SUMMARY: This project presents a custom rule-based variant filtering pipeline for Next-Generation Sequencing (NGS) data, focused on clinical relevance and quality assurance. Using a structured logic framework, variants were assessed across multiple dimensions, including:

- **Gene-specific filters** (e.g., *CHEK2*, *KMT2C*, *ARID1A*)
- **Variant caller agreement** (Mutect and Pindel)
- **VAF and depth thresholds**
- **COSMIC hematologic evidence**
- **Row-level concordance between matched files**
- **Column-specific subfilters** (e.g., small indels, subset of metadata)

The filtering logic was encoded into decision rules to categorise each variant as **Retain** or **Discard**. Visual summaries, including tables and bar plots, were generated to illustrate:

- The proportion of retained vs discarded variants
- Match vs mismatch statistics
- Distribution of filtered variants across rule categories

Key Outcomes

- **60 variants retained** after stringent filtering
- **40 variants discarded** due to rule violations
- **100 rows** identified with mismatches
- **Variant caller agreement** and **COSMIC evidence** were critical factors in retention

This repository includes:

- Filtering logic implementation scripts
- Summary tables and plots
- Text outputs for review and reproducibility

This framework supports transparent and reproducible variant filtering decisions in clinical bioinformatics workflows.