



Tutorial

Finding a Rare Disease Causing Variant

June 20, 2019

— Sample to Insight —

Finding a Rare Disease Causing Variant

Hugh Rienhoff's daughter was born with unusual features usually associated with syndromes such as Marfan and Loeys-Dietz, although she was not affected by these syndromes. Trained as a clinical geneticist, Hugh Rienhoff spent a decade sequencing whole exomes of his family members and analyzing the data to find the mutation responsible for his daughter's rare genetic disease (see <http://www.nature.com/news/father-s-genetic-quest-pays-off-1.13269>). This tutorial uses a subset of the whole exome data provided by Hugh Rienhoff and the capabilities of Biomedical Genomics Analysis plugin and the Ingenuity Variant Analysis plugin to recapitulate within minutes the search for a unique deleterious variant within the affected family.

This tutorial covers the following:

- Create a Custom Reference Data Set to support the functionalities of the workflow with reads mapping only to chromosomes 14 and 18.
- Find a list of variants specific to the proband's exome using the workflow Identify and Interpret Causal Variants in Trio using IVA (WES).
- Set up filters in Ingenuity Variant Analysis to discriminate among variants generated by the workflow.
- Update the results in the workbench after variants were filtered in Ingenuity Variant Analysis.

Prerequisites For this tutorial, you must have installed the Ingenuity Variant Analysis plugin, and you must have access to Ingenuity Variant Analysis services. How to install plugins is described here: <http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html>. A free trial account for Ingenuity Variant Analysis can be created using this link: <http://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/>. Please note that setting up an account can take 1-4 business days.

Download and import data

In this tutorial we will use a subset of the whole exome data provided by Hugh Rienhoff. The original data contains whole exome reads for three family members: the mother, father, and the proband. To allow you to run this tutorial within minutes, we have created a subset of the whole exomes to contain only the chromosomes 14 and 18, as well as a reference data set specific to these 2 chromosomes.

Go through the following steps to download and import the data into the Workbench.

1. Download the sample data from our website: http://resources.qiagenbioinformatics.com/testdata/Finding_rare_disease_causing_variant.zip.
2. Start the workbench.
3. Import the data via the toolbar: **Import**  | **Standard Import** 
4. Choose the zip file called Finding_rare_disease_causing_variant.zip. Leave the import type set to Automatic.

5. Save the imported data.

The data set includes:

- the reads from the father, mother and proband. These would normally be obtained by importing the .fastq files directly produced by the sequencer.
- a target regions file called *SeqCap_EZ_Exome_v2_BED_chr14-chr18*. Please note that when working with your data, you can obtain the relevant target region tracks from the vendor of the amplicon or hybridization kit.
- reference sequences:
 - *Homo sapiens (hg19) sequence-chr14-chr18*
 - *Homo sapiens (hg19) mRNA-chr14-chr18*
 - *Homo sapiens (hg19) CDS-chr14-chr18*
 - *Homo sapiens (hg19) Gene-chr14-chr18*

Once the data has been downloaded and imported, you should see the folder and files in the Navigation Area as shown in figure 1.

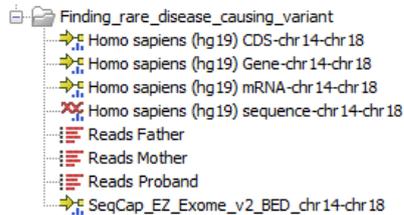


Figure 1: The data set has been imported.

Running the Ready-to-Use workflow

The **Identify and Interpret Causal Variants in Trio using IVA (WES)** workflow is an automated way to run a series of tools using the output of one as input for the next. During the workflow, The **Map Reads to Reference** tool maps the reads of each family member to the reference we selected for this tutorial. This is followed by the **Indels and structural Variants** and the **Local Realignment** tools that aim to improve the reads mapping. Variants are then detected using the **Fixed Ploidy Variant Detection** tool, and false positive variants are removed. The **Backcheck** tool is then used to check each family member genome thoroughly for any of the variants identified in the other family members. Finally, using the variant lists of each family member, **Ingenuity Variant Analysis for Hereditary Diseases** applies four different filter cascades, each to identify causal variants with a specific inheritance pattern: dominant, de novo, recessive compound heterozygous, and recessive homozygous. The variants of the proband that pass through the filter cascades are possibly responsible for the genetic disease under study, and these variants are displayed in a **Track List** for easy inspection. You can see a schema of the workflow by selecting and then right clicking on the name of the workflow. Choose to **Open Copy of Workflow** to see how the tools are connected to each other.

To run this workflow, go to:

Toolbox | Ready-to-Use Workflows | Whole Exome Sequencing (WES) | Hereditary Disease (HD) | Identify and Interpret Causal Variants in Trio using IVA (WES)

1. If you are connected to a server, you will first be asked where you would like to run the analysis.
2. Select the **sequencing reads** for the proband (figure 2) and click **Next**.

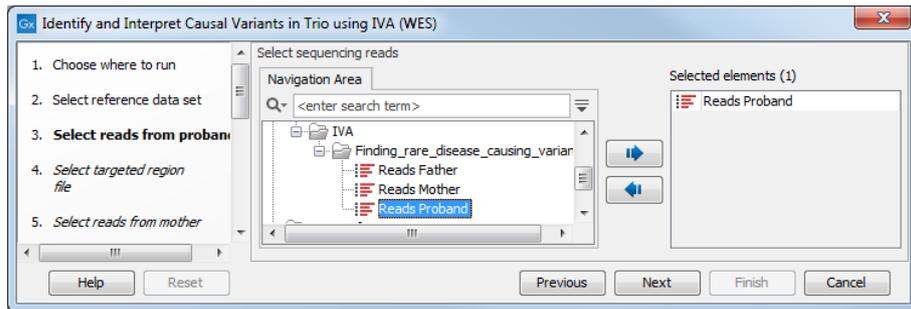


Figure 2: Specify the sequencing reads for the proband.

3. Specify the **target region** file *SeqCap_EZ_Exome_v2_BED_chr14-chr18* (figure 3), and click **Next**.

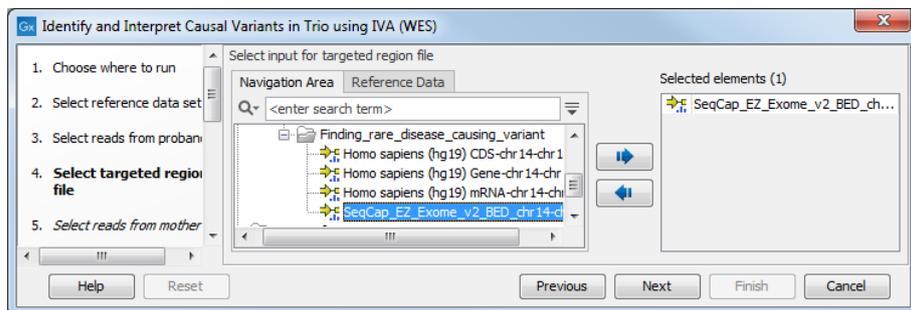


Figure 3: Specify a target region file

4. Select the **sequencing reads** for the Mother and Father (as done previously for the proband).
5. You have to specify the correct data set to use as reference. As we are working with data restricted to chromosome 14 and 18, we will choose to work with the data set created specifically for this tutorial as shown in (figure 4)
6. Leave the parameters for the Fixed Ploidy Variant Detection tool as default for the proband, mother, and father successively (figure 5).
7. In the **Ingenuity Variant Analysis for Hereditary Diseases** dialog, write in your Ingenuity Variant Analysis login information (figure 6), i.e., the email address and password you use to sign in the Ingenuity Variant Analysis. The parents in this family are unaffected, so leave those boxes unchecked. Note: Certain input combinations can lead to wrong results in IVA. Please refer to the plugin manual for supported input combinations, especially if you are analysing a family with an affected parent.

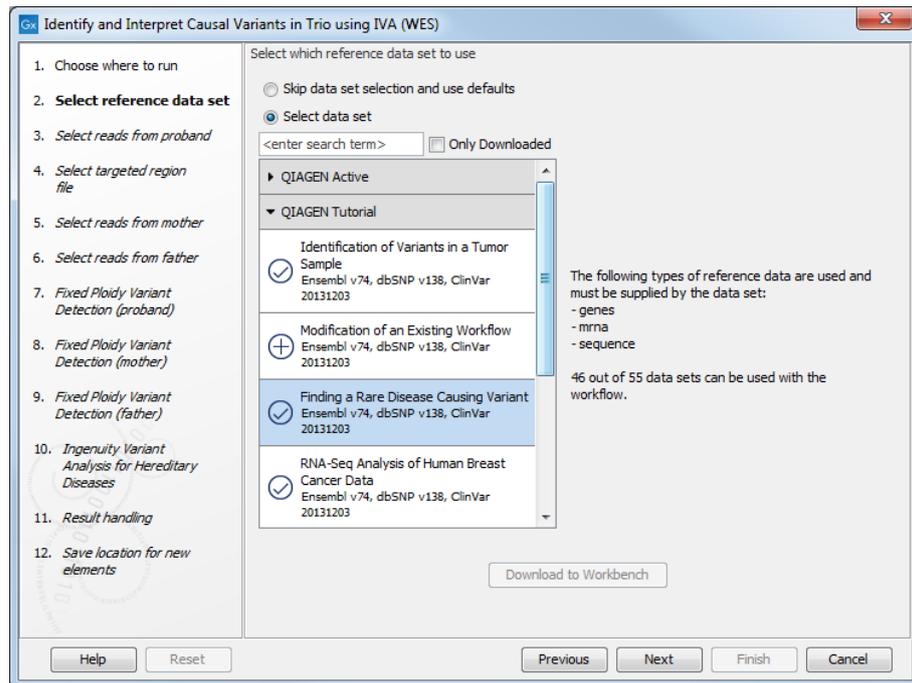


Figure 4: Specify the sequencing reads for the proband.

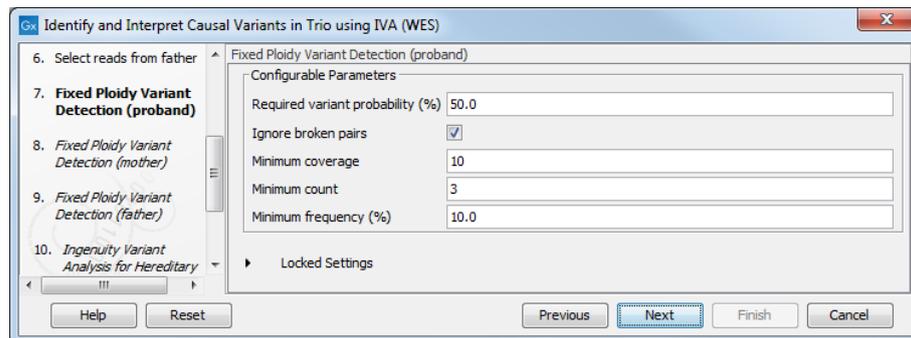


Figure 5: Specifying the parameters for the Fixed Ploidy Variant Detection tool.

- On the last wizard window, pressing the button **Preview All Parameters** allows you to preview all parameters. At this step you can only view the parameters, it is not possible to make any changes. Choose to save the results and click on the button labeled **Finish**.

After the workflow is done running, you will receive an email from support-ingenuity@qiagen.com with a link to your results. Alternatively you can obtain the link from the Navigation Area (figure 7).

Analyzing results in the workbench and Ingenuity Variant Analysis

Once the analysis is over, you can open the Track List by double clicking on the file with that name in the Navigation Area. It opens a Genome Track and a table listing all 315 variants identified in the proband prior to the application of Ingenuity Variant Analysis filter cascades. Those of the 315 proband variants that are located on chromosome 14 are shown in red in the proband track of figure 8. Use the drop down menu in the top right corner to choose which chromosome you

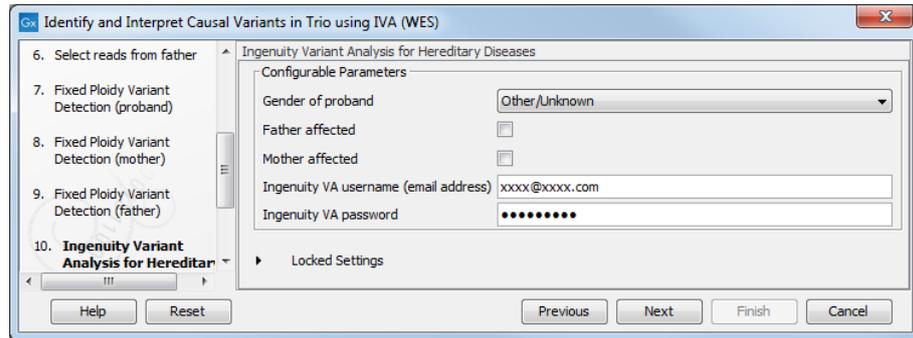


Figure 6: Specify login information to Ingenuity Variant Analysis.

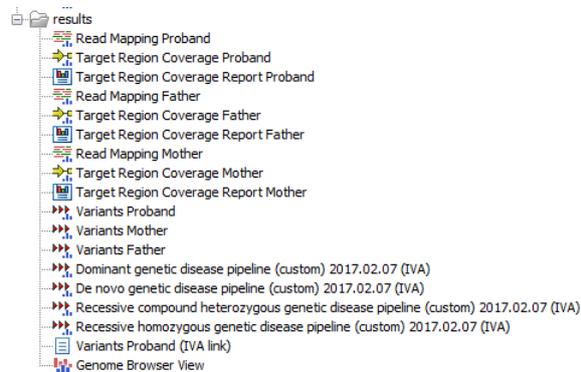


Figure 7: The results of the workflow are saved in your Navigation area, including a file containing the link to Ingenuity Variant Analysis.

want to look at. Drag tracks up and down in the Track List to create the most informative view for your purpose.

At the bottom of the Track List you find four tracks, each with causal variants (if any) exhibiting a specific inheritance pattern: dominant, de novo, recessive compound heterozygous, and recessive homozygous. By applying the Ingenuity Variant Analysis filter cascades, the 315 variants of unknown importance have been narrowed down to a few candidate causative variants (here 2 in the de novo genetic disease pipeline). To inspect the potentially disease causing variants, first double-click on the track name to open a variant table below. Note that the table includes 6 entries to describe the 2 de novo variants found: for example, some entries describe the reference allele, and two entries for the variant present on chromosome 18 differ by their Transcript ID.

The Track List and the table below are linked, so clicking on a variant in the table will display that variant in the Browser View at the nucleotide level, allowing you to look closely at the read mapping and coverage at a particular position. In the example below (figure 9), we have chosen to show the variants of all family members at the position of a potentially disease causing variant with de novo inheritance pattern.



Figure 8: Track List of chromosome 14.

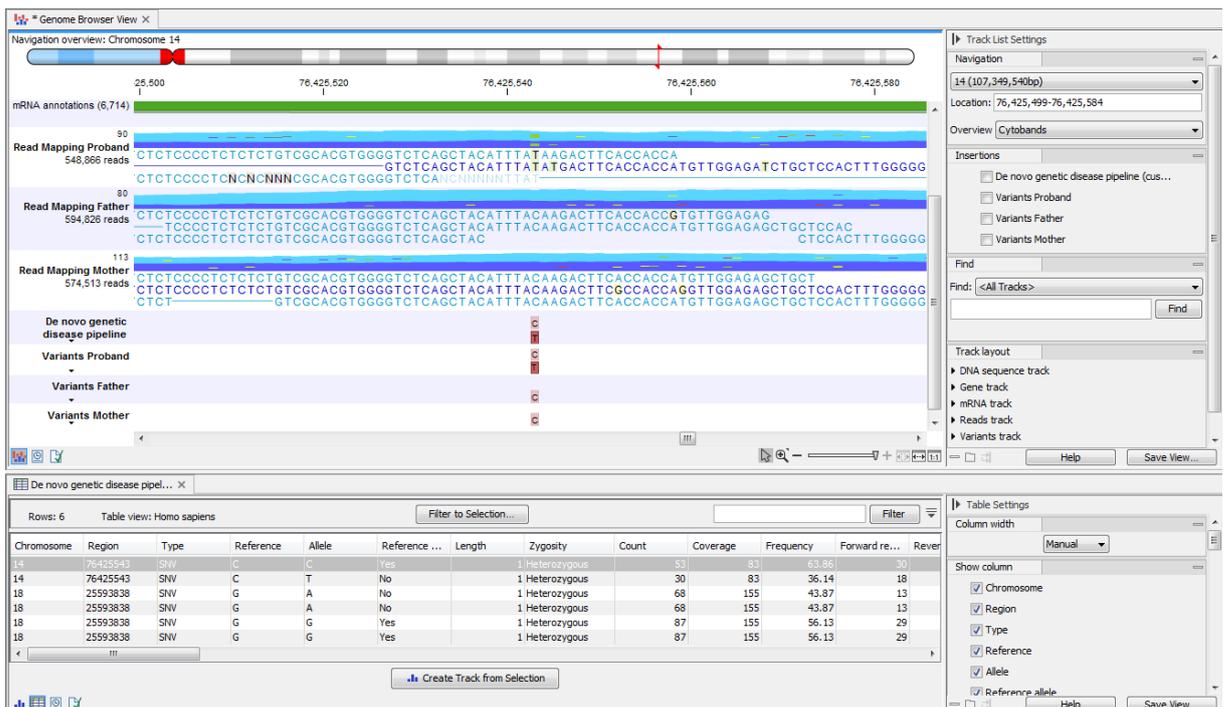


Figure 9: Details of the de novo variant found on chromosome 14.

To check the variants identified as potentially causal in Ingenuity Variant Analysis, go to the "Ingenuity Variant Analysis Link" column of the variant table and click on the link provided for either one of the variant. In the internet browser that opens, you can see the information about that particular variant. To the left, in the filter cascade, click on the last filter called "Genetic Analysis" to see both potentially disease-causing variants (figure 10). According to the Ingenuity Variant Analysis interpretation, the variant found in TGFB3 is potentially damaging.

Click on the TGFB3 row in the variant table. A window appears at the right hand side of

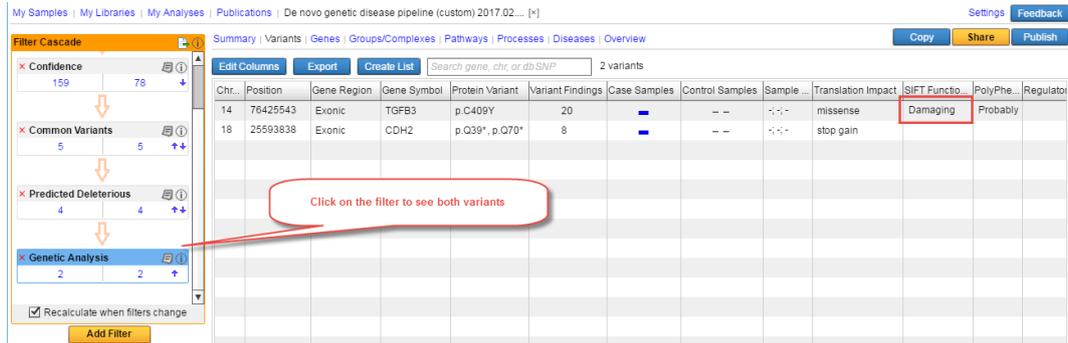


Figure 10: Ingenuity Variant Analysis interface showing the 2 genes where variants were found.

the table, offering links to additional information. In particular, click on "Variant Findings" or "Pharmacogenetics" to get access to scientific articles mentioning the gene TGFβ3 (figure 11).

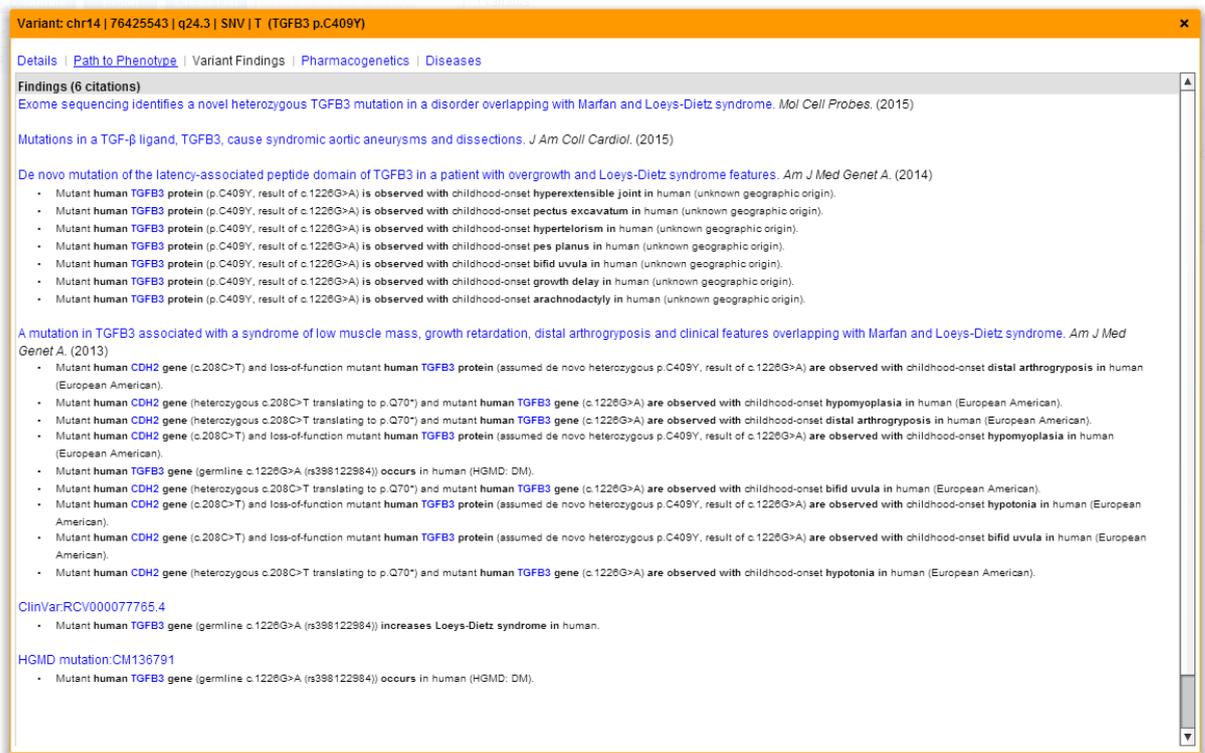


Figure 11: Ingenuity Variant Analysis gives you access to scientific articles on TGFβ3 in the context of a syndrome or a treatment.

To look further into which of the two variants are the most likely cause of the rare disease, we will expand the filter cascade with one additional filter, a phenotype filter that will rank the remaining variants based on the proband's phenotypes.

1. Below the Filter Cascade, click on the button called **Add filter** and choose the option "Phenotype-Driven Ranking" (figure 12).

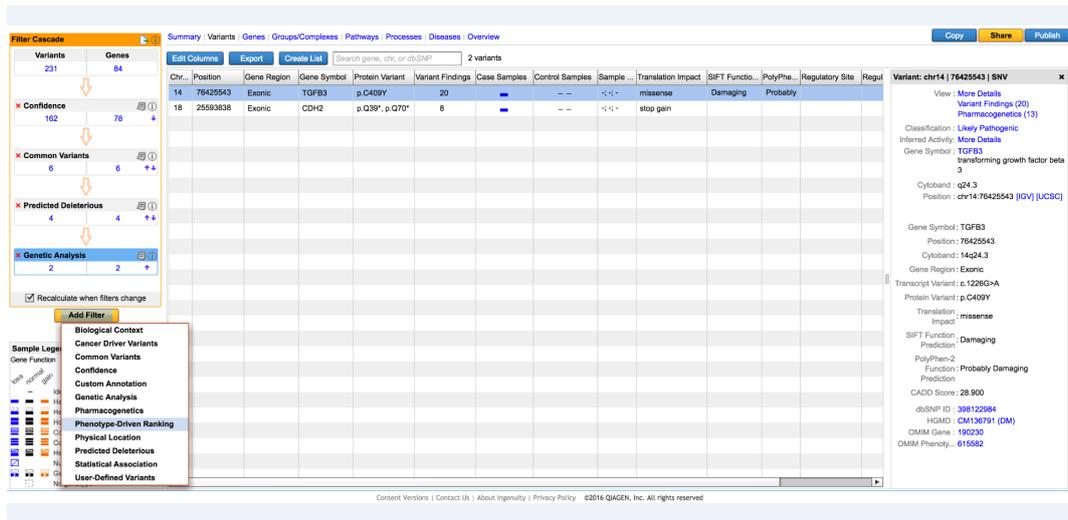


Figure 12: Add a Phenotype-Driven Ranking filter.

2. In the pop up window (figure 13), enter some of the proband's phenotypes, e.g., "hypertelorism" and "bifid uvula" (see <http://www.nature.com/news/father-s-genetic-quest-pays-off-1.13269>). Click on **Apply**.



Figure 13: Search for the relevant phenotypes in the drop down menu.

Selecting the phenotypes will result in a list reduced to a single variant, the one situated in TGFβ3 on chromosome 14 (figure 14). The Ingenuity Variant Analysis interface now shows a ranking of the most likely disease caused by the variant that was identified in the proband.

Feel free to explore additional features offered by Ingenuity Variant Analysis with online training and video tutorials available at <http://www.qiagenbioinformatics.com/ingenuity-variant-analysis-resources/>.

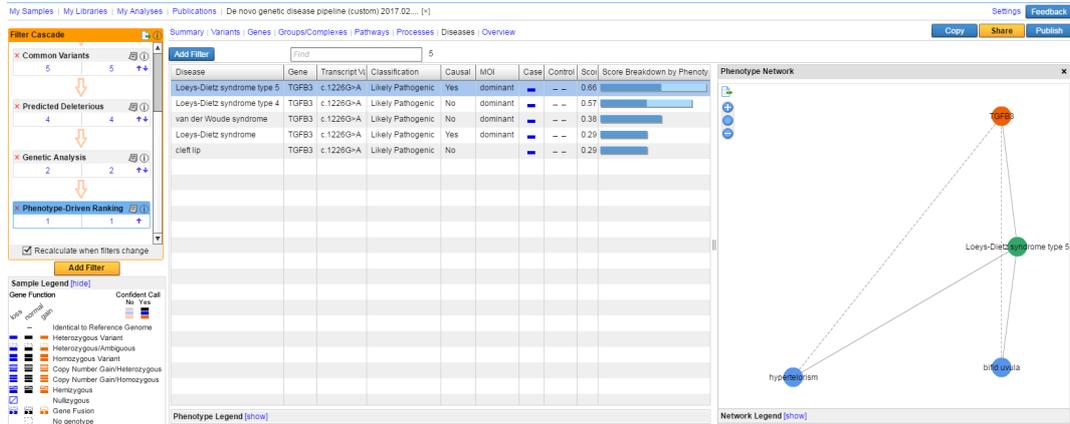


Figure 14: List of disease caused by the variant associated with some of the proband's phenotypes.

Updating the results in the workbench

Once the Filter Cascade has been adjusted in Ingenuity Variant Analysis, it is possible to update the results in the workbench. To do this:

1. Open the "De novo genetic disease pipeline (custom) <date> (IVA)" track and right click on the blue area to choose the option "Variant Analysis Update..." in the drop down menu (figure 15). This can also be done from the De novo track included in the Track List.

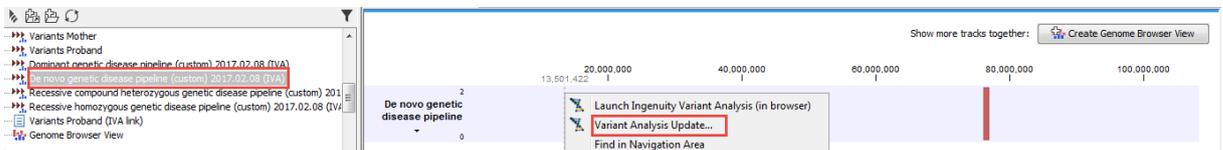


Figure 15: Update your results after you have applied filters in Ingenuity Variant Analysis.

2. In the wizard that opens, select the *Homo sapiens (hg19) sequence-chr14-chr18* and write in your Ingenuity Variant Analysis login information (figure 16).

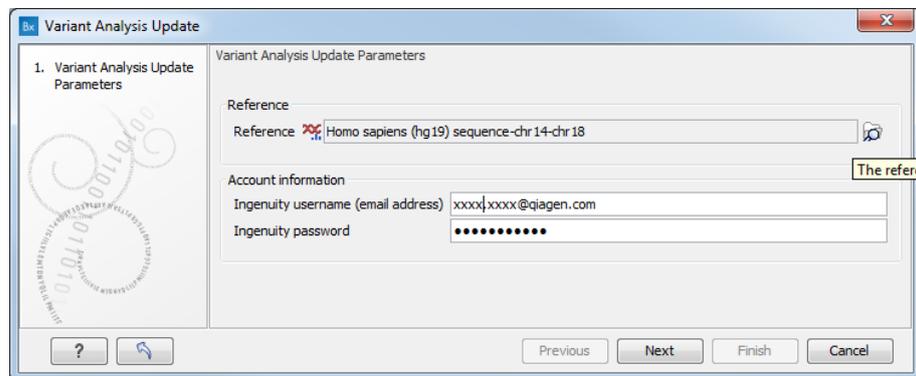


Figure 16: Variant Analysis Update wizard window.

3. Save the results in your Navigation Area (figure 17).

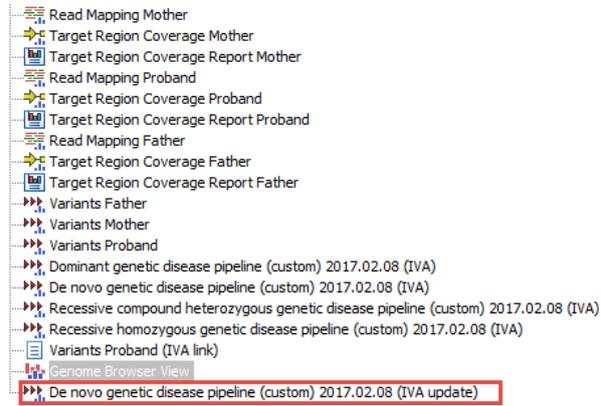


Figure 17: The updated file "De novo genetic disease (custom) <date> (IVA update)" is now saved in your Navigation Area.

A file called "De novo genetic disease (custom) <date> (IVA update)" will appear in the Navigation Area. You can drag this new variant track into the Track List to see that the updated variant track contains only one variant in TGFB3. Note that TGFB3 appears twice in the table because it lists both the variant and the reference allele.
