由于现在测序方法和算法有很大不一致，需要一个高精度的基因型标准

In this report, we describe a highly confident set of genome-wide genotype calls that can be used as a benchmark.

NA12878是什么？

属于人类千人基因组计划，NA12878是母本系，除了GIAB， Illumina公司也构建了类似的NA12878高可信度参考突变集合，称为“白金基因组，  白金基因组则是综合考虑了家系，做出了更大面积和数量的高可信度参考集合，瓶中基因组联盟发布了标准基因型的参考数据集，为人们提供了一组高度可信的单核苷酸多态性（SNP）、插入缺失和纯合子参考基因型的测序检测结果。

NA12878, the pilot genome for the Genome in a Bottle Consortium.

通过对5个测序技术的14个数据集、7个读取器和3个不同的调用者之间的集成和仲裁来减少对任何方法的偏见

比较五个测序平台上整合11个完整的人类基因组和三个外显子组数据集

HapMap/1000 Genomes CEU female NA12878, which is a prospective Reference Material (RM) from the National Institute of Standards and Technology (NIST)

To develop these data, we are developing methods to arbitrate between results from multiple sequencing and bioinformatics methods. The resulting arbitrated integrated genotypes can then be used as a benchmark to assess rates of false positive (FPs, or calling a variant at a homozygous reference site), false negatives (FNs, or calling homozygous reference at a variant site), and other genotype calling errors (e.g., calling homozygous variant at a heterozygous site).

基因组分析工具包（GATK）包括使用链偏置相关的VQSR模块，映射质量，等位基因平衡，reads的位置等来过滤潜在的错误。GATK采用怀疑真阳性变种找到过滤的FPS最佳方式，同时保留可能真阳性变种指定的敏感性高斯混合模型。

A couple methods have recently been proposed by the 1000 Genomes Project to integrate multiple variant call sets, but these methods have not been used to arbitrate between datasets from different sequencing methods on the same genome.