bioinformatics.ca

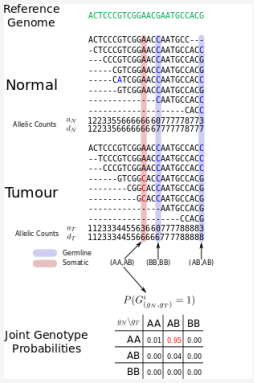
Canadian Bioinformatics Workshops

www.bioinformatics.ca

Lab Module 6
Somatic Mutations

Sohrab Shah, Fong Chun Chan
Bioinformatics for Cancer Genomics
May 25-29, 2015

bioinformatics.ca



BC Cancer Agency
CARE & RESEARCH
An agency of the Provincial Health Services Authority

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Many slides adapted from Gavin Ha, Andrew Roth, and Andrew McPherson

Lab Module 6: Learning Objectives

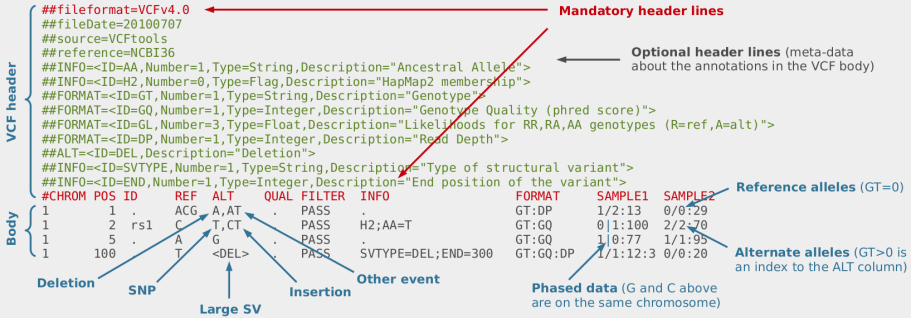
1. Apply methods to identify SNVs from sequencing data
2. Understand the Variant Call Format (VCF)
3. Visualize SNV and understand some common technical artifacts

Calling Variants

- Strelka (Saunders et al., Bioinformatics, 2012)
  - Realignment around indels
- MutationSeq (Ding et al., Bioinformatics, 2011)
  - Supervised learning based on ~1000 validated events

VCF Format

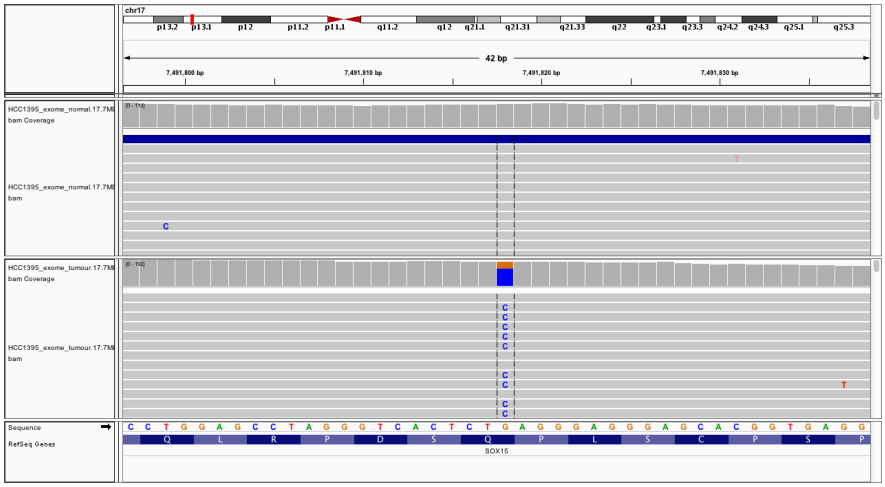
- Tab separated
- 1 line per variant
- Structured fields as detailed in header



Jose Blanca & Joaquin Cañizares
http://bioinf.comav.upv.es/courses/sequence_analysis/snp_calling.html

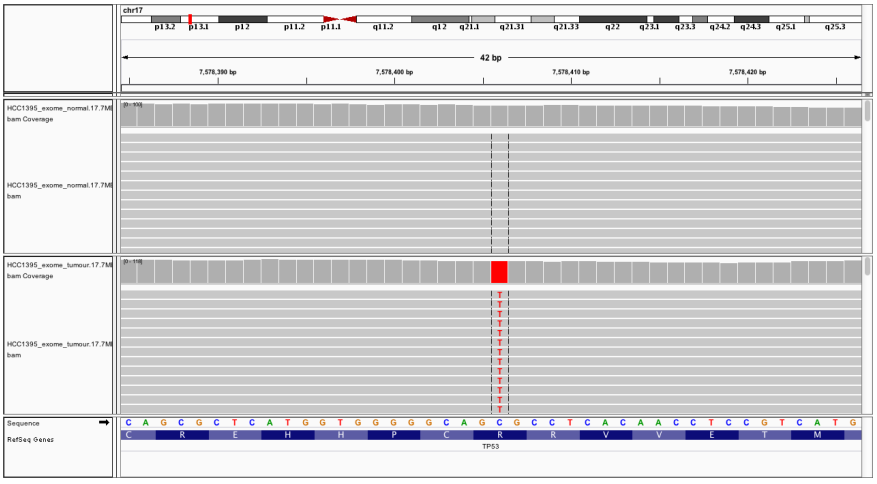
Browsing Somatic Mutations In IGV

- A somatic mutation (17:7491818) predicted by both Strelka and MutationSeq



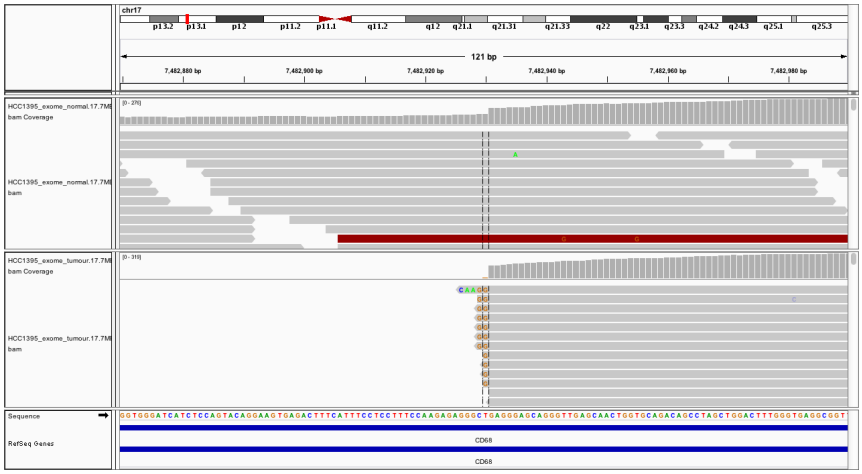
Browsing Somatic Mutations In IGV

- A somatic mutation (17:7578406) predicted only by Strelka (MutationSeq PR = 0.68)



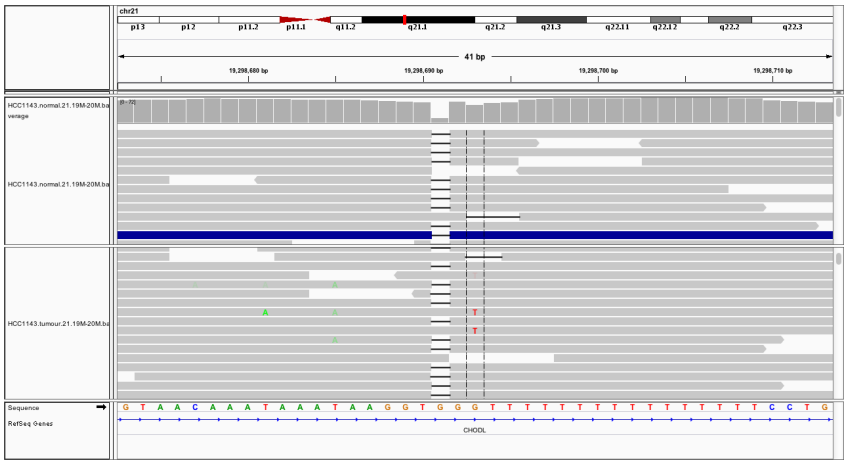
Browsing Somatic Mutations In IGV

- Likely false positive (17:7482929) identified by Strelka, but occurs near the end of reads (MutationSeq PR = 0.5)



Browsing Somatic Mutations In IGV

- Alignment issues with result in false positive; MutationSeq identified this as a "INDL" region (nearby by indels)



Summary Of Somatic Mutation Calling

- Joint analysis of both tumour and normal genomes removes germline false positives effectively
- False positives can still arise:
  - Sequencing artifacts
  - Alignment issues
- Post-prediction filter can be further applied to reduce false positives
- When possible, manually inspect predicted mutations in IGV

Summary Of Somatic Mutation Calling

Many different software for calling somatic mutations:

- Strelka (https://sites.google.com/site/strelkasomaticvariantcaller/)
- MutationSeq (http://compbio.bccrc.ca/software/)
- Mutect (http://www.broadinstitute.org/cancer/cga/mutect)
- EBCall (https://github.com/friend1ws/EBCall)
- …

Tools designed for normal tissue work poorly for cancer genomes:

- GATK
- Samtools

No matter what software you use, all predictions are just predictions. Without validation on an orthogonal platform they should never be considered true.