Parliament: The Consensus Structural Variant Calling Framework

SUMMARY

DESCRIPTION

Parliament is a consensus structural variant (SV) calling framework, which combines multiple data types and runs a variety of SV detection methods.

DEVELOPERS

Human Genome Sequencing Center, Baylor College of Medicine

CHALLENGE

- Tedious to install & maintain multiple discovery tools
- Requires substantial systems admin & computational resources
- Harmonizing disparate data types and multiple SV detection methods

SOLUTION

DNAnexus simplifies SV analysis by providing a polished cloud-based Parliament application with robust parameters that is easy to run and cost-efficient.

OVERVIEW

Characterizing large genomic variants is a difficult problem. Although multiple data types and methods are available to detect structural variants (SVs), they remain less understood and more challenging to detect and characterize than smaller variants. This is because of SV diversity, complexity, and size.

Parliament is a publicly available consensus structural variant calling framework that combines multiple data types (e.g. Illumina and PacBio reads) and runs a variety of SV detection methods (e.g. PBHoney, Breakdancer, Delly). By using both long-read and short-read data, Parliament calls SVs with an accuracy not possible with short-reads alone and while only requiring a modest coverage of long-read data. This extensible framework also allows for the incorporation of novel SV detection methods. Co-developed by Baylor College of Medicine's Human Genome Sequencing Center with DNAnexus, Parliament has high concordance when evaluated with gold standard truth sets (e.g. HS1011 and Genome in a Bottle).

THE CHALLENGE

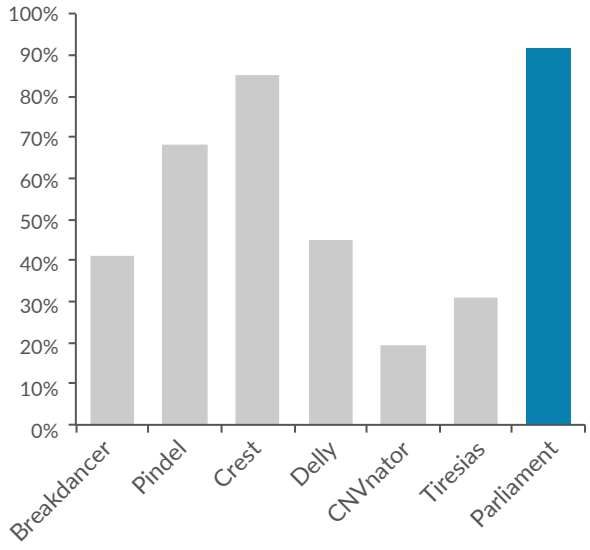
Typically, short-read sequencing technologies are not able to span the entire structural variant, which can be tens of thousands of base pairs. In addition, the amount of data from long-read sequencing technologies in human genomes is sparse. The lack of a reference set for structural variants by which to compare calls further complicates matters. Parliament helps alleviate these challenges by casting a wide net to discover SVs and then leverages multiple data types to refine them.

The many complex short-read SV calling tools, which compose the Parliament framework, impose time consuming installation and the ongoing burden of maintenance. Rather than wrangle dependencies and manage local compute resources, DNAnexus allows scientists to shift their focus on discovering and interpreting SVs. In order to make this tool available for research labs of all sizes, DNAnexus installed Parliament on its leading cloud-based genome informatics and data management platform to simplify the SV calling and analysis process.

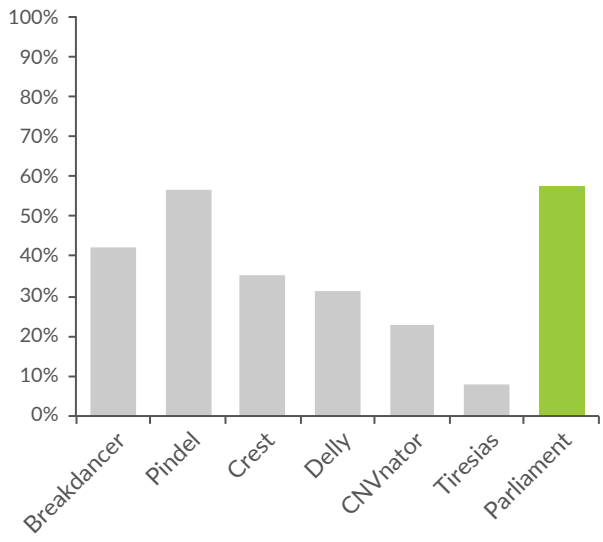
THE RESULTS

Parliament has been shown to identify nearly 10,000 putative SVs through hybrid assembly, long-read PacBio data, and multi-source heuristics. By combining data from a variety of SV callers, Parliament offers higher specificity and sensitivity than other SV detection methods, increasing the probability of discovering true structural variants.

Specificity of Callers



Sensitivity of Callers



Data from Illumina only SV callers.

THE BENEFITS

- Expert bioinformatics support for complex structural variation analyses
- Fast, accurate, and cost-efficient
- Eliminate complicated long-read applications installations
- No local compute infrastructure requirements
- Easily deploy additional bioinformatics tools (e.g. de novo assemblers)

GET STARTED

1. Direct data upload of Illumina FASTQ files or BLASR mapped BAM files
2. Copy featured Parliament Project on DNAnexus to your own project
3. Call SV candidates on short-read data
4. Run Parliament workflow with outputs of STEP 3
5. Completion time: 1 day for whole genome (30x Illumina + 10x PacBio)

Reference Article: English, AC, Salerno, WJ, et al. Assessing structural variation in a personal genome--towards a human reference diploid genome. BMC Genomics. 2015;10.1186/s12864-015-1479-3

DNAnexus®