Disease Variant Prioritization using IVA: Integrative Variant Analysis tool

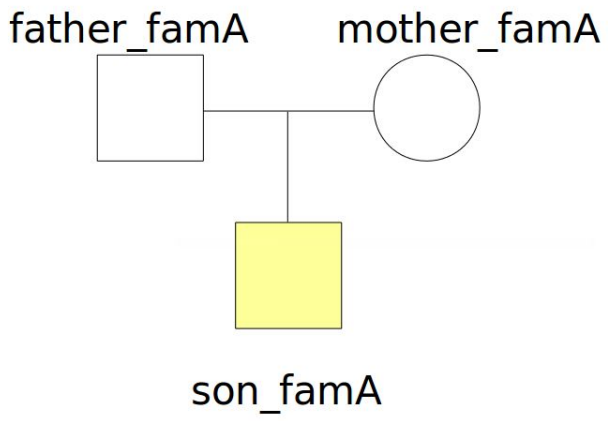
Url IVA: http://iva-courses.clinbioinfosspa.es/
OpenCB: http://docs.opencb.org/

Users list: http://www.clinbioinfosspa.es/files/courses/gatk19/users_list.xlsx

Exercises: Metabolic Diseases

Study A (guided):

Study: metabolic_diseases
Samples: son_famA, mother_famA, father_famA
Phenotype: Cystic Fibrosis



Candidate variant: 7:117171029 G>A

Study B (self-paced):

Study: metabolic_diseases
Samples: daughter_famB, mother_famB, father_famB

A girl has the following symptoms: hepatic steatosis, hepatic hemangioma, hypothyroidism and arrhythmia.
Both parents are healthy. After considering clinical and familiar history, a fructose intolerance is suspected.

1) Draw the family pedigree.
   Write down the suspected inheritance mechanisms, given the information you have on the family, and the possible individuals' genotypes.

2) Search for the phenotypic terms in HPO (Human Phenotype Ontology), of the subject's symptoms.
http://compbio.charite.de/hpoweb/showterm?id=HP:0000118#id=HP_0000118

3) What disorder is suspected to have the affected individual? Search for it on OMIM database (Online Mendelian Inheritance in Man). http://omim.org/
What other names has this disorder?
What is its OMIM code?
This disorder can be caused by compound heterozygous mutations?
Do they know the driven gene?

4) Search for the disease on ORPHANET database.
https://www.orpha.net/consor/cgi-bin/index.php?lng=ES
What is its Orpha code?
What is its ICD-10 code (International Statistical Classification of Diseases - 10)?
What is its prevalence?

5) Create an analysis with IVA tool to study this family. (Studies > Metabolic_diseases)
- Affected daughter: daughter_famB
- Healthy mother: mother_famB
- Healthy father: father_famB

Prioritize the variants in this family. Remember that filters are on the left panel.
How many variants are in the family?
How many variants are in PPT1 gene?
How many variants are in X chromosome? And between positions 38330500 and 38340000 from chromosome 21?

In the Population Frequency filter, identify the following populations:

- European population from 1000 genomes project.
- Spanish population from 1000 genomes project.
- Populations from ExAC consortium.
- Healthy spanish population from Medical Genome Project.

How many variants have a population frequency lower than 0.02.

What is the 1000G IBS frequency of the first variant? And MGP frequency?

From the variants left after applying population frequency filter, search for the ones that are LoF (Loss of Function) variants. What terms are selected after choosing LoF terms on the SO terms filter?

Each of the SO terms are detailed in http://www.sequenceontology.org/
You can find the ranking of their impact here:
https://www.ensembl.org/info/genome/variation/predicted_data.html

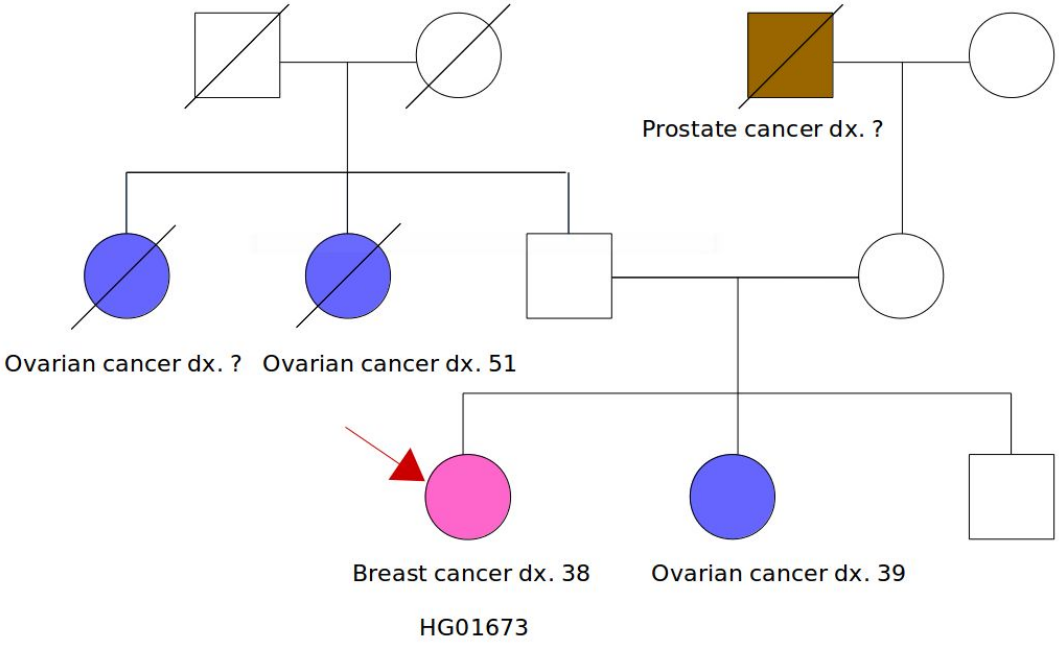
Do any of the resulting variants has ClinVar information? What does this mean?
What is its CADD value? Considering the phenotype of the subject, would you consider this variant as a good candidate?

Apply Genotype filter, so that the individuals' genotypes agree with the suspected inheritance mode. Search for all the variants in coding region with a Scaled CADD value higher than 15. Remember to apply population frequency filter (in this case, try lower than 0.02).

Considering the results, is there any candidate variant?

Exercise (bonus): Hereditary Cancer

Study: hereditary_cancer
Sample: HG01673
Phenotype: Breast Cancer (HP:0003002)



Create an individual analysis using this sample and perform an analysis and interpretation in order to obtain candidate variants explaining the patient's phenotype (marked with a red arrow).

Databases

ORPHANET: The portal for rare diseases and orphan drugs
https://www.orpha.net/consor/cgi-bin/index.php

OMIM: Online Mendelian Inheritance in Man
http://omim.org/

HPO: Human Phenotype Ontology
https://hpo.jax.org/app/

ClinVar: Database for Clinical Variation
https://www.ncbi.nlm.nih.gov/clinvar/

COSMIC: Catalogue Of Somatic Mutations In Cancer
https://cancer.sanger.ac.uk/cosmic

1000Genomes
http://www.internationalgenome.org/

dbSNP: Short Genetic Variations
https://www.ncbi.nlm.nih.gov/projects/SNP/

Readings

1: Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet. 2017 Oct;18(10):599-612. doi: 10.1038/nrg.2017.52. Epub 2017 Aug 14. Review. PubMed PMID: 28804138; PubMed Central PMCID: PMC5935497.

2: Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PubMed PMID: 25741868; PubMed Central PMCID: PMC4544753.

3: Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy

statement of the American College of Medical Genetics and Genomics. Genet Med. 2017 Feb;19(2):249-255. doi: 10.1038/gim.2016.190. Epub 2016 Nov 17. Erratum in: Genet Med. 2017 Apr;19(4):484. PubMed PMID: 27854360.
4: ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet Med. 2015 Jan;17(1):68-9. doi: 10.1038/gim.2014.151. Epub 2014 Nov 13. PubMed PMID: 25356965.

5: Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81. doi: 10.1038/nprot.2009.86. Epub 2009 Jun 25. PubMed PMID: 19561590.

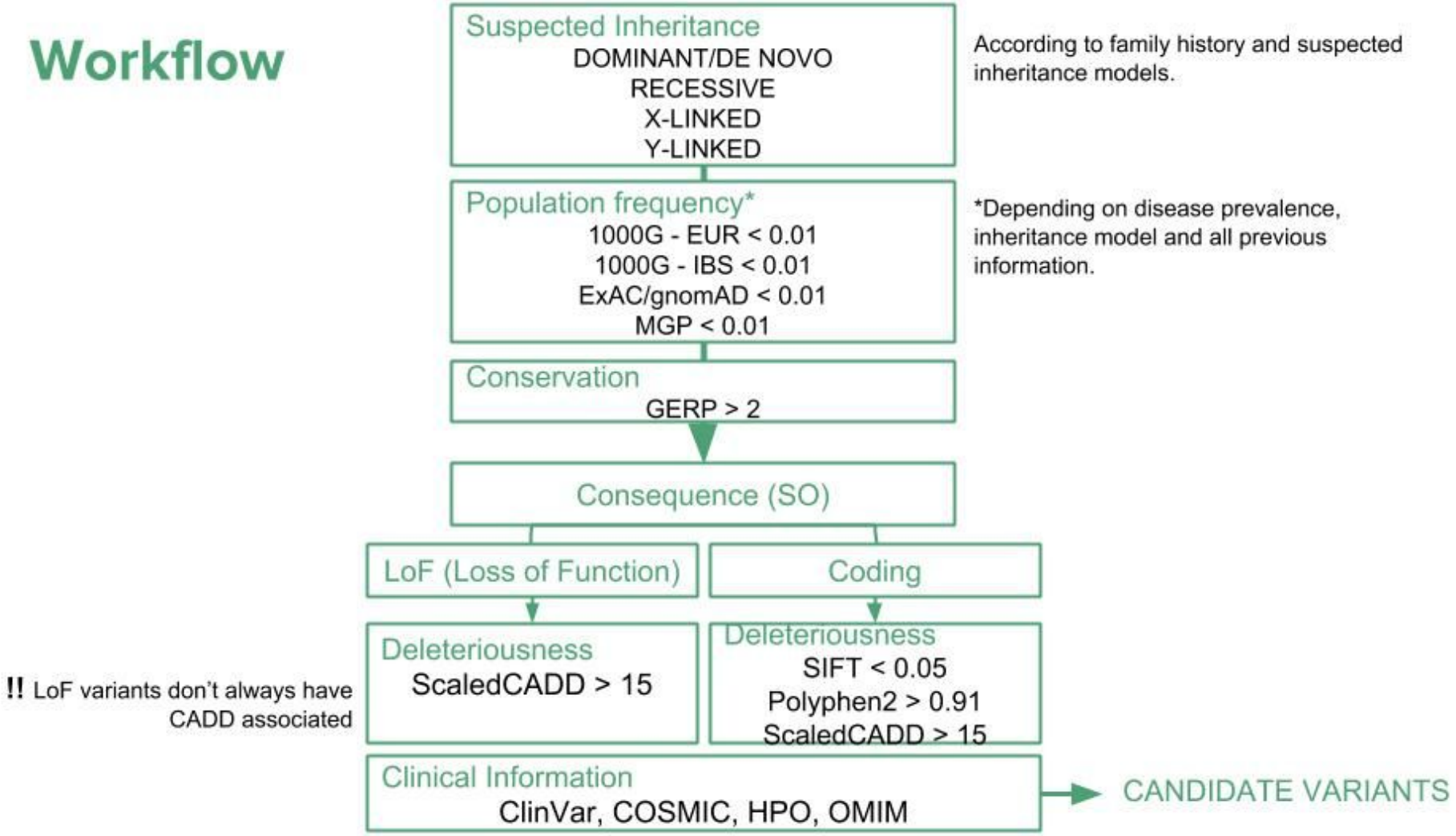
6: Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013 Jan;Chapter 7:Unit7.20. doi: 10.1002/0471142905.hg0720s76. PubMed PMID: 23315928; PubMed Central PMCID: PMC4480630.

7: Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9. doi: 10.1038/nmeth0410-248. PubMed PMID: 20354512; PubMed Central PMCID: PMC2855889.

8: Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Mar;46(3):310-5. doi: 10.1038/ng.2892. Epub 2014 Feb 2. PubMed PMID: 24487276; PubMed Central PMCID: PMC3992975.

9: Alemán A, Garcia-Garcia F, Medina I, Dopazo J. A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W83-7. doi: 10.1093/nar/gku472. Epub 2014 May 26. PubMed PMID: 24861626; PubMed Central PMCID: PMC4086136.

Workflow


Suspected Inheritance
DOMINANT/DE NOVO
RECESSIVE
X-LINKED
Y-LINKED
According to family history and suspected inheritance models.
Population frequency*
1000G - EUR < 0.01
1000G - IBS < 0.01
ExAC/gnomAD < 0.01
MGP < 0.01
*Depending on disease prevalence, inheritance model and all previous information.
Conservation
GERP > 2
Consequence (SO)
LoF (Loss of Function)
Coding
Deleteriousness
ScaledCADD > 15
Deleteriousness
SIFT < 0.05
Polyphen2 > 0.91
ScaledCADD > 15
!! LoF variants don't always have CADD associated
Clinical Information
ClinVar, COSMIC, HPO, OMIM
CANDIDATE VARIANTS