etailed Examples

otype Determination

t Calling

G)] / P(D)

P(D | G): Likelihood of data given genotype

10177 (A→G)

Genotype: G/G

Observed: 1A, 5G reads

P(D|G/G) = 0.413

P(G/G|D) ≈ 0.81 (81%) ✓

nd selects the one with the highest
read (A), the model correctly identifies G/G
equencies. The low probability for A/A

3. Read Depth & Allele Frequency Analysis

Interpreting Coverage and Allele Balance

Example 1: Heterozygous SNV (A/G)

Aligned Reads (DP=20):

Allele A: (9 reads, 45%)

Allele G: (11 reads, 55%)

Total Depth (DP): 20
Allele Frequency (AF): 0.55 (55%)
Expected for Het: ~0.5 (50%)
✓ **Consistent with A/G genotype**

Example 2: Homozygous Variant (G/G)

Aligned Reads (DP=25):

Allele A: (1 read, 4%)

Allele G: (24 reads, 96%)

Total Depth (DP): 25
Allele Frequency (AF): 0.96 (96%)
Expected for Hom: ~1.0 (100%)
✓ **Consistent with G/G genotype**

Example 3: Low Coverage (Unreliable)

Aligned Reads (DP=3):

Allele A: (1 read, 33%)

Allele G: (2 reads, 67%)

Total Depth (DP): 3
Allele Frequency (AF): 0.67 (67%)
Minimum recommended: DP ≥ 10
✗ **Insufficient coverage - unreliable call**

Example 4: Strand Bias (Artifact)

Aligned Reads (DP=20):

Forward (+): 10 A reads 0 G reads

Reverse (-): 0 A reads 10 G reads

Fisher Strand Bias (FS): 85.2
Threshold: FS < 60 for SNVs
Variant only on one strand
✗ **Likely sequencing artifact**

Coverage & Allele Balance Interpretation

Read Depth (DP): Higher coverage provides more statistical power. Minimum 10x recommended, 30x or higher ideal for clinical applications. **Allele Frequency (AF):** For diploid organisms, heterozygous variants should have AF ≈ 0.5, homozygous variants AF ≈ 1.0. Deviations may indicate sequencing bias, copy number variations, or sample contamination. **Strand Bias:** True variants should appear on both DNA strands equally; systematic bias suggests technical artifacts.

4. Haplotype-Based Variant Callin

Local A

Step 1:

Reference: ...ATCGATCG AAAA TCGATCGATCG

Active Region (H

S

Reads in active region:

Read1: ...ATCG AAA TCGATC G GATCG...

Read2: ...ATCG AAAA TCGATC C GATCG...

...and 13 more reads supporting these patterns...

Assembled Haplotypes:

H1: ...ATCG-AAA-TCGATC-G-GATCG... (1bp deletion +

Step 3:

H2/H2 (Ref/Ref)

Both chromosomes match reference

Likelihood: 0.001

P(H2/H2|reads) = 2%

Advantages of Haplotype-Based Calling

Traditional variant callers evaluate each posi
variants. Haplotype-based methods like GAT
haplotypes in regions of variation. This appr
phases nearby variants that occur on the sa

n Variant Calling

rmula

rror)

l or variant call is incorrect

Table

Accuracy	Interpretation
90%	Low quality
99%	Moderate
99.9%	Good (Standard)
99.99%	High quality
99.9999%	Very high quality

10-fold decrease in error probability. A
neaning we accept only variants with
are combined with mapping quality and

Better handles complex variants like MNPs (
than multiple independent SNVs.