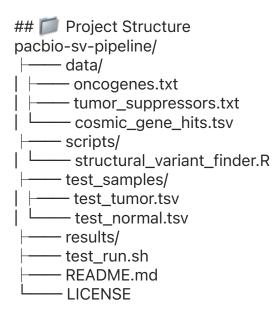
## # PacBio Structural Variant (SV) Pipeline

## ## 8 Overview

This pipeline is designed for the annotation and clinical prioritization of structural variants (SVs) detected from PacBio HiFi long-read sequencing data. It is tailored for somatic cancer variant discovery and classification, particularly in leukemia cases.

## ## **V** Key Features

- Germline filtering using tumor-normal comparison
- Functional annotation via known oncogenes, tumor suppressors, and cancer driver genes
- Clinical relevance scoring using real-time API queries to:
- CIViC
- OncoKB
- COSMIC (local fallback)
- SV-specific scoring based on size, type (e.g., fusions, deletions, duplications)
- ACMG-style clinical classification: Pathogenic, Likely Pathogenic, VUS, Benign
- Excel output with scoring breakdown and logs for each sample



## # How to Run ### Prerequisites

- R ≥ 4.0
- R packages: `dplyr`, `stringr`, `openxlsx`, `readr`, `httr`, `jsonlite`, `clusterProfiler`, `org.Hs.eg.db`, `ggplot2`
- API keys:
- CIViC: https://civicdb.org
- OncoKB: https://oncokb.org

## ### Command

```bash

Rscript scripts/structural\_variant\_finder.R data/sample01\_tumor.tsv

data/sample01\_normal.tsv

Outputs

sample01\_clinical\_significance.xlsx: Annotated SV table with gene, SV type, clinical scores, and classification

sample01\_pipeline\_log.txt: Full log of the run with error handling notes

Customization

Replace gene lists in data/ with custom panels

Replace COSMIC annotations or expand fallback sources

Adjust scoring thresholds inside the R script

Example Test Run Run the test samples with:

bash test\_run.sh