

# **PairHMM Test Framework**

# Description of Test Data

- Test file: verificationTest.bz2
- Test file needs to be unzipped (bzip2)
- Unzipped test file: verificationTest.txt
- HaplotypeCaller on the NA12878 chromosome 20: 100000000-200000000
- List of lines containing arguments to PairHMM.  
computeReadLikelihoodGivenHaplotypeLog10() along with the likelihood
- Separated by whitespace, ends with newline
- haplotypeBases readBases readQuals readInsQuals  
readDelQuals overallGCP haplotypeStart  
reachedReadValue likelihood

# Use of Test Data

- Each line represented by TestRow
- PairHMM requires initialization for each unique batch
- batches are composed by a set of rows with the same haplotypeBases
- Initialization of a batch performed by PairHMM.  
initialize() requires two parameters:
  - Maximum length of haplotypeBases
  - Maximum length of readBases
- we determine the grouping of each batch by :
  - check for `haplotypeStart == 0 && haplotypeBases == currentHaplotypeBases`

# Checking Test Results

- PairHMM is ran with the arguments extracted from the test file
- Result is compared with the expected result extracted from the test file
- Precision = 0.001, can be adjusted
- Error > Precision = test case rejected

# Modifying Test Framework

- Possible to use different implementation of PairHMM
- Initialization calls in RunTest.java must be verified
- Likelihood calculation calls in RunTest.java must be verified