

QC per SampleID Summary Report

Jacquemont Lab

Dataset: SPARK_Array_GRCh37

Date: 08/09/2025

Sample Count by Sex

Table 1: Number of Samples by Sex

| Sex | Count |
|---------|-------|
| female | 30019 |
| male | 39492 |
| unknown | 414 |

Filter Quality Control

According to Liu, J., Zhang, L., Xu, L. *et al.*

Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array.

BMC Genomics 14, 229 (2013). <https://doi.org/10.1186/1471-2164-14-229>

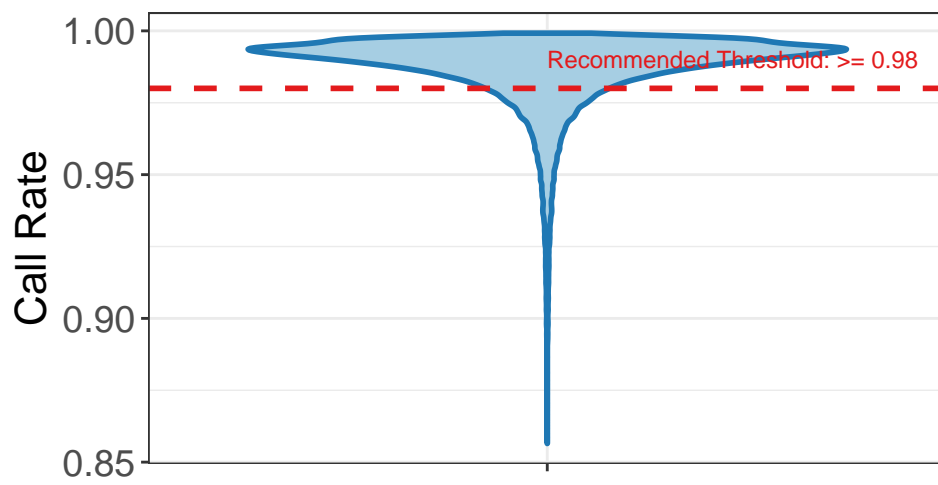
The quality filters recommended were:

- **LRR standard deviation (SD) < 0.30**
- **BAF drift < 0.01**
- **Waviness factor between -0.05 and 0.05**

In addition, based on common practice in CNV detection studies (e.g., PennCNV and Illumina genotyping guidelines), we also recommend applying:

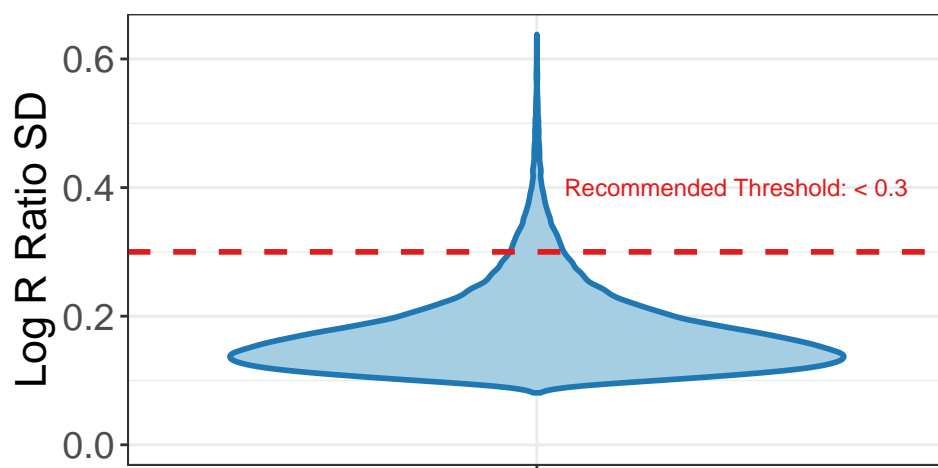
- **Call Rate ≥ 0.98**

Call Rate Distribution



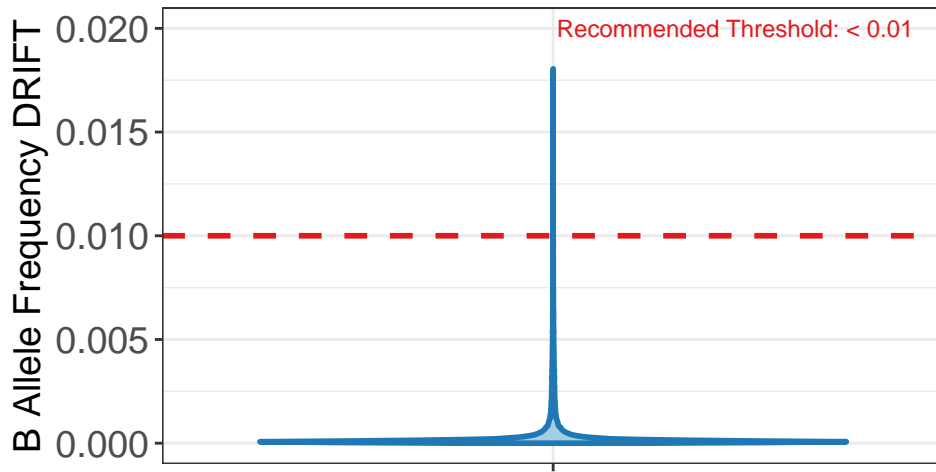
Number of samples NOT passing Call Rate threshold (≥ 0.98): 13620

LRR_SD Distribution



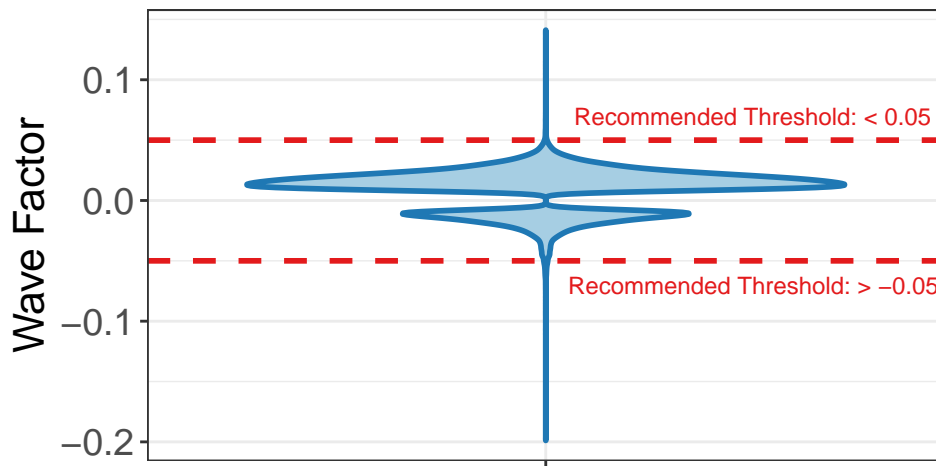
Number of samples NOT passing LRR_SD threshold (< 0.3): 3829

BAF_DRIFT Distribution



Number of samples NOT passing BAF_DRIFT threshold (< 0.01): 231

WF (Wave Factor) Distribution



Number of samples NOT passing WF threshold (between -0.05 and 0.05): 242

Table 2: Summary of Samples Failing QC Thresholds

| total_samples | fail_Call_Rate | fail_LRR_SD | fail_BAF_DRIFT | fail_WF | pass_all |
|---------------|----------------|-------------|----------------|---------|----------|
| 69925 | 13620 | 3829 | 231 | 242 | 56184 |