From: John West
To: Mark Pratt
Subject: Comparing methods to determine SNP error rate
Date: Thursday, May 3, 2012 12:13:18 PM
Attachments: Methods to determine SNP error rate 3May2012 JW.pptx

---

Hi Mark,

I would like our accuracy development options to be prioritized by quantitation of errors by type & underlying mechanism. So far, I don't feel we have a handle on the SNP error rate, and the data we do have (Rick Dewey's MIE analysis from late 2010) seems to be very different from what Hugo et al found when comparing Illumina vs Complete Genomics. I took a few minutes and put together a slide with what seem to be some of the options to address this, and am attaching it here. Can we discuss this when you get a chance ?

Thanks,

John

Determining SNP Error Rate

Quartet data

- Process with modern HugeSeq (done)
- VCF's aligned across four family members, including non-calls, coverage, allele Q-scores
- MIE Analysis
- Link MIE rate to error rate, type, mechanism

CEPH1463 CGI data

- MIE Analysis
- Link MIE rate to error rate & type

NA12878 SRA data

Comparisons

- Same library, two runs
- Same protocol, two libraries
- Broad HiSeq vs CGI (vs array ?)
- Max quality alignment by combining 10 paired-end insert lengths & very high coverage; Then compare individual runs to this standard

Single genome

- Where is coverage insufficient ?
- Where are allele Q-scores too low ?
- Tri-allelic / triple haplotype sites
- Allelic bias (requires very high coverage)

CEPH1463 family new data

- Comparison ILMN / CGI / Genotyping
- MIE path like quartet

Personalis EX2037