ONH Pipeline Outline

Kevin Stachelek, Jennifer Aparicio

10/19/2017

1. [Trimmomatic](http://www.usadellab.org/cms/?page=trimmomatic)
   * Remove adapters
   * Remove leading low quality or N bases
   * Remove trailing low quality or N bases
   * Scan the read with a n-base wide sliding window, cutting when the average quality per base drops below k
   * Drop reads below a given length
2. [BwaMem](http://bio-bwa.sourceforge.net/)
   * local alignment

# Picard

1. [SamFormatConverter](http://broadinstitute.github.io/picard/command-line-overview.html#SamFormatConverter)
   * Convert a BAM file to a SAM file, or SAM to BAM. Input and output formats are determined by file extension.
2. [SortSam](http://broadinstitute.github.io/picard/command-line-overview.html#SortSam)
   * Sorts a SAM or BAM file
3. [MarkDuplicates](http://broadinstitute.github.io/picard/command-line-overview.html#MarkDuplicates)
   * Identifies duplicate reads.
4. [AddOrReplaceReadGroups](http://broadinstitute.github.io/picard/command-line-overview.html#AddorReplaceReadGroups)
   * Replace read groups in a BAM file
5. [BuildBamIndex](http://broadinstitute.github.io/picard/command-line-overview.html#BuildBamIndex)
   * Generates a BAM index ".bai" file.
6. [Mosdepth](https://github.com/brentp/mosdepth)
   * fast BAM/CRAM depth calculation for WGS, exome, or targeted sequencing.

# GATK

1. [BaseRecalibrator](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_bqsr_BaseRecalibrator.php)
   * Detect systematic errors in base quality scores
2. [PrintReads](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_readutils_PrintReads.php)
   * Write out sequence read data (for filtering, merging, subsetting etc)
3. [VariantFiltration](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_filters_VariantFiltration.php)
   * Filter variant calls based on INFO and FORMAT annotations
4. [SelectVariants](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_variantutils_SelectVariants.php)
   * Select a subset of variants from a larger callset
5. [HaplotypeCaller](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php)
   * Call germline SNPs and indels via local re-assembly of haplotypes
6. [GenotypeGVCFs](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_variantutils_GenotypeGVCFs.php)
   * Perform joint genotyping on gVCF files produced by HaplotypeCaller
7. [VariantRecalibrator](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_variantrecalibration_VariantRecalibrator.php)
   * Build a recalibration model to score variant quality for filtering purposes
8. [ApplyRecalibration](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_variantrecalibration_ApplyRecalibration.php)
   * Apply a score cutoff to filter variants based on a recalibration table
9. [CalculateGenotypePosteriors](https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_variantutils_CalculateGenotypePosteriors.php)
   * Calculate genotype posterior likelihoods given panel data
10. [VariantAnnotator](https://software.broadinstitute.org/gatk/gatkdocs/3.6-0/org_broadinstitute_gatk_tools_walkers_annotator_VariantAnnotator.php)
    * Annotate variant calls with context information
11. [TableAnnovar](http://annovar.openbioinformatics.org/en/latest/user-guide/startup/)
    * takes an input variant file (such as a VCF file) and generate a tab-delimited output file with many columns, each representing one set of annotations. Additionally, if the input is a VCF file, the program also generates a new output VCF file with the INFO field filled with annotation information.
12. [VcfAnno](https://github.com/brentp/vcfanno)
    * vcfanno allows you to quickly annotate your VCF with any number of INFO fields from any number of VCFs or BED files. I am using it to annotate
      1. gnomad minor allele frequency
      2. dbsnp ids
13. [Genmod](https://github.com/moonso/genmod)
    * GENMOD is a simple to use command line tool for annotating and analyzing genomic variations in the VCF file format. GENMOD can annotate genetic patterns of inheritance in vcf:s with single or multiple families of arbitrary size.