Article overview: "Analysis and benchmarking of small and large genomic variants across tandem repeats"

TR Catalog Development and Genome-Wide Study

Objective:
Create a comprehensive catalog of Tandem Repeat (TR) regions to promote the development of TR analysis methods.
Focus on 86 haplotype-resolved long-read human assemblies.

Key Steps:

1. Curated variants from the Genome in a Bottle (GIAB) HG002 individual to generate a robust dataset.
2. Developed a TR dataset to serve as a benchmark for future TR analysis methods.
3. TRs often excluded due to variant calling and representation challenges, now addressed by this catalog.

New Variant Comparison Method and Truvari

Improved Variant Comparison:
Introduced a method to handle variants greater than 4 bp in length and variable allelic representations.
Essential for large-scale genome-wide studies including TRs.

Truvari Tool:
Used to benchmark structural variants (SVs) and validate TRs.
Improves accuracy in identifying complex genomic variants like indels and larger structural changes.
Key to comparing and assessing variant calls in benchmarking projects.

- "Analysis and benchmarking of small and large genomic variants across tandem repeats"

Utility of LaTyr and Study Results

LaTyr Pipeline:

- Tailored for Tandem Repeat (TR) detection in long-read sequencing data.
- Essential for analyzing larger repeat expansions that may be missed in short-read sequencing.

Study Results:

TR regions represent 8.1% of the genome, but contain ~24.9% of variants per individual.
The GIAB HG002 benchmark set includes 124,728 small variants and 17,988 large variants.
Applicable across short-read and long-read sequencing technologies, supporting versatile analysis.

- "Analysis and benchmarking of small and large genomic variants across tandem repeats"

Application

TR Data Integration for Rare Disease Variant Detection:
The TR catalog and variant analysis methods can improve identification of causative variants across different variant types, including TRs, SNVs, indels, CNVs, and SVs.

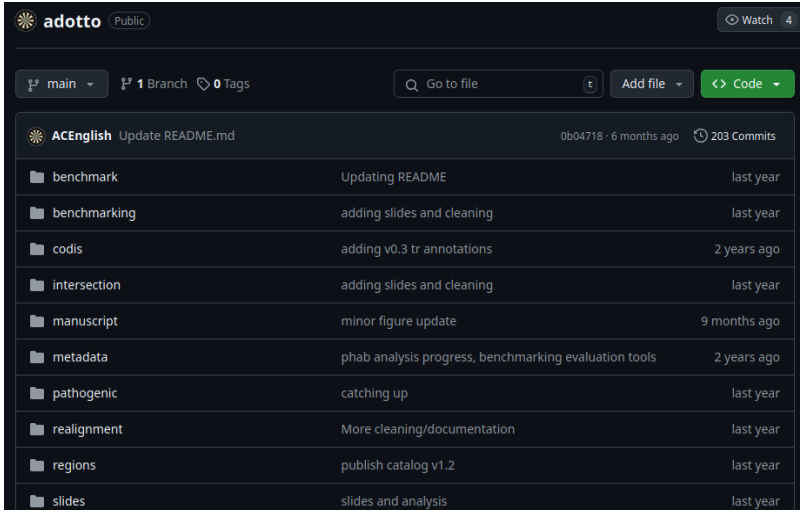
Truvari and LaTyr Use:
Truvari helps validating structural variants and improve accuracy in variant calling, especially with complex indels and SVs in our patients.
LaTyr will support the detection of repeat expansions in TRs, potentially uncovering previously missed variants relevant to developmental disorders.

Application to Whole Genome Sequencing (WGS):
Improving our variant discovery and interpretation.

- "Analysis and benchmarking of small and large genomic variants across tandem repeats"

Supplementary materials

- **TR catalog and pVCF files**: These contain the **full tandem-repeat catalog** and **variant call format files** for the **86 haplotype-resolved genome assemblies** used in the study. The files can be downloaded from **Zenodo** as indicated in the repository.
- **Benchmarking Scripts**: The repository contains scripts for generating the **benchmark** used in the paper, allowing users to reproduce or extend the analysis.
- **Manuscript Materials and Plotting Workflows**: This includes summaries, plotting workflows, and slides for the GIAB team meetings.



Benchmark link (git hub): git_hub_repository