Precision medicine: NGS variant analysis and interpretation for translational research

Exercise 2: Variant annotation with VEP

Fátima Al-Shahrour ● Javier Perales ● Elena Piñeiro

November 15, 2017

cnio Centro Nacional de Investigaciones Oncológicas

INSTITUTO GULBENKIAN DE CIÊNCIA

Study case

Tumor type: hepatic met – colon adenocarcinoma

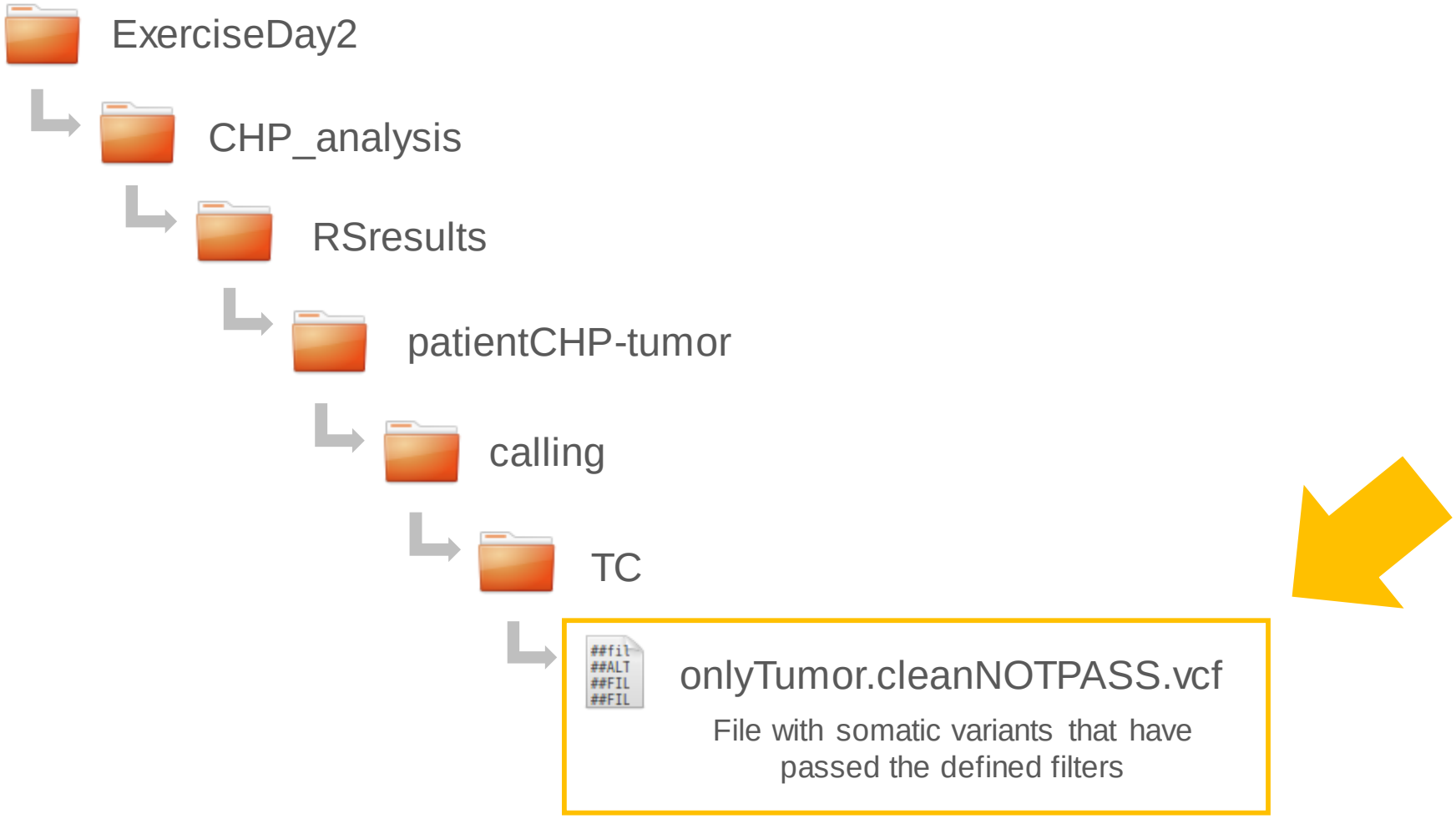
Sequencing: Ion Torrent technology

Panel: Ion Ampliseq Cancer Hotspot Panel v2

Tumor – control paired sample

File with somatic variants: variants detected in tumor sample but not in the corresponding control

Path to the input file



Execution of VEP through the web page

- We want to obtain the annotations using the ensembl transcript set
- We want in the output the following annotations:
  - HUGO gene symbol
  - The HGVS identifiers for cDNA and protein
  - The Global Minor Allele Frequency of 1000 genomes project
  - gnomAD frequencies
  - SIFT and PolyPhen prediction and score
  - Condel prediction and score
- You can choose any other parameters to explore the results

HINTS: Remember to use the same assembly used in the variant detection
http://www.ensembl.org/info/docs/tools/vep/online/input.html (further info)

Answer the following questions

How many variants were in the input file?

How many of them are not known in the database?

How many genes and transcripts are affected by the variants?

Is there any regulatory region overlapping some variant?

Which is the most represented consequence category?

Which is the most represented coding sequence?

Answer the following questions

How many variants fall in a coding region in some gene?

What does the HGVS identifiers mean in each case?

Does the prediction tool agree in the prediction of functional impact?

Is there any clear polymorphism within the data?

… and explore the results

Download the file

Save the file in vcf format, copy it in the CHP_analysis folder and rename it (e.g. patientCHP.onlyTumor.cleanNOTPASS.vep.vcf)

Check that the following annotations have been added to the INFO field:

- Consequence
- Existing_variation
- Feature
- PolyPhen
- Condel
- SIFT
- SYMBOL
- Protein_position
- Amino_acids
- HGVSc
- HGVSp
- AF
- CDS_position
- Allele
- Gene
- Feature_type
- cDNA_position
- Codons
- gnomAD_AF
- gnomAD_NFE_AF