Basic SNP calling procedure:

- Map reads to reference
- Remove/mark PCR duplicates (artefacts from library prep)
- Use probability model to distinguish SNPs (or other variants) from mismatches due to sequencing errors etc…
- Post processing of SNPs using flags in Variant Call File (VCF)


Re-align within multi-read context

- Uses information from known SNPs/indels

  (dbSNP, 1000 Genomes)

- Uses information from other reads

- Smith-Waterman exhaustive alignment on select reads

Consider:

- Coverage at position
- Number independent reads supporting variant
- Observed allele fraction vs expected (somatic,germline)
- Strand bias
- Base qualities at variant position
- Mapping qualities of reads supporting variant
- Variant position within reads (near ends or at centre)

SA PATHOLOGY