

# BayesQual

IMPROVES THE BQSR STAGE OF GENOMICS PIPELINES  
SMALLER FILES, BETTER GENOTYPING ACCURACY

MAPPING

LOCAL REALIGNMENT

DUPLICATE MARKING

BQSR

VARIANT CALLING

## Overview

BayesQual improves upon the BQSR stage of genomics pipelines. It is a command line tool that is easy to deploy into existing pipelines.

By using BayesQual, users benefit from improved genotyping accuracy, as well as much smaller storage footprints.

## Background

Base Quality Score Recalibration (BQSR) is a standard stage in many bioinformatics pipelines that is used to correct quality scores just in case a sequencer is producing incorrectly calibrated results. It is a standard component in widely used pipelines including GATK and htlib genotyping pipelines.

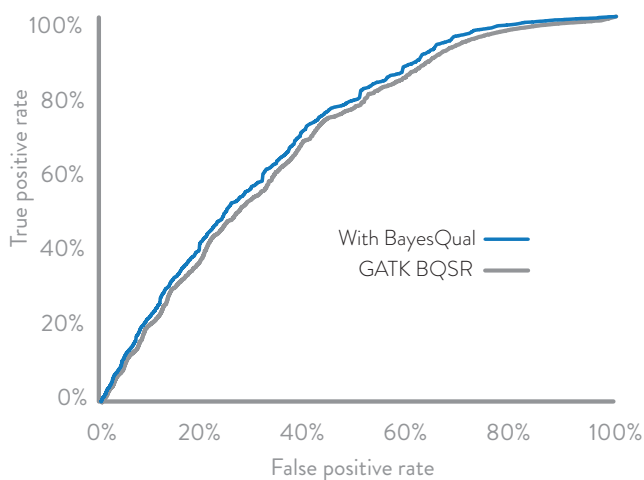
## Smaller and more accurate

BayesQual improves upon GATK's BQSR by adjusting quality scores using a Bayesian model of sequencing error. It combines k-mer analysis with a background corpus to better estimate a posterior probability of sequencing error.

The resultant files typically have better genotyping accuracy than with GATK's BQSR, and the BAM files are also 3-4x smaller. When stored as CRAM files they are 7-8x smaller than the BAM file from GATK's BQSR, and 5-6x smaller than the CRAM equivalent. We have demonstrated that this method doesn't introduce reference bias effects or adversely affect the discovery of rare variants.

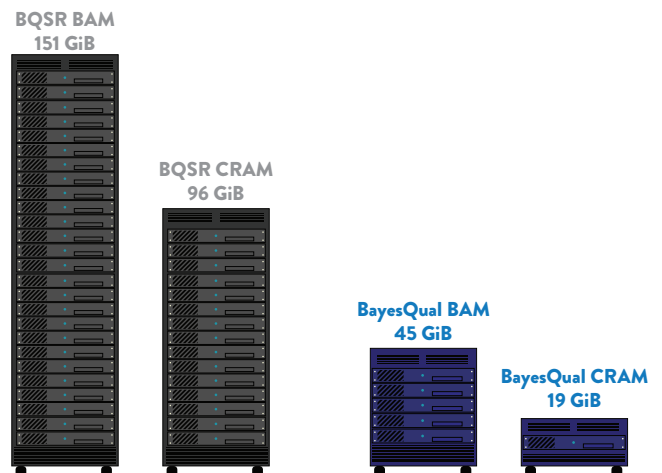
### Genotyping Accuracy ROC

SRR622461 (NA12878) human on GATK HC pipeline



### BayesQual reduces file sizes

NA12878J human WGS at 30x (HiSeq X)



For more information, please contact us:  
[info@petagene.com](mailto:info@petagene.com)  
[www.petagene.com](http://www.petagene.com)

 **PetaGene**  
SMALLER, FASTER GENOMICS DATA