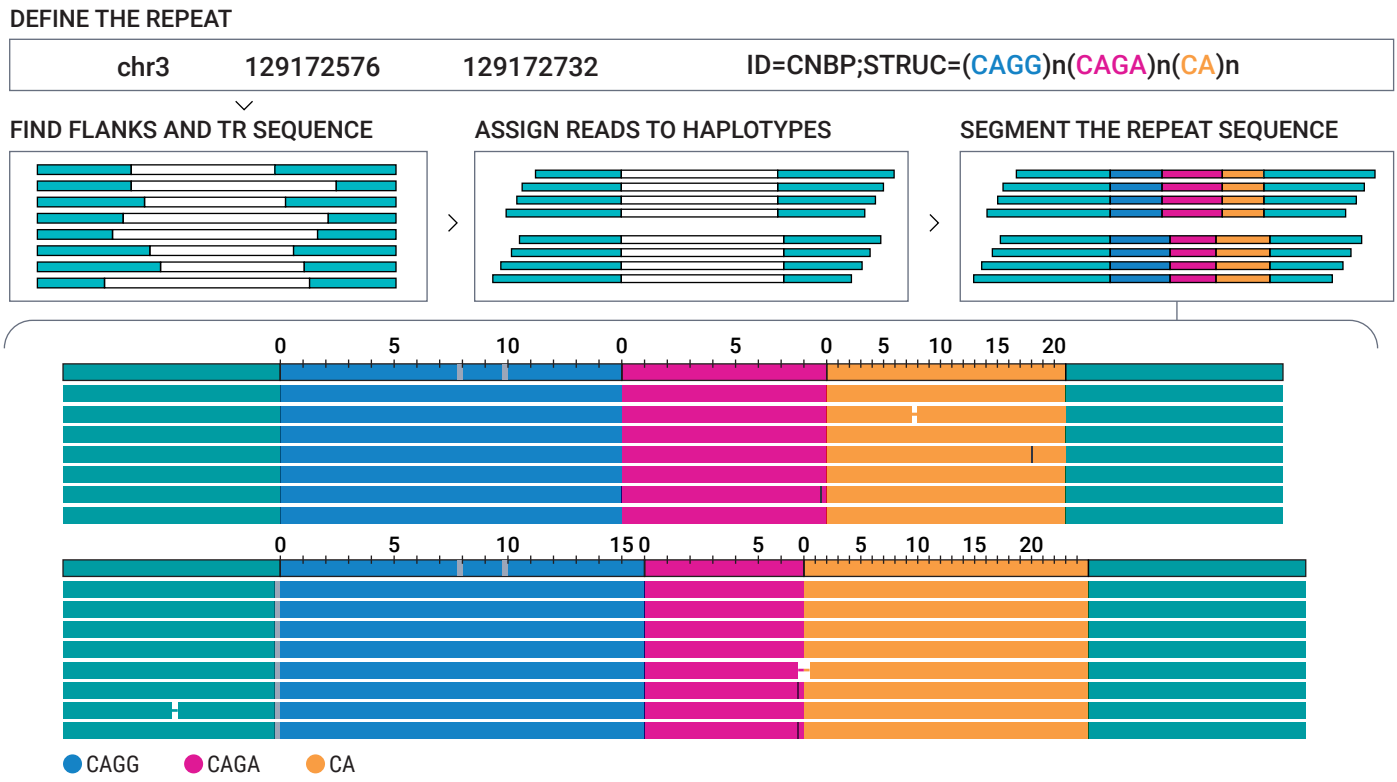


Figure 1: TRGT includes multiple TR catalogues to enable and simplify analysis. As an example, the definition (location and structure) of the CNBP repeat is included within the catalog of pathogenic repeats. TRGT locates reads that span the region, assigns them to haplotypes, and determines the structure of the resulting repeat alleles, which can be visualized using TRVZ.

KEY REFERENCES

1. Madsen, B. E., Villesen, P., & Wiuf, C. (2008). Short tandem repeats in human exons: a target for disease mutations. *BMC genomics*, 9, 410.
2. Ryan, C. P. (2019). Tandem repeat disorders. Evolution, medicine, and public health. (1), 17.
3. Paulson, H. (2018). Repeat expansion diseases. *Handbook of clinical neurology*, 147, 105–123.
4. Dolzhenko, E., et al. (2023). Resolving the unsolved: Comprehensive assessment of tandem repeats at scale. bioRxiv 2023.05.12.540470.

102-326-546 REV03 26MAY2023