Supplemental Materials

Estimating the size of long tandem repeat expansions from short reads with ScatTR

Rashid Al-Abri[3,2] and Gamze Gürsoy[1,2,3*]

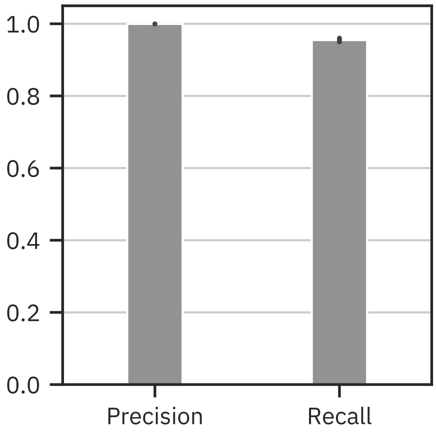
[1]Department of Biomedical Informatics, Columbia University, New York, USA
[2]New York Genome Center, New York, USA
[3]Department of Computer Science, Columbia University, New York, USA
[*]Corresponding author: Gamze Gürsoy, gamze.gursoy@columbia.edu

List of Supplemental Materials

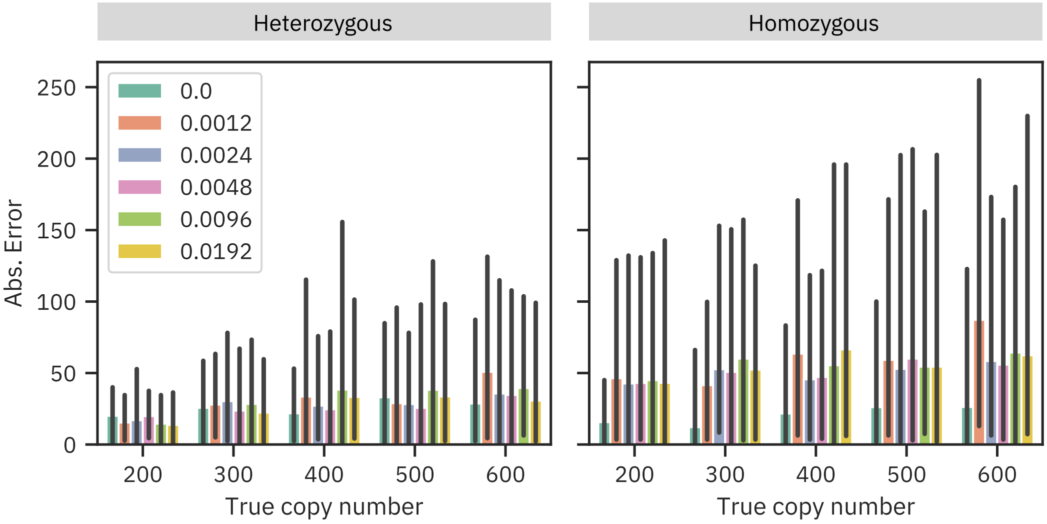
- **Supplemental Figures**
  - **Supplemental Figure S1.** Precision and recall of IRR extraction from 540 simulated WGS samples.
  - **Supplemental Figure S2.** Absolute error of ScatTR across varying sequencing error rates in simulated WGS samples. Error bars represent 95% confidence intervals across 30 TR loci.
  - **Supplemental Figure S3.** Comparison of the estimated sum of alleles for the *C9orf72* repeat expansion by ScatTR, ExpansionHunter, GangSTR, and STRling. The analysis was performed on short-read WGS data from 15 ALS patients with confirmed expansions.
- **Supplemental Tables**
  - **Supplemental Table S1.** Regions and motifs of tandem repeat loci used for benchmarking RMSE of ScatTR and state-of-the-art methods

- **Supplemental Table S2.** Regions and motifs of tandem repeat loci with long motifs (21-50 bp) used for benchmarking RMSE of ScatTR and state-of-the-art methods
- **Supplemental Table S3.** Regions and motifs of tandem repeat loci used for benchmarking effect of mutations and sequencing error on ScatTR's performance
- **Supplemental Table S4.** Sample IDs from the ALS cohort used in this study

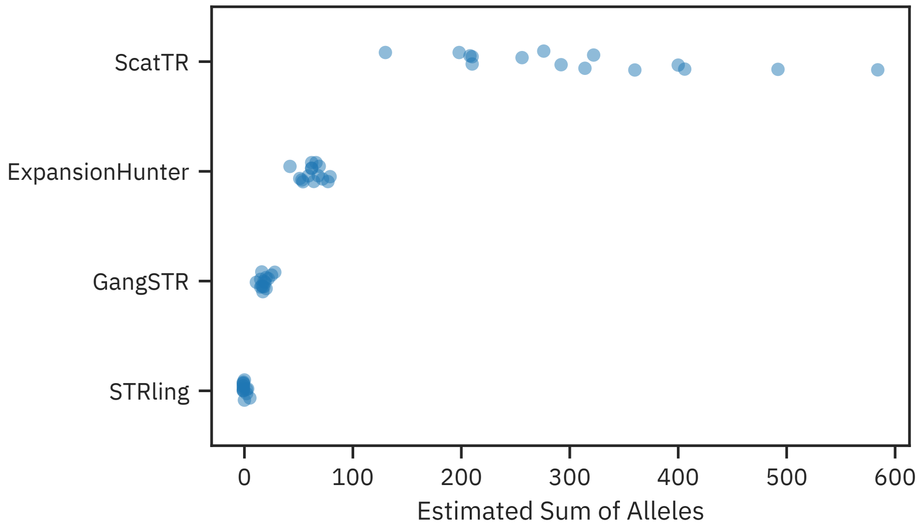
Supplemental Figures



Supplemental Figure S1. Precision and recall of IRR extraction from 540 simulated WGS samples.



Supplemental Figure S2. Absolute error of ScatTR across varying sequencing error rates in simulated WGS samples. Error bars represent 95% confidence intervals across 30 TR loci.



Supplemental Figure S3. Comparison of the estimated sum of alleles for the *C9orf72* repeat expansion by ScatTR, ExpansionHunter, GangSTR, and STRling. The analysis was performed on short-read WGS data from 15 ALS patients with confirmed expansions.