Variant Calling Pipeline

Primary source for lastest pipeline

- Broad / GATK Best practices
- Very complete but not 100%
  - Alternative Callers (consensus callers; Speed)
  - Stock Somatic one: missing indels
  - Annotation
    - Annotation is a difficult and rapidly changing area
    - Lots of stuff not shared (proprietary)

BP are evolving: WARNING

Trust, but verify.

If someone hands you a script and tells you "this runs the GATK Best Practices", start by asking what version of GATK it uses, when it was written, and what are the key steps that it includes. Both our software and our usage recommendations evolve in step with the rapid pace of technological and methodological innovation in the field of genomics, so what was Best Practice last year (let alone in 2010) may no longer be applicable. And if all the steps seem to be in accordance with our docs (same tools in the same order), you should still check every single argument in the commands. If anything is unfamiliar to you, figure out what it does. If you can't find it in the docs, ask us in the forum. It's one or two hours of your life that can save
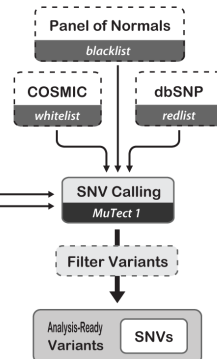
Overview
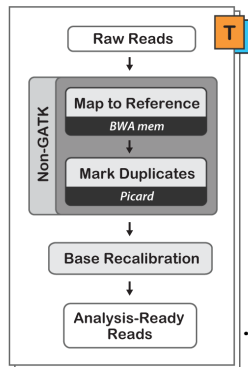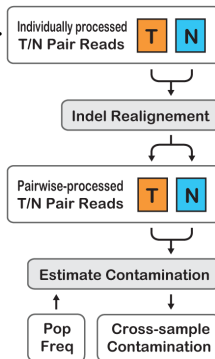
Best Practices for Somatic SNPs in Whole Genomes and Exomes

Be Explicit!

- ▶ Whenever and as much as possible want to be explicit in your pipelines:
    - ▶ Paths
    - ▶ Versions
    - ▶ Output
    - ▶ Version Stamps
- ▶ That goes for documentation of both pipeline and results (*Readability*)
- ▶ Implicitness is nice for development but can be painful in the long run

Annotation

- Annotation of mutations is