An automatic end-to-end solution for disease-causing variant detection in rare and hereditary diseases with a high case solve rate and a much reduced false positive rate

A. Joecker[1], S. Shah[2], F. Schacherer[1], R. Yip[2], D. Richards[2], B. Oester[1], G. Eley[3], B. Solomon[3], J.G. Vockley[3]
[1]QIAGEN Aarhus, Denmark; [2]QIAGEN Silicon Valley, Redwood City, USA
[3]Inova Translational Medicine Institute, VA, USA

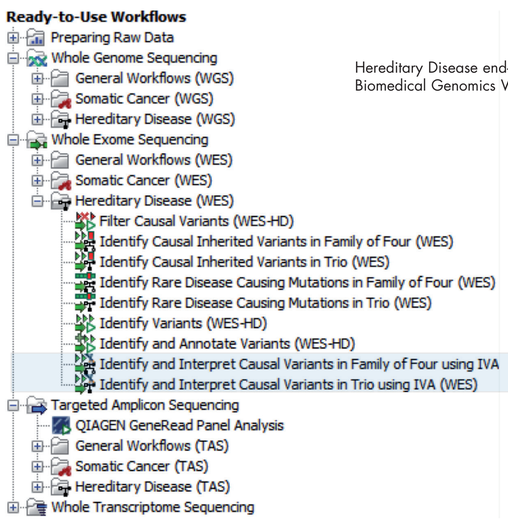
Motivation & Goals

Identification of causal variants in undiagnosed diseases can be both challenging and time consuming. Often a lot of time and resources are wasted trying to validate false positive candidates. Furthermore, sometimes no causal variant can be identified at all.

QIAGEN Bioinformatics' new hereditary disease solution delivers increased sensitivity for identifying causal variants, while reducing the false positive rate close to zero percent. In addition, it is fast and easy to execute as all steps are embedded in an end-to-end workflow with optimized parameter settings.

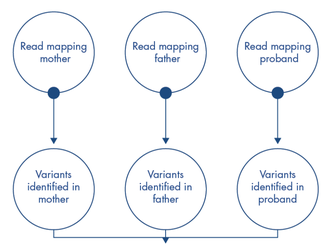
In this study we show the first results from a benchmarking study with four whole genome trios and one whole exome trio.

Materials & Methods



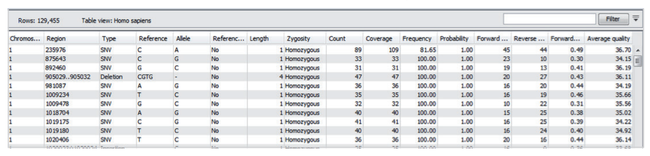
Hereditary Disease end-to-end workflows in Biomedical Genomics Workbench (BxWB)

The back-check part of the hereditary disease workflows increases the sensitivity when identifying variants

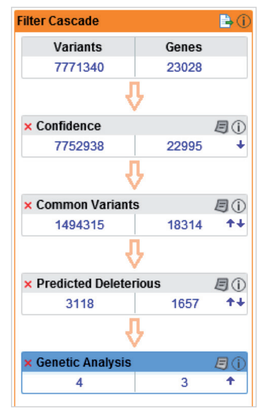


Do backcheck!
Have I really not missed anything?

Combined variant list from all family members

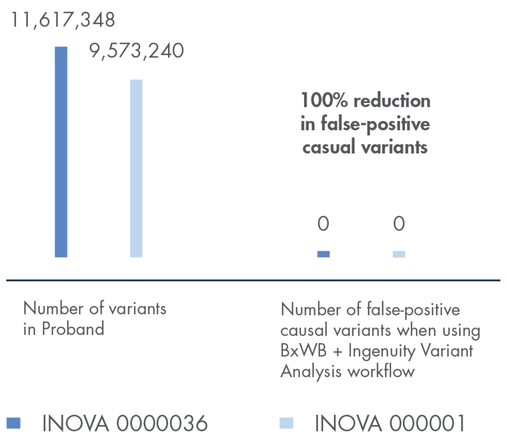


Filter cascade in Ingenuity Variant Analysis



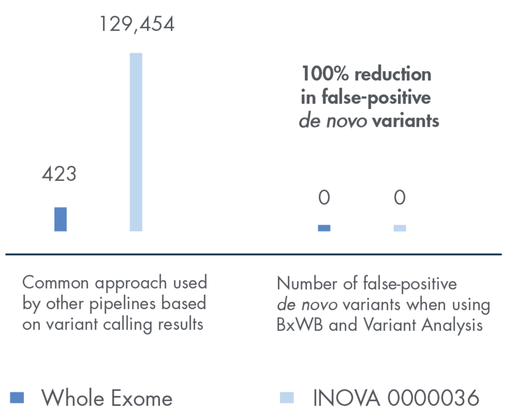
The highlighted workflows are part of the Ingenuity Variant Analysis plugin available for Biomedical Genomics Workbench. They perform an end-to-end analysis-to-interpretation workflow and were used for the study here on four whole genome trios of the INOVA Genomes and one whole exome dataset[1] for which the disease causing variants are known.

QIAGEN's hereditary disease end-to-end workflow is able to identify the disease causing *de novo* variant from a TRIO with zero false positives



In two of the whole genome TRIOs and the single exome dataset, parents were not affected and had no symptoms, meaning that a de novo variant was the disease causing mutation.

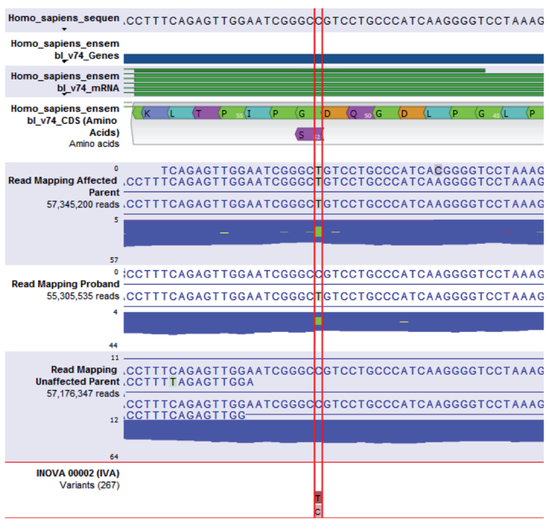
In all cases the disease causing variant could be identified. In addition, the false positive rate could be significantly reduced, which made it possible to concentrate on the most likely candidates only.

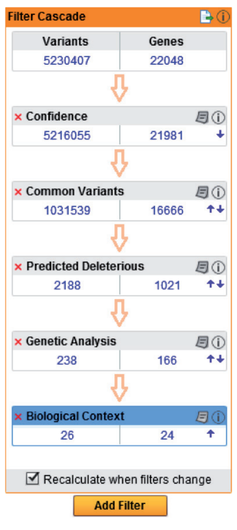


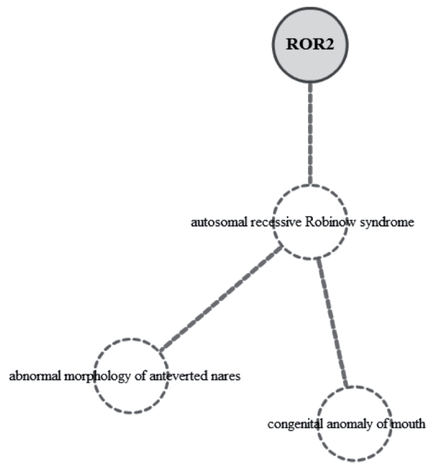
A common approach to the discovery of *de novo* variants is based on the variant call results of all family members. To identify *de novo* variants that potentially are disease causing, all variants present in mother and father are substracted from the list of variants found in the child.

Due to low read coverage and allelic dropout, many of the variants are usually not called in all individuals, which leads to a high number of false positives. The workflow with Biomedical Genomics Workbench and Ingenuity Variant Analysis removes this bias in the investigated whole exome and whole genome trio to 100%.

QIAGEN's workflow allows for 99% false positive reduction on dominant inherited cases







The result is achieved by seamlessly integrating Biomedical Genomics Workbench and Ingenuity Variant Analysis in a one step workflow. Results are shown in a Genome Browser View, enabling easy validation of a variant. Right clicking on the variant opens the filter cascade in Ingenuity Variant Analysis.

It can be optimized to include, for example, phenotype information in the analysis.

Updated results can afterwards be fetched from Ingenuity Variant Analysis. They can then again be visualized in the Genome Browser View in Biomedical Genomics Workbench. The example to the left shows the results of the analysis of the 000002 INOVA whole genome trio.

Among the 26 candidate causal variants found by the workflow, and taking into account the disease phenotype, ROR2 is the most promising one.

Summary and Discussion

In this study we analyzed four whole genome and one whole exome TRIO using QIAGEN's new hereditary disease solution. The end-to-end workflow in Biomedical Genomics Workbench includes a backcheck and optimal filter settings for Ingenuity Variant Analysis. This results in a high case solve rate and a much reduced false positive rate. We show that for diseases caused by *de novo* variants, the complete workflow results in the identification of the disease causing variant without calling any false positives. This is achieved using default parameters and providing phenotype information to the filter cascade. On dominant inherited diseases we were able to reduce the number of candidates up to 99%. In addition, we are allowing the easy validation of the candidates and an optimization of the filter cascade. As a result, less time and resources have to be spent on additional validation steps.

References

1. Rienhoff HY Jr. et al. (2013) **A mutation in TGFB3 associated with a syndrome of low muscle mass, growth retardation, distal arthrogryposis and clinical features overlapping with Marfan and Loeys-Dietz syndrome.** Am J Med Genet A. 2013 Aug;161A(8):2040-6

Sample to Insight