

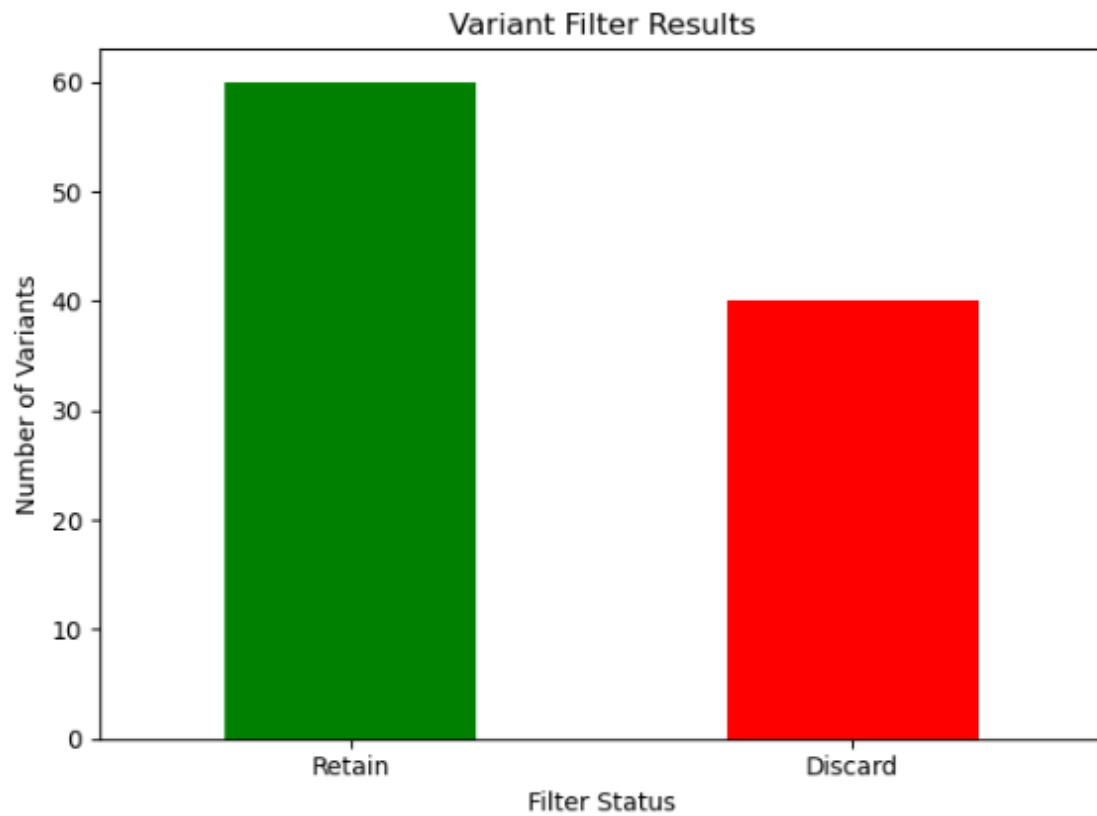
**Table 1.** Rule-based Criteria for NGS Variant Filtering and Classification.

This table summarises the logical conditions and rules applied to classify next-generation sequencing (NGS) variants as **Retain** or **Discard** during filtering. The filtering logic integrates multiple variant-level features such as gene association, variant allele frequency (VAF), variant length, caller confidence (e.g., from MuTect or Pindel), and biological relevance based on databases like COSMIC.

Rule/Condition	Action
<b>CHEK2 + panhaem_myeloid</b>	Discard
<b>Variant caller ≠ both &amp; vcf_mutect &gt; 10 or vcf_pindel &gt; 10 + (ASXL1 specific variant or cosmic_haem ≥ 5)</b>	Retain
Otherwise under above rule	Discard or Retain accordingly
<b>Variant caller = both &amp; both vcf_mutect and vcf_pindel &gt; 10 + (ASXL1 specific or cosmic_haem ≥ 5)</b>	Retain
Otherwise under above rule	Retain
<b>KMT2C or ARID1A + (vaf &lt; 0.20 or length &gt; 10)</b>	Discard
Otherwise	Retain
Default	Retain

**Fig 1.** Final Variant Filtering Outcome.

Bar plot showing the total number of variants categorised as **Retain** (green) or **Discard** (red) after applying rule-based filtering criteria. **60 variants were retained** for downstream analysis based on gene relevance, variant caller agreement, VAF/depth thresholds, and clinical evidence. **40 variants were discarded** due to insufficient support from filtering rules or mismatched criteria.



**Table 2.** Breakdown of Variant Filtering Metrics Based on Custom Logic. This table summarises the distribution of filtered variants across various rule-based categories applied during quality control. Key filtering criteria include gene-specific flags (e.g., CHEK2, KMT2C/ARID1A), mismatch detection between datasets, variant length and depth checks, caller consistency, and other sub-feature thresholds. Notably, mismatch rows and variant caller validations accounted for the largest portions of filtered outputs.

Metric	Count	Description
CHEK2	9	Variants involving CHEK2
KMT2C or ARID1A	16	Variants involving KMT2C or ARID1A
Exact Matches	0	Exact matches between datasets
Mismatches	100	Rows with differences
Small Variants	1	Possibly small length differences
Sub Columns	2	Subset of filtered columns
VAF/Depth Checks	6	Variants passing vaf/depth filters
Variant Calls	96	Variants passing caller-related checks

**Figure 2.** Variant Filtering Summary by Rule Category. This bar plot provides a descriptive overview of each filter metric represented in Table 2 above. It details the number of variants falling into each category and a short explanation of the filtering rationale, helping contextualise the variant exclusion and retention process.

