VarFish Training Handout

Course September 5, 2022
Current Version: September 2, 2022

Nadja Ehmke[1], Manuel Holtgrewe[2], Peter Krawitz[3]

[1]Humangenetik, Charité - Universitätsmedizin Berlin
[2]Core Unit Bioinformatics (CUBI) - Berlin Institute of Health @Charité
[3]Institut für Genomische Statistik und Bioinformatik, Universitätsklinikum Bonn

Contact: manuel.holtgrewe@bih-charite.de

1 Introduction

1.1 VarFish

VarFish is a web-based software for the analysis of **genetic variant data** for **rare disease genetics**. VarFish allows genetics specialists (physicians, *Fachhumangenetiker (de)*, genetic councelors, or similarly qualified) to filter and prioritize variants from exome, genome, or panel sequencing assays.

VarFish was originally developed by Core Unit Bioinformatics (CUBI) at Berlin Institute of Health @Charité in close collaboration with the Charité Institute of Medical Genetics with the key contributors being (ordered alphabetically) Dieter Beule, Nadja Ehmke, Manuel Holtgrewe, Oliver Stolpe. If you want to cite VarFish, please use:

> Manuel Holtgrewe, Oliver Stolpe, Mikko Nieminen, Stefan Mundlos, Alexej Knaus, Uwe Kornak, Dominik Seelow, Lara Segebrecht, Malte Spielmann, Björn Fischer-Zirnsak, Felix Boschann, Ute Scholl, Nadja Ehmke, Dieter Beule, VarFish: comprehensive DNA variant analysis for diagnostics and research, Nucleic Acids Research, Volume 48, Issue W1, 02 July 2020, Pages W162–W169, https://doi.org/10.1093/nar/gkaa241

VarFish development is now continued as open source software with a group of developers based in Berlin and Bonn with contributions from other sites such as Aachen and Göttingen. Organizations using VarFish include (ordered alphabetically by city) Uniklinik RWTH Aachen, Charité Berlin, Labor Berlin, Universitätsklinikum Bonn, Universitätsmedizin Göttingen, and Universitätsklinikum Schleswig-Holstein.

1.2 This Course

This course is aimed at participants who have at least basic experience with the analysis of NGS variant data. The course starts with a walkthrough of solving a simple case with VarFish. You can find a version in text form in Section 2.

When being done online/on-site, the participants are given this training material, an account for the VarFish installation at https://varfish-ext.cubi.bihealth.org, and access to the video before the actual course. Participants must ensure that their login works and are recommended to solve the first case with the walkthrough information. We provide a copy of the example cases for each user for their training.

The course itself starts by an expert VarFish user going through the walkthrough and explaining useful VarFish users. Participants are invited to ask any questions they have. The course then continues with participants solving the remaining cases on their own while the trainers (expert VarFish users) are available to offer any help. Of course, users can also pair/team up to solve cases together. The course then closes with a discussion round where participants can reflect on their experience and learnings and discuss any pitfalls that they found.

1.3 More Information

VarFish contains an online manual that can be found at the following location:

- https://varfish-server.readthedocs.io/en/latest/

Also, you can access this by clicking the “Manual” link on the top right of the VarFish web application, see Figure 1.

Figure 1. Manual link in VarFish web application.

1.4 Getting VarFish for your Own Data

Unlike services such as Limbus varvis or Illumina TruSight Software Suite, VarFish is not running in a central instance in the cloud but it can be installed locally in your organization. Such an installation will require an appropriate (virtual) Linux server and knowledge about operating such server systems.

If you want to test VarFish on your own data, we can create a space with controlled access on the server https://varfish-ext.cubi.bihealth.org for you and other users from your organization so you can use VarFish with your own data without having to set up your own local instance. If you are interested to do so, please send us an email to cubi-helpdesk@bih-charite.de.

1.5 Getting in Touch

VarFish developers can be reached at cubi-helpdesk@bih-charite.de.

You can subscribe to the mailing list for VarFish end users (using VarFish to solve cases) here:

- https://mailman.charite.de/mailman/listinfo/varfish-users

There also is a mailing list for VarFish operators (those who are responsible for running a local VarFish server):

- https://mailman.charite.de/mailman/listinfo/varfish-operators

1.6 IGV Installation and Usage

IGV (Integrative Genome Viewer) is used for display of BAM (binary alignment and mapping) files. This section gives a short overview of the installation of IGV and its use in the context of this course. IGV can be downloaded from the following website:

- https://software.broadinstitute.org/software/igv/download

Download the appropriate version for your operating system that has “Java included”, e.g., “IGV for Windows (Java included)”.

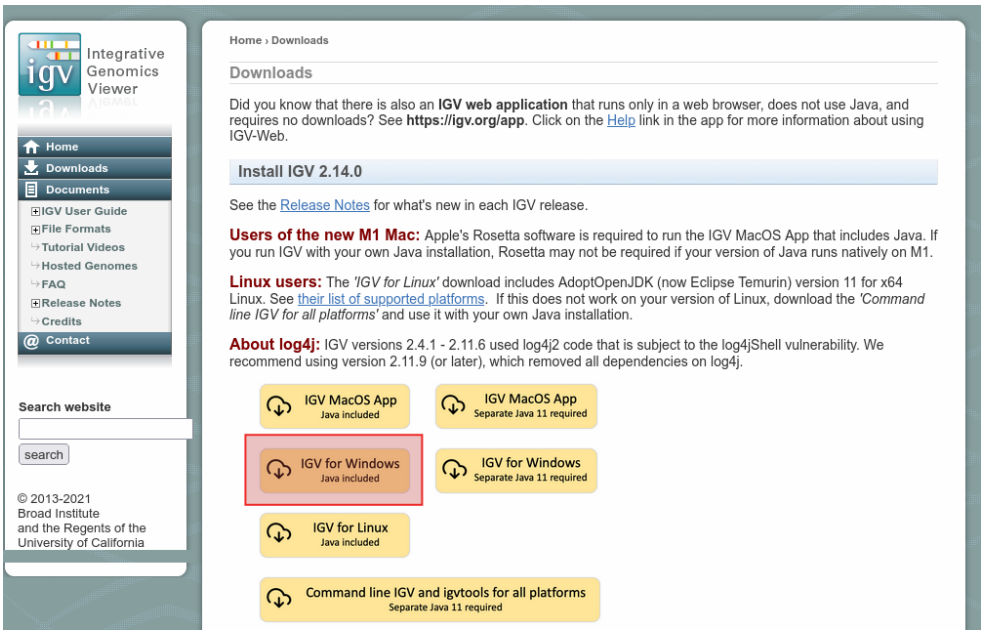


Figure 2. IGV Download site with “IGV for Windows (Java included)” highlighted.

On Windows, after downloading the installer, accept the license agreement, click “next”, and “install” to the default directory. You will get an icon “IGV_2.14.0” (or the installed version, respectively). Start the application with the icon and you will see the IGV genome browser main window.

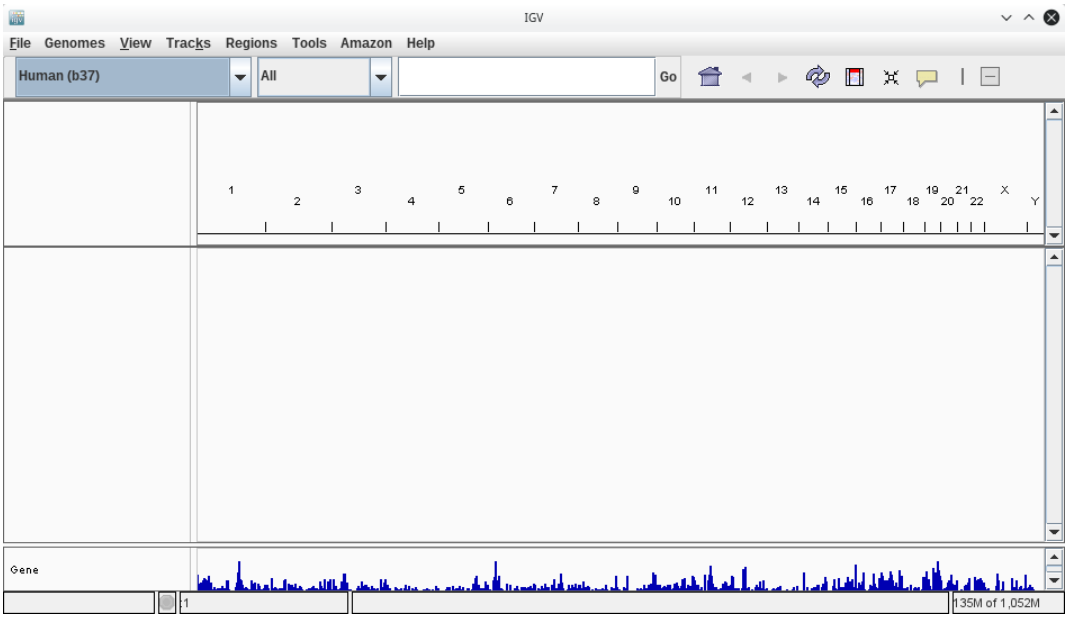


Figure 3. IGV genome browser main window.

In the case that IGV complains that it could not download the genome data on startup, you will have to properly configure the proxy. This will be the case in many hospital networks. The actual setting depends on your organization. You can open the “Preferences” window by using the “View > Preferences” menu of the IGV main window. Go to the “Proxy” tab as shown below. You will need to check “Use proxy” and configure the appropriate proxy host and port (you can get this information from your organization’s IT department).

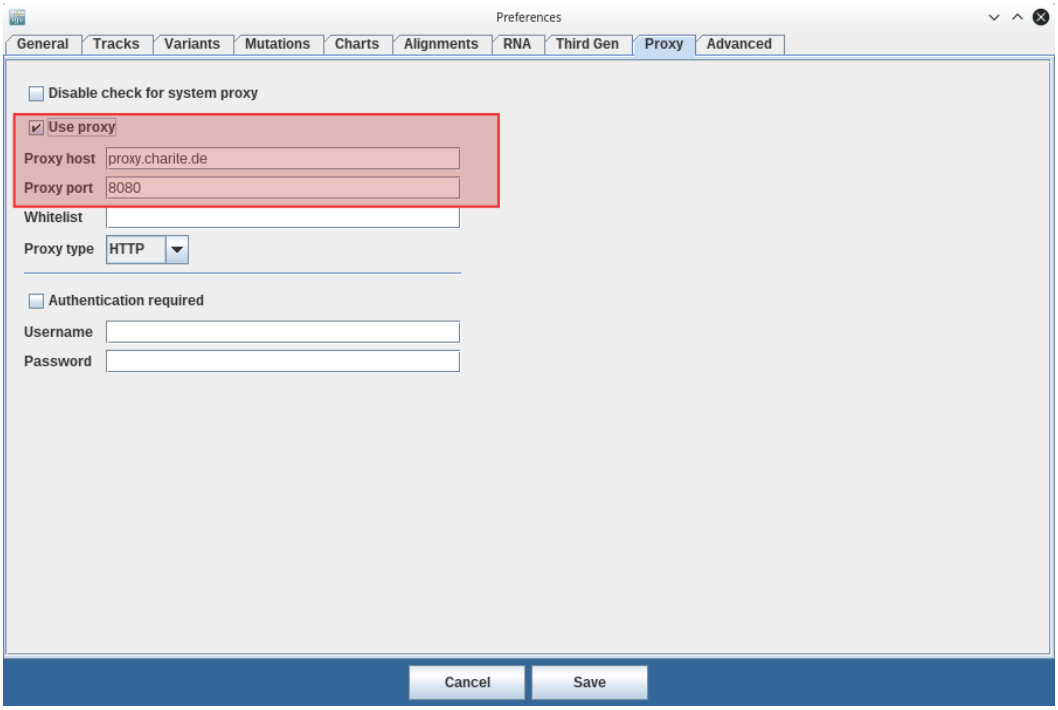


Figure 3. IGV proxy configuration setting in IGV “Preferences”.

Make sure that you click “Save” so the setting is actually applied. Also make sure that in the “Advanced” tab of Preferences, you have “Enable port” activated and the “Port number” set to 60151.

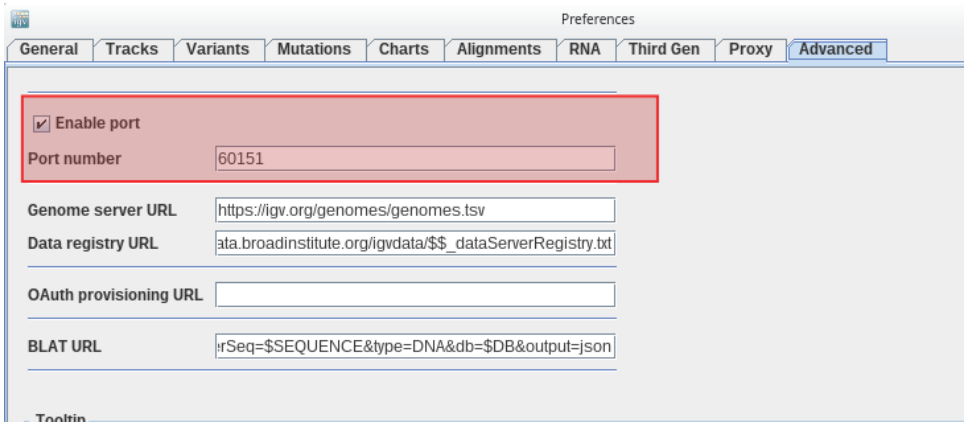


Figure 4. IGV advanced configuration to configure “remote control” via port.

To check that IGV has been setup correctly, try to open one of the session files that we provide. You can do so using the “File > Open Session” menu or by drag and drop of the file to the IGV window. Once the session has loaded, copy “15:48,936,524-48,937,521” into the location field to jump to the second exon of FBN1 and check that the reads of the cases load. Your browser session should look similar to the following figure.

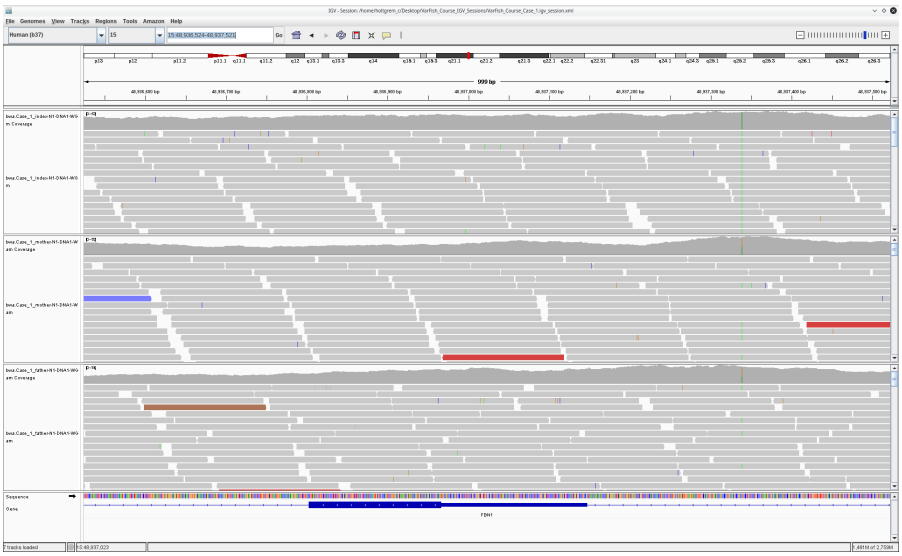


Figure 5. IGV session displaying the read alignments of the trio of case 1 in the first exon of gene FBN1.

2 VarFish Walkthrough

This section provides a very brief walkthrough of using VarFish using the first example case from Section 3. The walkthrough will demonstrate the general approach of solving rare disease cases with VarFish. Note that this is not a comprehensive description of all VarFish features but it demonstrates a large number of useful features that VarFish provides.

First, open the session XML file from the ZIP file we sent you via email for case 1 in IGV. You can also find a copy here: https://file-public.cubi.bihealth.org/transient/varfish-course/. The IGV window will initially look as follows

Figure 6. IGV session in initial state.

You can enter a gene name and IGV will load the read alignments at the given position. For “TGDS”, this will look as follows.

Figure 7. IGV session of case 1 for gene “TGDS”.

Next, open the VarFish web app at https://varfish-ext.cubi.bihealth.org. Here, login with the credentials that we sent to you via email.

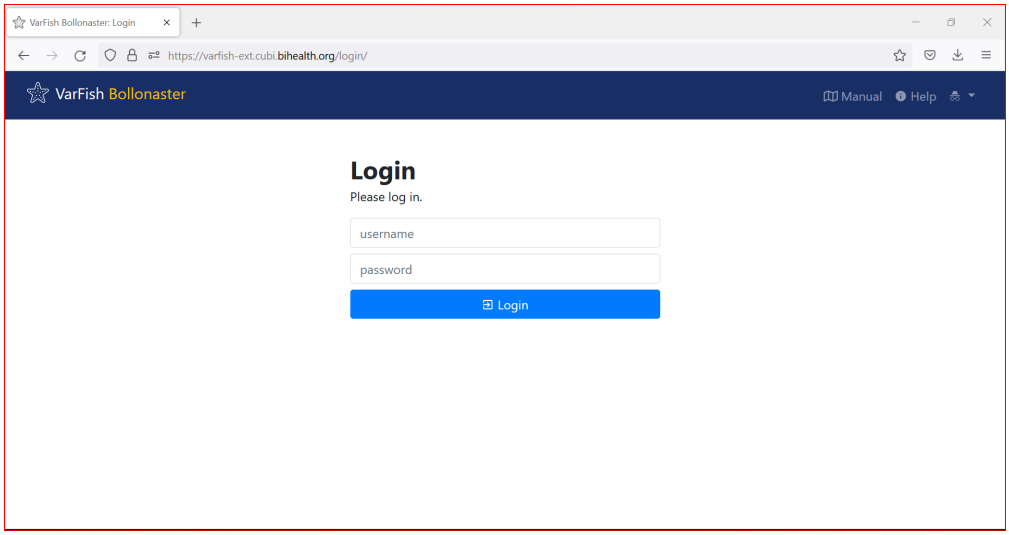


Figure 8. VarFish login screen.

When using a small screen, e.g., as on a laptop, you might want to use the “Zoom out” option of your browser to decrease the font size so more information fits on your screen (e.g., Ctrl+”-” on Windows). Initially, you will see the project overview that will look similar to the following. You will only have access to your own project that has your email address in its name.

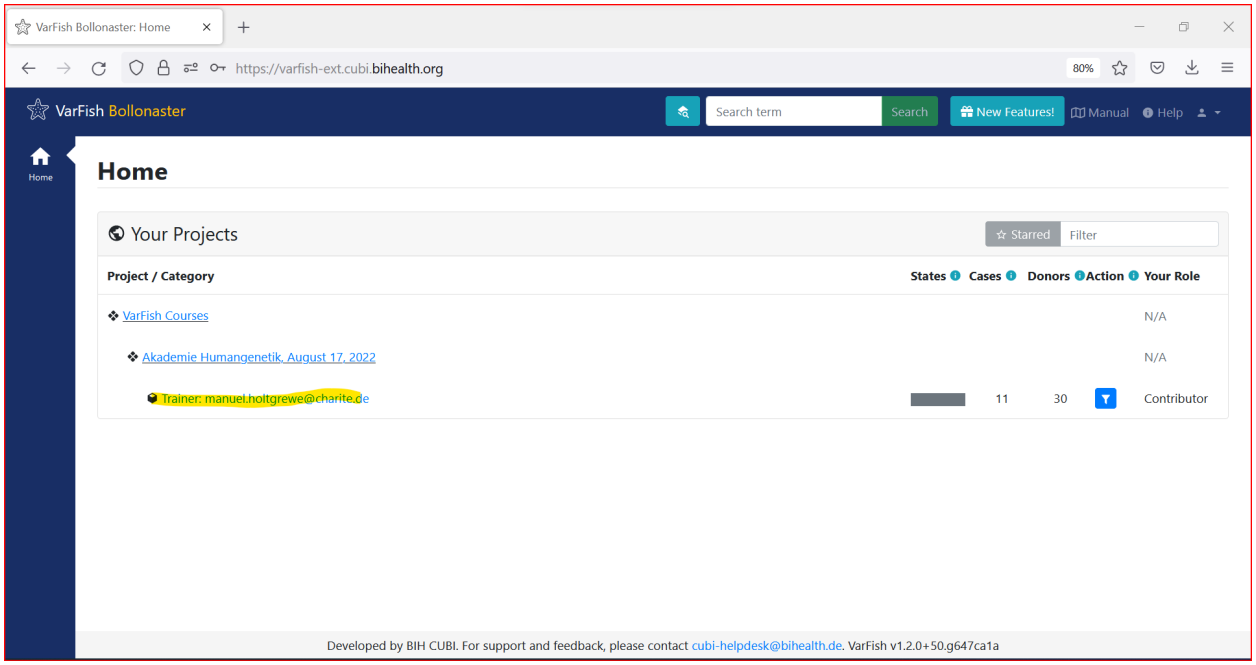


Figure 9. Project list.

Above, I see that my project has 11 cases with overall 30 donors. Before the workshop, we will only upload one case to your project. During and after the workshop, we will provide 10 cases for you to solve. Next, click on the project title and you will see the project overview with the most recent 5 cases. Click “See list of all cases” to continue to the full case list.

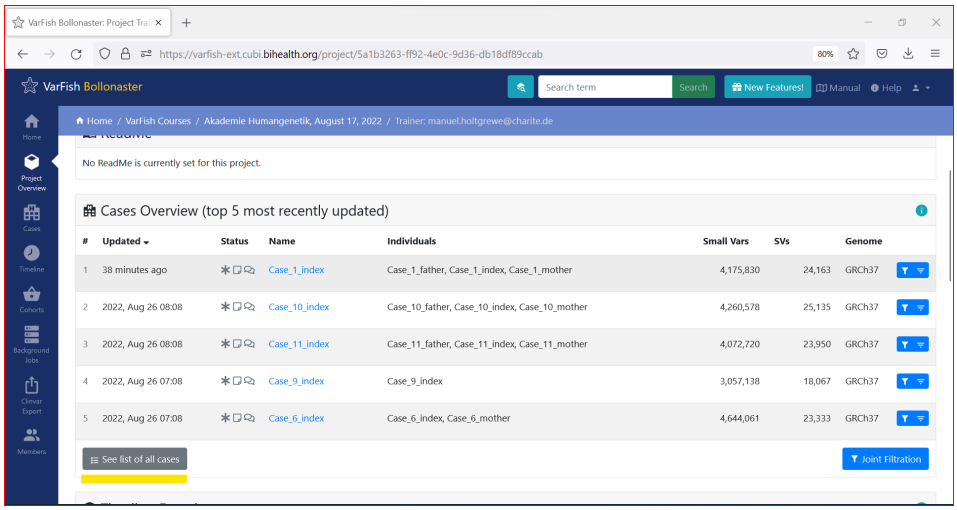


Figure 10. Project overview.

Click the name of the case (here Case_1_index, described in Section 3.1.1) to get to the case overview page.

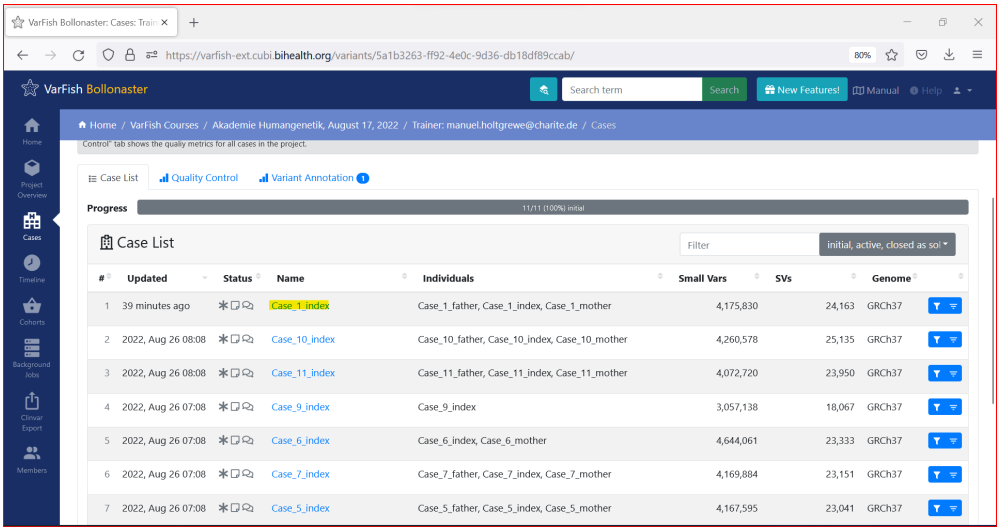


Figure 11. List of all cases in the project.

On the case detail screen, you can inspect various meta and quality control data for your case that will not be explained in detail in the walk-through. Next, click on the “Filter Variants” button to go to the small variant filtration screen.

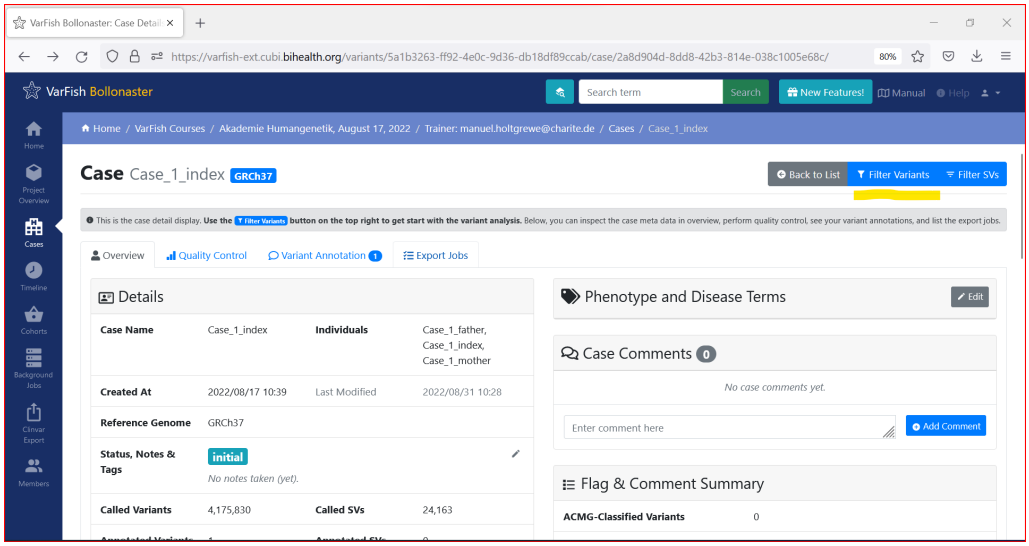


Figure 12. Case overview for Case_1_index.

The variant filtration screen (Figure 13) is quite complex when you see it the first but don't despair. You can use this screen to configure criterias that you want to filter your variants for. First of all, the top row allows you to quickly apply sensible filter presets (yellow markings in Figure 13). You can start out with overall presets, such as configuring filtration for *de novo* variants, applying dominant/recessive filter strategies, etc.

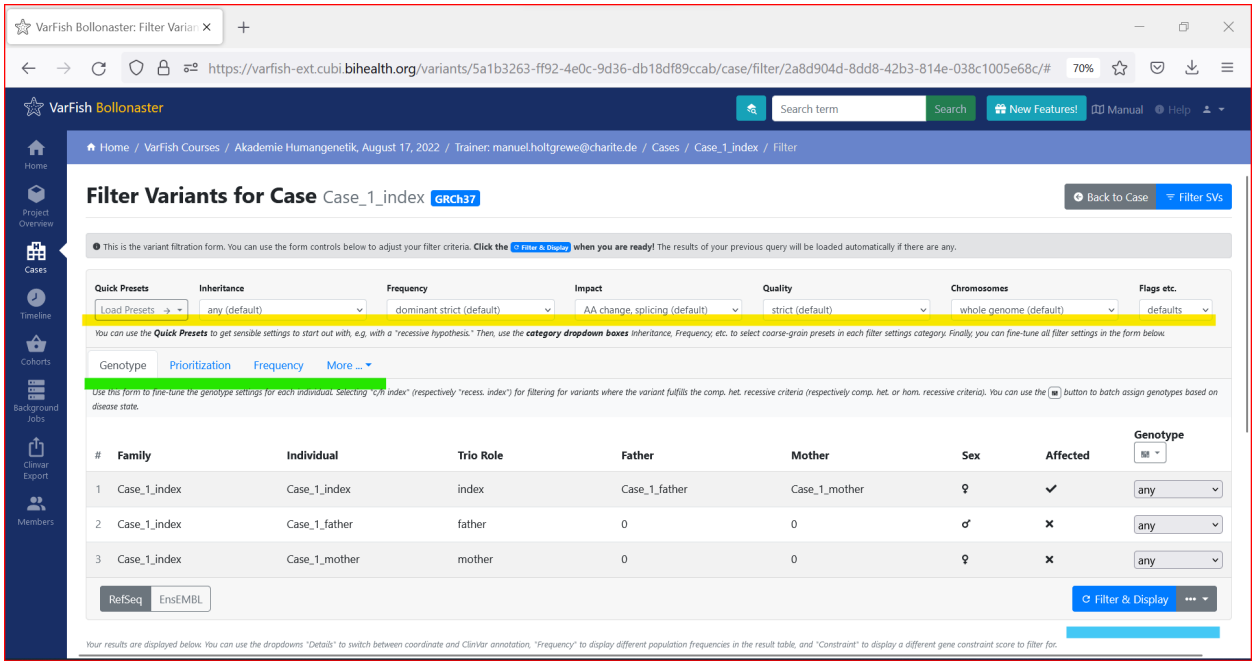


Figure 13. Variant filtration screen.

Selecting a such a quick preset will update the other preset categories in this row: genotype pattern compatible with modes of inheritance, frequency filter settings, molecular impact of variants, quality of variants, possible restriction to certain loci or genes, and requiring certain

user flags. You can view the different sections of the filter configuration form by using the tabs (marked as green in Figure 13). Finally, you can run the variant filtration by clicking on the “Filter & Display” button (marked as blue in Figure 13).

For case 1, let us first try a *de novo* preset. Click “Quick Presest => de novo”. This will apply appropriate configuration in terms of frequency etc. Then, click “Filter & Display”. VarFish will now start to filter the Variants.

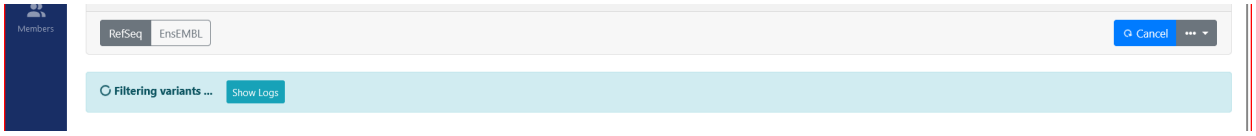


Figure 14. Variant filtration indicator.

Finally, the results will be displayed at the bottom as shown in Figure 15. The result table shows information on user flags/comments, presence in dbSNP and Clinvar, chromosomal position of the variant as well as genome reference and alternative base. It displays the frequency and number of homozygous in gnomAD exomes as well as the gnomAD pLI score of the corresponding gene. The row goes on with the gene name and a little “doctor” icon indicating whether the gene is in the ACMD incidental findings list or not. If disease gene association is in Human Phenotype Ontology or OMIM is available then the “DG” column will contain a red light bulb and the annotated mode of inheritance is shown (here “AD” for gene NSD1). The row further displays the impact on the protein (or transcript if outside of coding regions), as well as the genotype in each individual of the case.

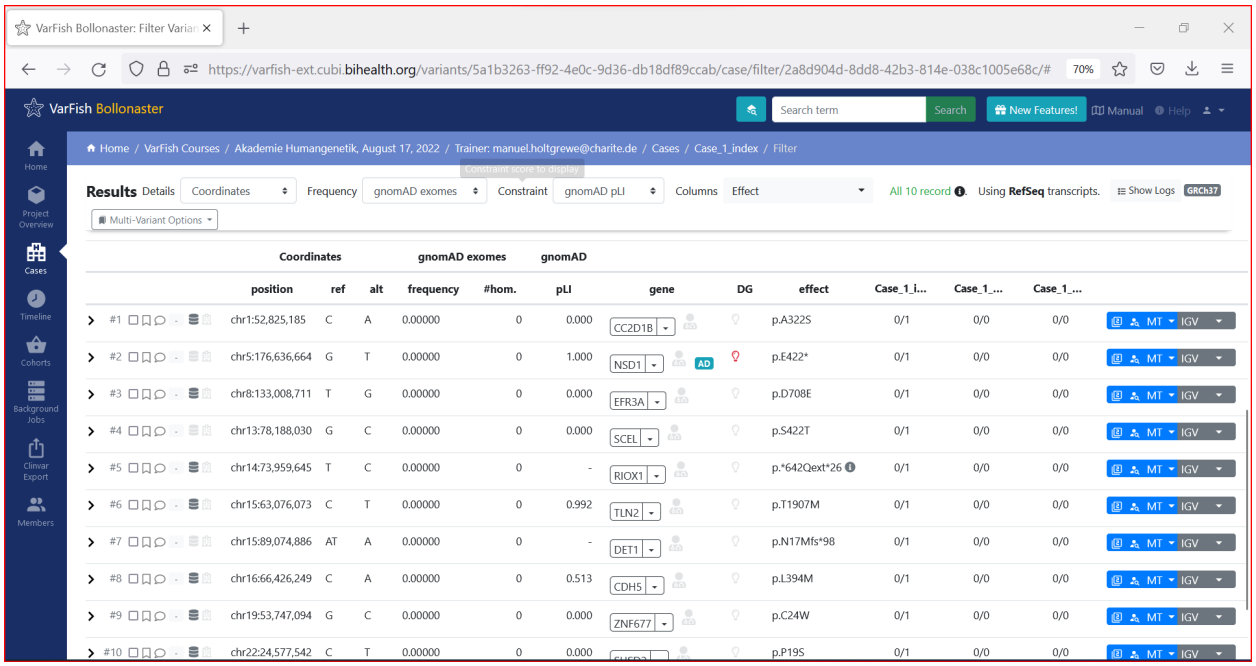


Figure 15. Filtration results.

At the end of the row, the “2” button allows to look for second hits of the same gene in the same case, the little “search” icon allows to search for other occurrences of the variant in cases that you have access to, and the “MT” button allows to query MutationTaster for the variant. The “IGV” button allows you to jump to the variant’s coordinates in IGV (if you have IGV open). The little arrows next to “IGV” and the gene name provide access to further databases related to this gene and this variant, e.g., gnomAD, or the UCSC genome browser.

Clicking on the little “>” sign on the left of a row shows further variant details. For example, the row for the “NSD1” gene will display the following.

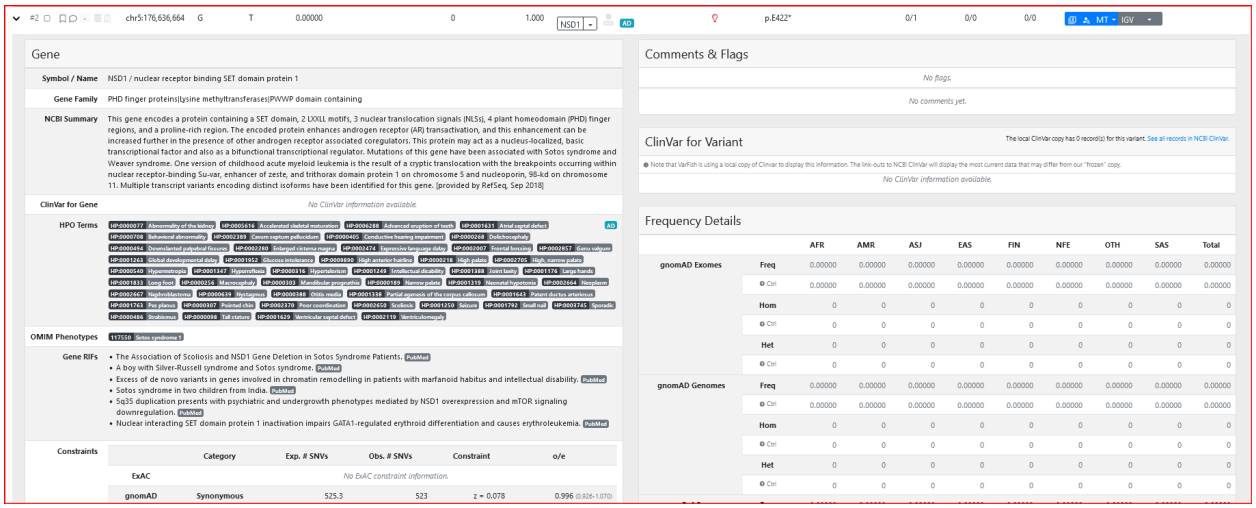


Figure 16. Detail display of NSD1 variant in case 1.

We can see that the gene defects in NSD1 cause Sotos syndrome 1, the variant is not present in ExAC, gnomAD, etc. The variant has a phred-scaled CADD Score of 40 and is highly conserved on the protein level in 100 vertebrates in UCSC genome browser. As Sotos syndrome 1 matches the phenotype of the case description, this is a good candidate for solving this case (and actually the variant that we spiked into it).

You can click the little bookmark icon mark the variant as pathogenic and thus colour it red (or as “unclear”/yellow or “benign/artifact”/green). This mark will be stored for this case and will also be available for other users that have access to the case. You can also leave text comments that will be displayed in the variant details and on the case summary page. Note that all participants in the course get their own project and do not get access to each other's cases. If you click the little “-” right of the speech bubble, you can use ACMG criteria to grade the variants. Figure 17 shows the according user interface dialogues for this.

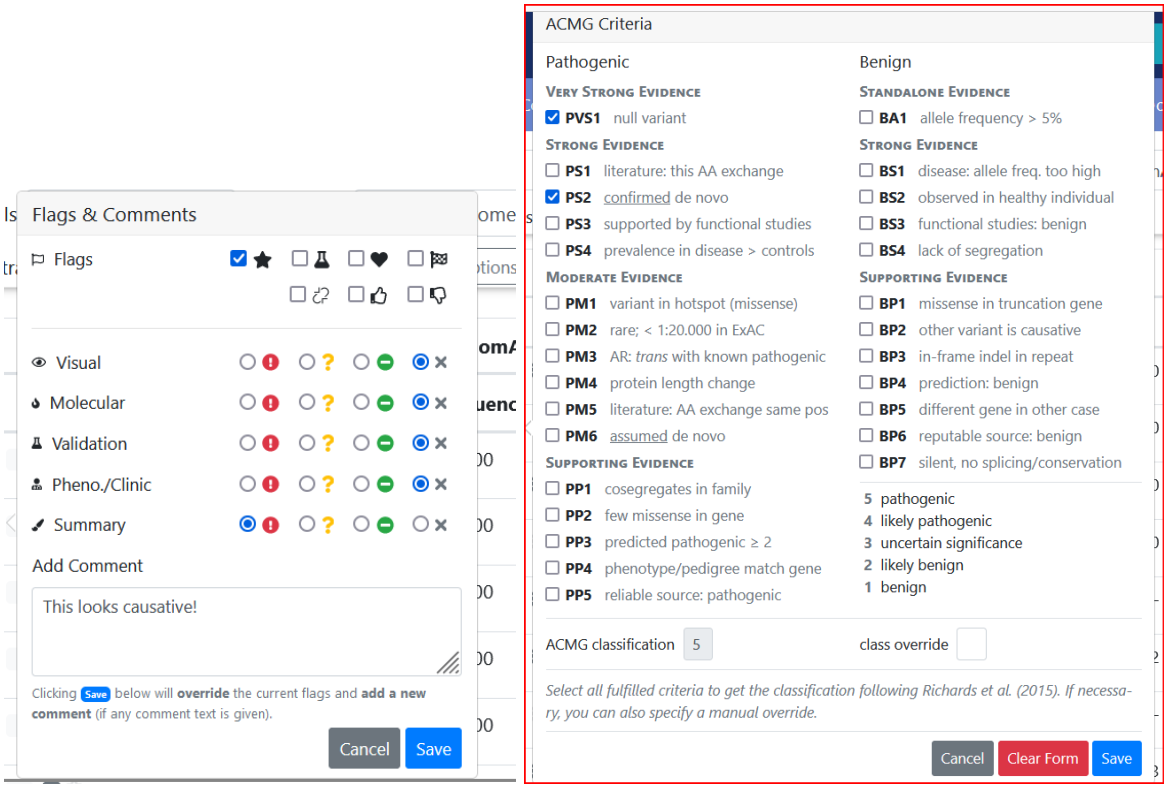


Figure 17. User interface for annotating variants.

Overall, a variant annotated as causative could look as shown in Figure 18.

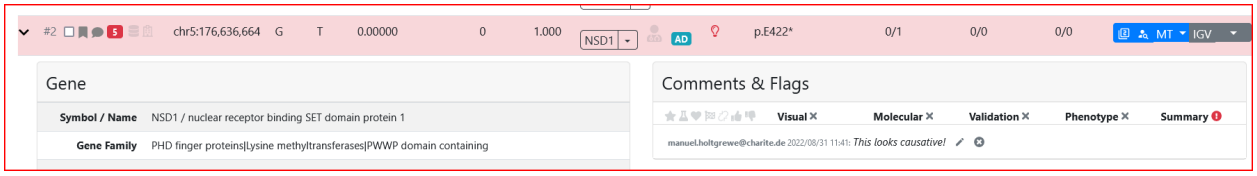


Figure 18. User annotation for causative variant.

This is the end of this walkthrough. We hope that we could give you a good impression about VarFish and its features and we look forward to you joining our training.

3 Example Cases

This section contains example cases. Sharing full exomes/genomes is problematic because of privacy reasons. Nevertheless, real-world genetic/sequencing data is required for proper training.

We thus created the following example cases by adding (“spiking”) variants into known disease genes into public domain data from the IGSR (International Genome Sample Resource, formerly known as 1000 Genomes Consortium). The selected pathogenic variants are thus embedded into the variant data of healthy individuals which gives us a realistic setting for the course. Also, we have both small (SNV and indel) and structural variant data (such as copy number variants) available. Images were taken from the literature.

Note: It is generally believed that the donors of the IGSR/1000 Genomes sequencing data consist of healthy individuals in the sense that the donors are not affected with monogenetic early-onset diseases. However, we are aware of some cases where carriers of at least likely pathogenic variants of highly penetrant diseases have been sequenced (e.g., ASLX1:p.G1132Vfs*31 in NA12386). We have attempted to exclude such variants from the training data set but some variants might have escaped our attention.

Section 3.1 contains a short description of the cases as would also be done in a clinical case conference. Section 3.2 then shows the causative variants with additional explanation.

3.1 Case Descriptions

3.1.1 Case 1

This case is also used for the Varfish walkthrough in Section 2.

Phenotype
4-year-old female with tall stature, advanced eruption of teeth, feeding difficulties

Family History
- Family from Western Europe.
- Parents are unaffected, index is affected.

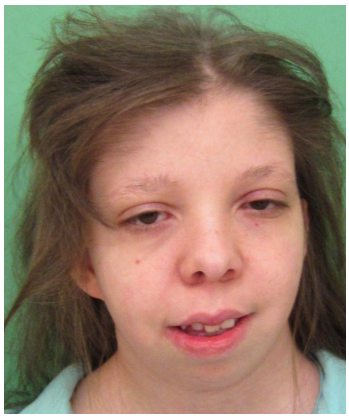
Technical Notes
- This is based on NA12386 which carries a likely pathogenic variant in ASXL1 (p.Glu1132Vfs*31). This variant has been removed by us from the variant data set.

3.1.2 Case 2

Phenotype
21-year-old female with intellectual disability, cerebellar vermis hypoplasia, hypertrichosis, ptosis, broad facial features, muscular hypotonia and short distal phalanges.

Family history

- Family from Western Europe.
- Parents are unrelated and unaffected, index is affected.



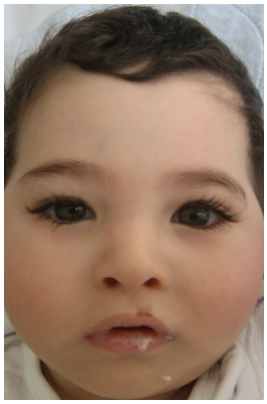
3.1.3 Case 3

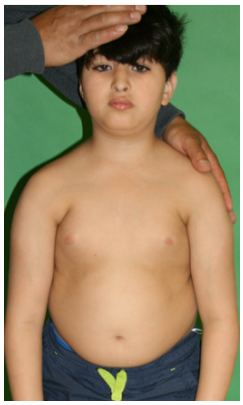
Phenotype

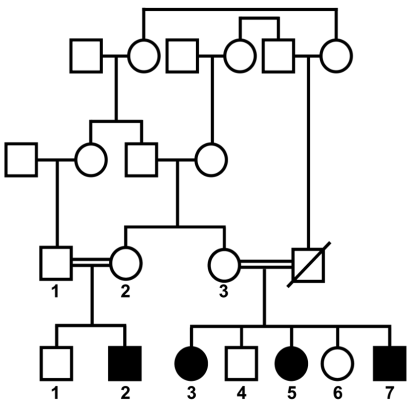
11-year-old male with intellectual disability, glaucoma, macula edema, myopia, mild postnatal microcephaly, mild short stature, truncal obesity, long eyelashes and short philtrum.

Family History

- Parents are first cousins from Jordan
- Three cousins with developmental delay, pigmentary retinopathy, microcephaly, thick hair and short philtrum.



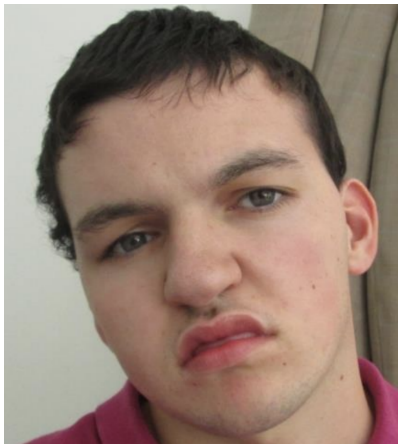




3.1.4 Case 4

Intentionally missing because of limitations in simulation

3.1.5 Case 5

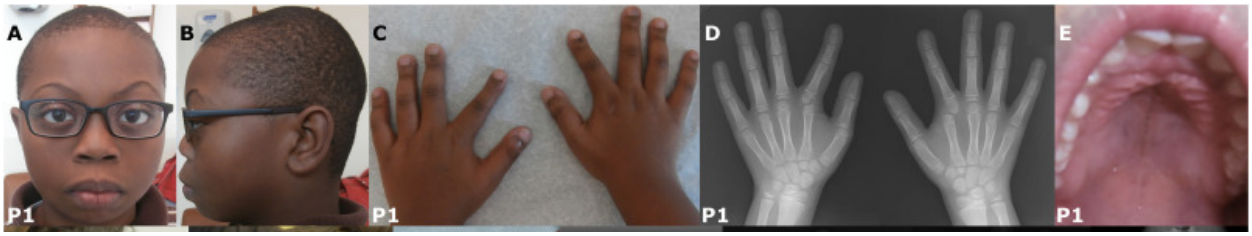


Phenotype
18 yo male with Intellectual disability, seizures, elevated alkaline phosphatase

Family History
- Family from Western Europe (non-consanguineous)
- Parents unaffected, index affected

Miscellaneous
Hint: Think about reasonable gene panels on PanelApp for the phenotypes.

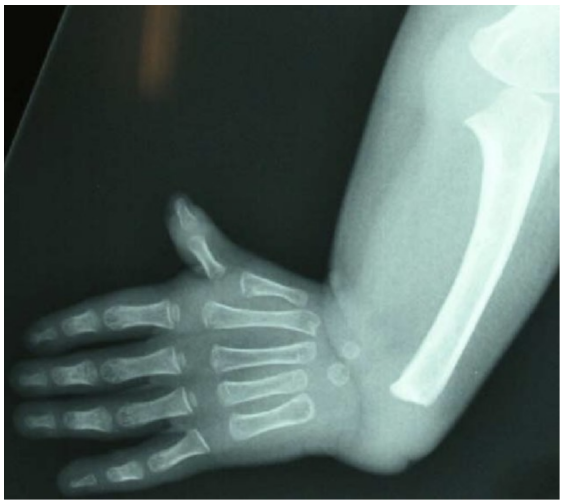
3.1.6 Case 6



Phenotypes
13-year-old boy with short stature, butterfly vertebrae, radial deviation and clinodactyly of the index fingers, microretrognathia, hypoplastic left heart, mild developmental delay.

Family History
- Non-consanguineous family
- Parents unaffected, index affected
- Father is not available

3.1.7 Case 7

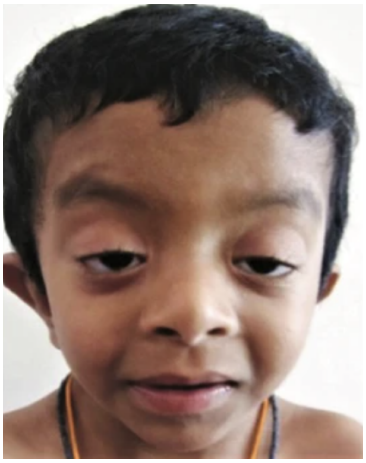


Phenotype
1 yo boy with 40 mio platelets per liter and abnormal findings in radiograph of the upper limb

Family History
- Family from Western Europe (non-consanguineous)
- Parents unaffected, index affected

Miscellaneous
Hint: Think about reasonable gene panels on PanelApp

3.1.8 Case 8

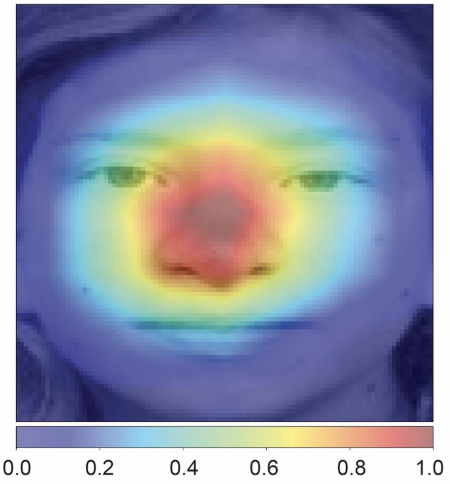


Phenotype
6 yo female with pulmonary stenosis, ptosis, and mild intellectual disability

Family History
Healthy, non-consanguineous parents from India

3.1.9 Case 9

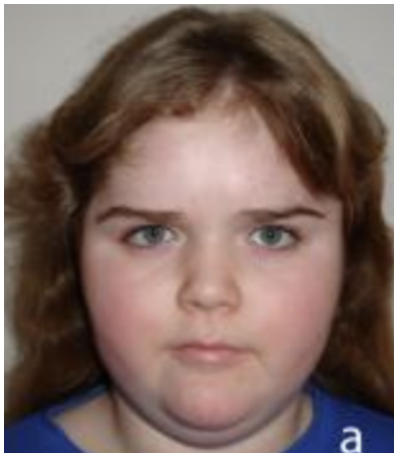
Phenotype
16 yo female with muscular hypotonia since early childhood and moderate intellectual disability.
Facial dysmorphism with bulbous tip of the nose. Very characteristic for a dominant ID syndrome. (heatmap from GestaltMatcher analysis).



Family History
Non-consanguineous parents and a sister. All not affected.

Hint:
Discuss how to reduce the search space effectively with the phenotype information

3.1.10 Case 10



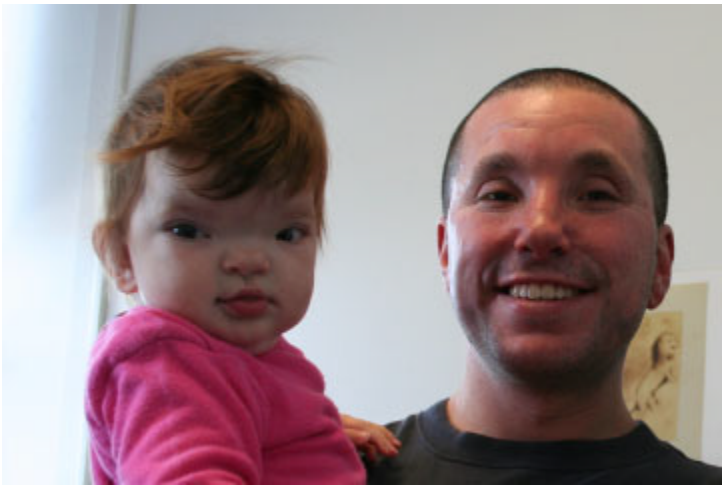
Phenotype
10 yo girl (height 155cm weight 65kg) with moderate ID

Family History
Only child of non-consanguineous parents of European descent

3.1.11 Case 11

Phenotypes
5-months-old girl with frontal bossing, hypertelorism, telecanthus, bifid nasal tip and one dystrophic nail.

Family History
- Unrelated parents from Europe
- Father has telecanthus and had surgery for pectus carinatum. Mother had one therapeutic abortion for fetal diaphragmatic hernia.



3.2 Case Solutions

3.2.1 Case 1

Synopsis
Sotos syndrome, *de novo* based

Disease
Sotos Syndrome OMIM: 117550 (https://www.omim.org/entry/117550)

Causative Variant

Gene	NSD1
HGVS(c)	NM_022455.5:c.1264G>T
HGVS(p)	NP_071900.2:p.(E422*)

Genome 37	GRCh37:5:176636664:G:T
Genome 38	GRCh38:5:177209663:G:T
Genotypes	index=0|1, father=0|0, mother=0|0
Allelic Balance	index~0.5, father=0.0, mother=0.0

Incidental Finding Variant

Gene	BRCA1
HGVS(c)	NM_007294.4:c.5503C>T
HGVS(p)	NP_009225.1:p.(R1835*)
Genome 37	GRCh37:17:41197784:G:A
Genome 38	GRCh38:17:43045767:G:A
Genotypes	index=0|1, father=0|0, mother=0|1
Allelic Balance	index~0.5, father=0.0, mother=~0.5

Solution

Approach

- Start with hypothesis: de novo (quick presets). The *NSD1* variant fits well with the phenotype.
- Continue with other presets (comp het, homozygous recessive, X-linked), no suitable candidates; no suitable SVs.
- ClinVar pathogenic preset: *BRCA1* variant (incidental finding)

Explanation

The *NSD1 de novo* variant causes a premature stop codon. The gene is known to cause Sotos Syndrome which matches the described phenotype and the image. The *BRCA1* variant is maternally inherited and has been reviewed by a ClinVar expert panel.

Image Source

Tatton-Brown, K., Rahman, N. Sotos syndrome. Eur J Hum Genet 15, 264–271 (2007). https://doi.org/10.1038/sj.ejhg.5201686

3.2.2 Case 2

Synopsis

Coffin-Siris Syndrome, Mosaik *de novo*

Disease

Coffin-Siris Syndrome 2 (https://www.omim.org/entry/614607)

Causative Variant

Gene	ARID1A
HGVS(c)	NM_006015.6:c.5532G>A
HGVS(p)	NP_006006.3:p.(W1844*)
Genome 37	GRCh37:1:27105921:G:A
Genome 38	GRCh38:1:26779430:G:A

Genotypes	index=0|1, father=0|0, mother=0|0
Allelic Balance	index~0.23, father=0.0, mother=0.0

Solution

Approach

- Start with hypothesis: de novo. Initially, no good candidate is found for the phenotype.
- Continue with hypothesis comp. Het. and homozygous recessive, no good candidate.
- Continue with structural variants: no candidates.
- Possible explanation: mosaic *de novo* variant
- Relax quality thresholds to “strict” shows *ARID1A* which is a good candidate

Explanation

A *de novo* base exchange causes a premature stop codon in *ARID1A*, associated with Coffin Siris syndrome. The variant is present in 23% of the reads, most likely a mosaic variant, which has been observed for several *ARID1A* nonsense variants. The variant can only be detected when changing the quality settings of the *de novo* filter to strict or relaxed. With the presets of the *de novo* filter the variant will not be displayed because the allelic balance is below 0.3.
Pitfall: low allelic balance (loosen qual)

Image Source

Wieczorek, D., Bogershausen, N., Beleggia, F., Steiner-Haldenstatt, S., Pohl, E., Li, Y., Milz, E., Martin, M., Thiele, H., Altmuller, J., Alanay, Y., Kayserili, H., and 44 others. A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum. Molec. Genet. 22: 5121-5135, 2013. https://academic.oup.com/hmg/article/22/25/5121/575160

3.2.3 Case 3

Synopsis

Cohen Syndrome, comp. het. SNV + deep intronic

Disease

Cohen Syndrome (https://www.omim.org/entry/216550)

Causative Variant

Gene	VPS13B (COH1)
Variant 1: CNV	
Genome 37	GRCh37:8:100246250-100460500:DEL
Genome 38	GRCh38:8:99234022-99448272:DEL
HGVS(c)	N/A
HGVS(p)	N/A
Genotypes	index=0|1, father=0|1, mother=0|0
Allelic Balance	index=0.5, father=0.0, mother=0.5
Variant 2: deep intronic	
Genome 37	GRCh37:8:100479619:T:G

Genome 38	GRCh38:8:99467391:T:G
HGVS(c)	NM_152564.5:c.3446-23T>G
HGVS(p)	NP_689777.3:p.?
Genotypes	index=0|1, father=0|0, mother=0|1
Allelic Balance	index=0.5, father=0.5, mother=0.0

Solution

Approach

- Start with hypothesis: homozygous recessive (indicated by pedigree). No suitable results.
- No suitable *de novo*, compound heterozygous or mitochondrial candidate SNVs.
- Filter structural variants: paternally inherited 214 kb multi exon deletion of VPS13B, the gene associated with Cohen syndrome (suitable diagnosis).
- Filter SNVs for maternally inherited alteration: genotype index 0/1, father 0/0, mother 0/1; frequency recessive strict; impact: whole transcript; more -> gene lists and regions -> gene allowlist: VPS13B shows c.3446-23T>G in VPS13B

Explanation

In this case, Cohen Syndrome is caused by a deep intronic SNV (affecting splicing) and a deletion that are present in compound heterozygous state. The deletion includes exons 18-23 of VPS13B. The second variant is a deep intronic variant which is predicted to affect splicing (predictions from Splice AI and VarSEAK splicing: formation of a cryptic splice acceptor site) and is annotated as pathogenic in ClinVar. It has been shown that this variant leads to inclusion of 22 bp intronic sequence, frameshift and premature stop codon (Boschann et al., 2020). The cousins are homozygous for the splice variant. Biallelic loss of function variants of this gene are known to be causative for the phenotype.

Image Source

Boschann, Felix, et al. "An intronic splice site alteration in combination with a large deletion affecting VPS13B (COH1) causes Cohen syndrome." *European Journal of Medical Genetics* 63.9 (2020): 103973. DOI: 10.1016/j.ejmg.2020.103973

3.2.4 Case 4

Intentionally Missing

3.2.5 Case 5

Synopsis

Recessive case with missense SNV + 5’ UTR variant.

Disease

Mabry Syndrome (https://www.omim.org/entry/239300)

Causative Variant

Variant 1: Missense variant

Gene	PGAP3
HGVS(c)	NM_033419.5:c.860G>T
HGVS(p)	NP_219487.3:p.(W287L)
Genome 37	GRCh37:17:37829343:C:A
Genome 38	GRCh38:17:39673090:C:A
Genotypes	index=0|1, father=0|1, mother=0|0
Allelic Balance	index=0.5, father=0.5, mother=0.0

Variant 2: 5' UTR variant

Gene	PGAP3
Genome 37	GRCh37:17:37828497:G:A
Genome 38	GRCh38:17:39672244:G:A
HGVS(c)	NM_033419.5:c.*559C>T
HGVS(p)	NP_219487.3:p.?
Genotypes	index=0|1, father=0|0, mother=0|1
Allelic Balance	index=0.5, father=0.0, mother=0.5

Note that we replaced the variant missense NM_033419.5:c.861G>T from the manuscript with the "invented" variant NM_033419.5:c.860G>T because the first is in ClinVar and we wanted to use a novel variant in the training. The original variant leads to p.W287C whereas the "invented" variant leads to p.W287L. The PhyloP100way, MutationTaster and CADD scores are the same for both variants. We retained the original 5' UTR variant as we did not consider it possible to "invent" an equivalent variant.

Explanation

With these clinical features, GPIBD (glycosylphosphatidylinositol biosynthesis defects) is likely. Therefore, a gene panel with PIG* and PGAP* genes allows to narrow down the variants.

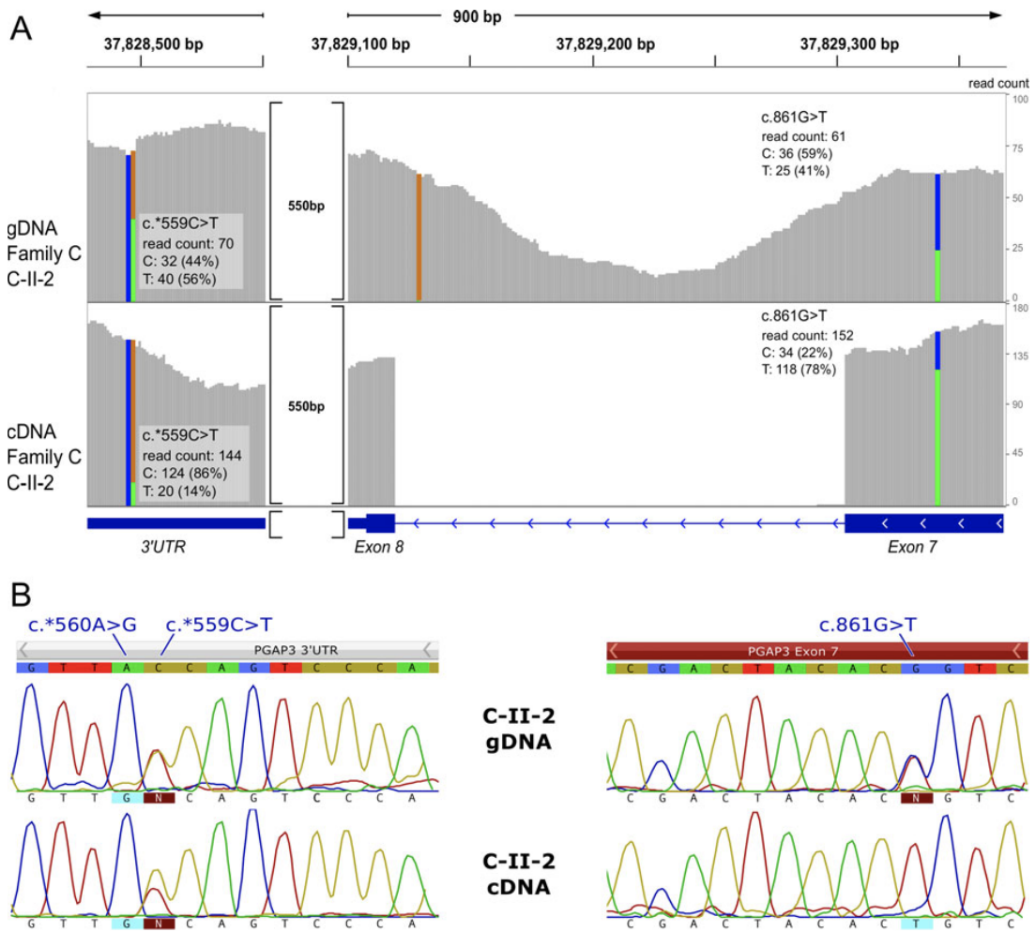


Figure. Depiction of the variants identified by Knaus et al. (2016).

Image Source
Knaus, Alexej, et al. "Rare noncoding mutations extend the mutational spectrum in the PGAP3 subtype of hyperphosphatasia with mental retardation syndrome." *Human mutation* 37.8 (2016): 737-744.

3.2.6 Case 6

Synopsis
Recessive case with SNV and CNV deletion; father unavailable

Disease
Vertebral, cardiac, renal and limb defect syndrome 2 (https://omim.org/entry/617661)

Causative Variant
Variant 1: Missense

Gene	KYNU
HGVS(c)	NM_003937.3:c.1283G>T

HGVS(p)	NP_003928.1:p.(R428L)
Genome 37	GRCh37:2:143799626:G:T
Genome 38	GRCh38:2:143042057:G:T
Genotypes	index=0|1, father=NA, mother=0|1
Allelic Balance	index=0.5, father=NA, mother=0.5
Variant 2: CNV Deletion	
Gene	KYNU
HGVS(c)	N/A deletion of exons 1-8
HGVS(p)	N/A
Genome 37	GRCh37:2:143632531-143721206:DEL
Genome 38	GRCh38:2:142874961-142963637:DEL
Genotypes	index=0|1, father=NA, mother=0|0
Allelic Balance	index=0.5, father=NA, mother=0.0

Note that we replaced the variant NM_003937.2:c.1282C>T from the publication with the "invented" variant NM_003937.3:c.1283G>T as we wanted the training to show a novel variant not in ClinVar yet. The original variant leads to p.R428W while the "invented" variant leads to p.R428L. The PhyloP100way score increases from -0.08 to 4.87 and CADD from 23.4 to 25.4 from the original to the "invented" variant. The MutationTaster result is "disease-causing" in both cases.

Solution

Approach

Option 1:

- Filter for structural variants (index: variant): detection of 88 kb mulit-exon deletion of *KYNU* (good candidate, autosomal recessive inheritance), not maternally inherited
- Filter SNVs for maternally inherited second alteration (genotype index 0/1, mother 0/1; frequency recessive strict; impact: whole transcript shows; more -> gene lists and regions -> gene allowlist: *KYNU*). This shows NM_003937.2:c.1283G>T;p.(R428L)

Option 2:

- Filter SNVs with phenotype prioritization (genotype: index variant, mother any; frequency: recessive strict; impact: AA change and splicing; quality: super strict; prioritization: HPO: HiPhive human and CADD), enter HPO terms; first candidate: *KYNU* NM_003937.2:c.1283G>T;p.(R428L)
- Filter SV for second hit, not maternally inherited

Explanation

Biallelic LOF variants in *KYNU* cause the described phenotype. The missense variant has been identified 5x het in gnomAD. A different aa change at the same position has been annotated in ClinVar as pathogenic (and has been published by Ehmke et al.). We cannot prove that the two variants are biallelic due to absence of paternal sequence data.

Image Source

Ehmke, Nadja, et al. "Biallelic variants in KYNU cause a multisystemic syndrome with hand hyperphalangism." *Bone* 133 (2020): 115219.

3.2.7 Case 7

Synopsis
Compound heterozygous case with CNV deletion and 5' UTR polymorphism (AF >1%)

Disease
Thrombozytopenia absent radii (TAR) syndrome (https://omim.org/entry/274000)

Causative Variant

Variant 1: 5' UTR	
Gene	RBM8A
HGVS(c)	NM_005105.5:c.-21G>A
HGVS(p)	NP_005096.1:p.?
Genome 37	GRCh37:1:145507646:G:A
Genome 38	GRCh38:1:145927447:C:T
Genotypes	index=0|1, father=0|0, mother=0|1
Allelic Balance	index=0.5, father=0.0, mother=0.5
Variant 2: DEL	
Gene	RBM8A
HGVS(c)	N/A deletion of exons 2-4
HGVS(p)	N/A
Genome 37	GRCh37:1:145507968-145508657:DEL
Genome 38	GRCh38:1:145926435-145927124:DEL
Genotypes	index=0|1, father=0|1, mother=0|0
Allelic Balance	index=0.5, father=0.0, mother=0.0

Solution and explanation
The mode of inheritance of TAR syndrome was unclear for quite some time. Did you find only the rare deletion and suspected a dominant mode of inheritance? This is also what the community thought first. However, unaffected carriers with a 200kb microdeletion were identified implying that haploinsufficiency of the deleted region is not sufficient to cause TAR syndrome. It took some while until the second hit, NM_005105.5:c.-21G>A, was identified to be pathogenic because it has an AF>1%. However, there are also unaffected homozygotes of NM_005105.5:c.-21G>A. Thus, you may discuss whether TAR fits to a classicial autosomal recessive mode of inheritance. In the literature it is described as a compound inheritance of a low-frequency noncoding SNP and a rare null allele in RBM8A (Albers, *et al.*)

Image Source
Albers, Cornelis A., et al. "New insights into the genetic basis of TAR (thrombocytopenia-absent radii) syndrome." *Current opinion in genetics & development* 23.3 (2013): 316-323.

Elmakky, et al. “Role of genetic Factors in the Pathogenesis of Radial Deficiencies in Humans” Current Genomics 2015, 16, 264-268

3.2.8 Case 8

Synopsis
Missense variant in index causes disease with parental mosaicism in unaffected father

Disease
Noonan Syndrome, NS (https://www.omim.org/entry/163950)

Causative Variant

Gene	PTPN11
HGVS(c)	NM_002834.5:c.166A>G
HGVS(p)	NP_002825.3:p.(I56V)
Genome 37	GRCh37:12:112888150:A:G
Genome 38	GRCh38:12:112450346:A:G
Genotypes	index=0|1, father=0|1, mother=0|0
Allelic Balance	index~0.5, father=0.3, mother=0.0

Solution and Explanation
The pedigree suggests a *de novo* mutation or an AR cause. However, the disease causing variant can only be found after relaxing the filter so that also rare variants with incomplete penetrance, or parental mosaics can pass. This is particularly important for syndromes such as Noonan in which a large phenotypic variability is observed.
Note that germline mosaicism is also age-dependent for mutations in the MAPKinase pathway indicating a selectional advantage.

Image Source
Athota, et al., Molecular and clinical studies in 107 Noonan syndrome affected individuals with *PTPN11* mutations, *BMC Medical Genetics 2020*
https://bmcmedgenet.biomedcentral.com/articles/10.1186/s12881-020-0986-5

3.2.9 Case 9

Synopsis
(mostly) Intronic microdeletion in single index

Disease
Koolen-de Vries syndrome (https://www.omim.org/entry/610443)

Causative Variant

Gene	KANSL1
HGVS	NM_015443.4:c.1849-4611_1895del
HGVS(p)	NP_056258.1:p.?

Genome 37	GRCh37:17:44128024-44132681:DEL
Genome 38	GRCh38:17:46050658-46055315:DEL
Genotypes	index=0|1, father=NA, mother=NA
Allelic Balance	index=0.5, father=NA, mother=NA

Solution and Explanation

The facial gestalt suggests Koolen de Vries syndrome, however, no pathogenic variants can be found with the default settings for variants and SVs. Reducing the minimal SV size to 500 shows a variant of size ~4.5kb. Small deletions (<5kb) are difficult to call in exome data, but easy in genome data. The lesson from this case is: if phenotype information indicates a certain disorder with high evidence, you should screen the alignment.

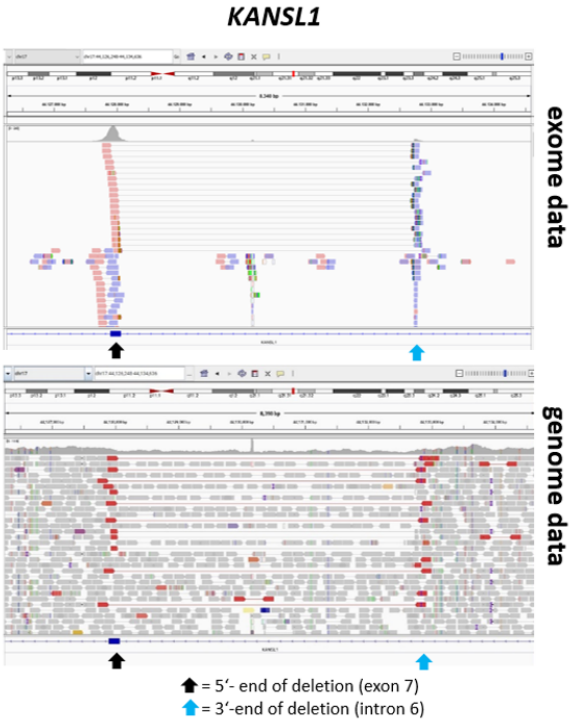


Image Source

Fabian Brand, Peter Krawitz, Claudia Perne. Next-generation phenotyping contributing to the identification of a non-coding deletion in *KANSL1* causing Koolen-de Vries syndrome. Human Mutation. (accepted)

3.2.10 Case 10

Synopsis

Variant that causes an ID syndrome if *de novo* in germline, but that also occurs in higher age due to clonal hematopoiesis

Disease

Tatton-Brown-Rahman syndrome (https://omim.org/entry/615879)

Causative Variant

Gene	DNMT3A
HGVS(c)	NM_022552.5:c.994G>A
HGVS(p)	NP_072046.2:p.(G332R)
Genome 37	GRCh37:2:25470480:C:T
Genome 38	GRCh38:2:25247611:C:T
Genotypes	index=0|1, father=0|0, mother=0|0
Allelic Balance	index=0.31, father=0|0, mother=0|0

Explanation

Variants that occur due to clonal hematopoiesis in elder individuals without clinical consequences can cause “false negative” filtering results if the de novo filter is very strict. The case could be solved by relaxing the population frequency filter. However, since in most IDs that are due to *de novo* mutations the penetrance is close to 100%, all carriers have to be carefully inspected. This variant only occurs in old people which is compatible with clonal hematopoiesis

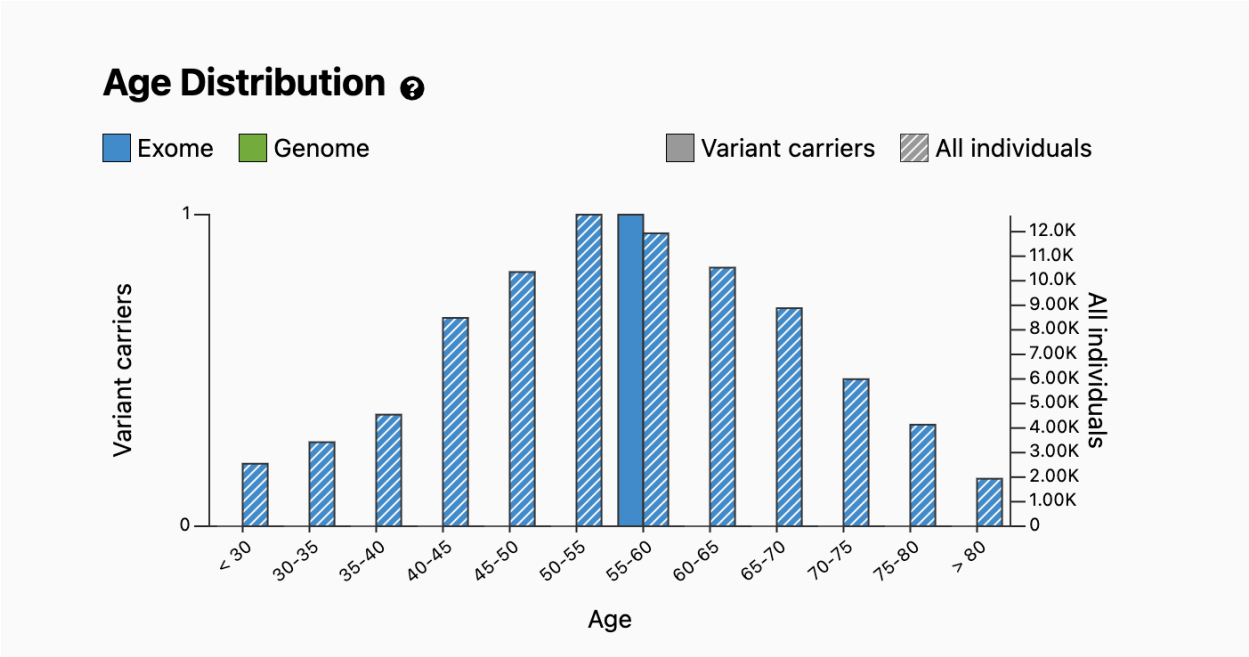


Image Source and literature
Screenshot from gnomAD
Brunet, et al. Clonal hematopoiesis as a pitfall in germline variant interpretation in the context of Mendelian disorders (https://pubmed.ncbi.nlm.nih.gov/35179199/)
Tatton-Brown, et al. Mutations in the DNA methylthrasnferase gene DNMT3A cause an overgrowth syndrome with intellectual disability.

3.2.11 Case 11

Synopsis
Paternally inherited heterozygous variant; X-linked dominant inheritance with more pronounced phenotype in affected females.

Disease
Craniofrontonasal syndrome (https://www.omim.org/entry/304110)

Causative Variant

Gene	EFNB1
HGVS(c)	NM_004429.5:c.409A>T
HGVS(p)	NP_004420.1:p.(T137S)
Genome 37	GRCh37:X:68059508:A:T

Genome 38	GRCh38:X:68839666:A:T
Genotypes	index=0|1, father=1|1, mother=0|0
Allelic Balance	index~0.5, father=1.0, mother=0.0

Note that we replaced the original variant NM_004429.5:c.407C>T with the ”invented” variant NM_004429.5:c.409A>T. Rather than p.S136L, this causes p.T137S but the predicted impact is at least the same (PhyloP100way of 9.06 instead of 7.682, MutationTaster disease-causing in both, CADD score falls from 31 to 10.9 but VarSome still gives 18/20 predictors as pathogenic).

Solution

Approach

- Filter SNVs with phenotype prioritization (genotype: index variant, parents any; frequency: recessive strict; impact: AA change and splicing; quality: super strict; prioritization: HPO: HiPhive human), enter HPO terms; first candidate: *EFNB1* NM_004429.5:c.409A>T

Explanation

None of the existing presets could find the variant because it is a rare case of a pathogenic variant inherited from the mildly affected father. The variant affects a highly conserved aa and has not yet been observed. Craniofrontonasal syndrome is a suitable diagnosis for the girl and the father’s clinical features represent the mild end of the phenotype, occasionally observed in affected males.

Image Source

Hogue, J., Shankar, S., Perry, H., Patel, R., Vargervik, K., Slavotinek, A. A novel EFNB1 mutation (c.712delG) in a family with craniofrontonasal syndrome and diaphragmatic hernia. Am. J. Med. Genet. 152A: 2574-2577, 2010.
https://onlinelibrary.wiley.com/doi/epdf/10.1002/ajmg.a.33596

Appendix

A Abbreviations

CUBI	Core Unit Bioinformatics
HGVS	Human Genome Variation Society
IGSR	International Genome Sample Resource

NGS	Next-Generation Sequencing
SNV	Single Nucleotide Variant
SV	Structural Variant