Increase diagnostic yield with class-leading variant prioritization

Congenica

Enabling genomic medicine

Congenica
Enabling genomic medicine

Achieve the highest diagnostic yield possible

Exomiser by Congenica

- Causal variants ranked 1st more than in any other tool
- Automates variant prioritization in all sample types from genomes to panels
- Whole genome variant prioritization in under 5-minutes

"Congenica provides a far higher variant detection rate compared to any other bioinformatic tools we've seen"

Dr Neil Morgan,
Chief Investigator, University of Birmingham

Executive summary

Next-generation sequencing (NGS) is revolutionizing the way diseases are characterized and diagnosed. However, sifting through the data often limits diagnostic yield and means many cases remain unsolved. Identifying causal variants can be like searching for a needle in a haystack, due to the sheer number of candidate variants remaining after common filtering strategies.

Diagnostic rates from whole-exome sequencing (WES) in rare pediatric disorders can vary dramatically across disease area with Cipriani et al. reporting positive diagnostic results in 17% for renal disease, 56% for non-syndromic deafness and inherited retinal disease (IRD), and 76% for ciliary dyskinesia.[1]

To interpret exome data for rare disease diagnostics, geneticists must identify the one or two mutations responsible for a patient's condition hidden amongst 40,000 variants. This process can take between 20 and 40 hours and requires highly skilled staff using specialized software – placing ever-increasing strain on the resources of diagnostic laboratories as they try to keep up with growing throughput demands, resulting in backlogs of cases and extended diagnostic odysseys.

Exomiser – a class-leading variant prioritization tool - greatly speeds up this process and enables geneticists to make faster, more accurate diagnoses and achieve a consistently higher [30-60%] diagnostic yield than industry averages and alternative solutions across all diseases, sample types and family structures.[2,3] This means increased throughput of cases and better analysis outcomes from NGS data.

Exomiser has been seamlessly integrated into the Congenica platform to provide optimized access to this powerful variant prioritization tool within an end-to-end automated clinical decision support platform for rapid scalable secondary and tertiary analysis of all next-generation sequenced data.

Foreword by Professor Damian Smedley

Damian Smedley is a Professor at Queen Mary University of London and until recently was Director of Genomic Interpretation; Associate Director of Bioinformatics at Genomics England. Together with Prof. Peter Robinson of the Jackson Laboratory, he led the creation of Exomiser and his pioneering research is helping to make genomic medicine a reality.

Exomiser was created back in 2014 through a collaboration between my team, then based at the Sanger Institute, and Professor Peter Robinson's. We had already shown through research studies that systematic collection of clinical data using the Human Phenotype Ontology and comparison of patient phenotypes to known disease and model organism phenotypes could accurately identify disease genes. Around this time, scientists were beginning to use newly developed next-generation sequencing technologies to try and discover the genetic cause of rare disease cases.

Landmark successes were published but talking to most scientists revealed they were struggling with most cases due to the overwhelming number of candidate variants. After some promising bespoke analyses, Peter and I decided to create an integrated tool that could filter a patient's exome for the most likely candidate variants and then prioritize these based on, not only their predicted pathogenicity, but also using our phenotype-comparison algorithms.

In our first paper we proved that Exomiser could reliably identify known diagnoses and we went on to work with the NIH Undiagnosed Disease Project to help them solve their particularly challenging cases. In the last few years, I have worked with Genomics England to ensure Exomiser has been a successful part of the 100,000 Genomes Project and has been applied to all 37k rare disease cases interpreted to date.

Our whole rationale for creating Exomiser has been to translate our academic work into a tool that can be used by researchers, diagnostic labs, rare disease platforms and companies. This approach has seen world-wide adoption and numerous, published reports of its use and has been one of the most satisfying achievements of my career. This success has driven us to continuously improve Exomiser over the years and keep its underlying databases up to date, which is not the case for most academic software.

In the clinical field, users increasingly require software to be used in a robust, accredited manner and the results presented back in easy to use, decision support tools and we partner with Congenica to achieve this widened access to gold-standard variant prioritization enabling rapid, reliable and repeatable genomic interpretation at scale.

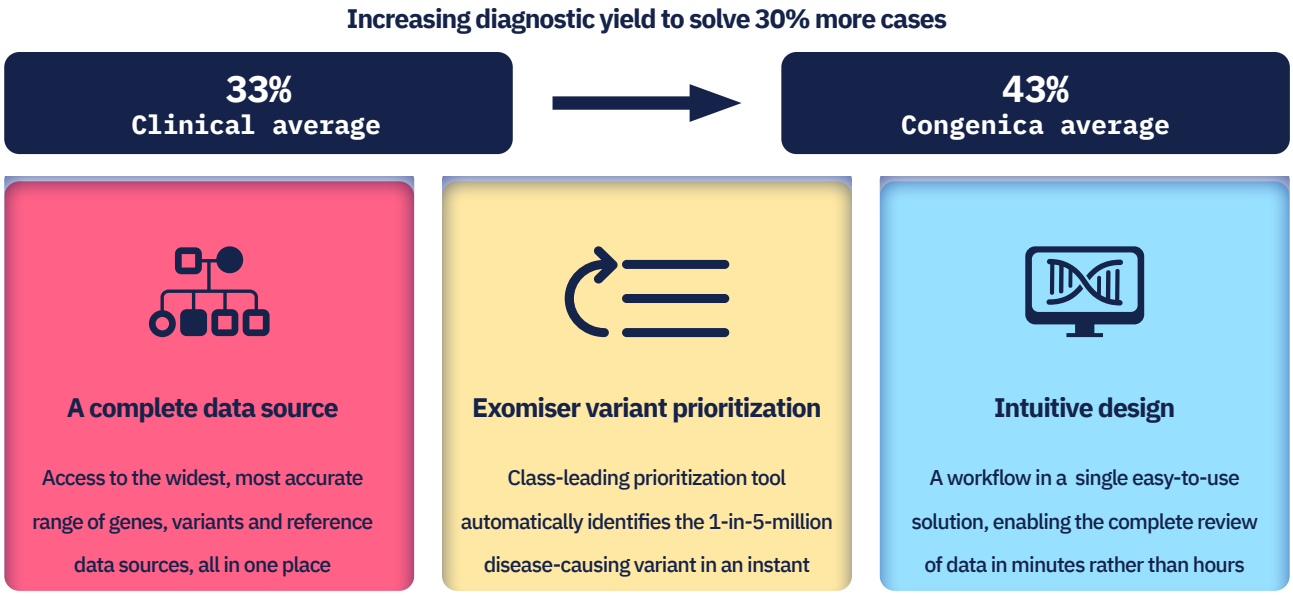
Challenges in analyzing genomic data

Numerous disorders remain unsolved after undergoing next-generation sequencing for different reasons. Often, many variants are found, their interpretation is extremely time-consuming and difficult and further complicated by incomplete knowledge on gene functions. Also, using very low frequency cutoffs may remove clinically relevant variants and lead to false negative findings, especially in hard to solve cases.[1]

These challenges can be overcome with Exomiser – a program developed as a collaboration between members of the Monarch Initiative [9] that enables geneticists to identify causal variants faster, with greater accuracy and confidence.

Exomiser performs an integrative analysis of sequence data and encodes genotypic variants to their phenotypes, encoding the variants with Human Phenotype Ontology (HPO) terms.[9] It prioritizes variants by leveraging information on variant frequency, predicted pathogenicity, and gene-phenotype associations derived from human diseases, model organisms, and protein–protein interactions.

Congenica with Exomiser has been demonstrated to support the greatest opportunity for identifying actionable insights, achieving a 30% higher diagnostic yield than industry averages across all diseases, sample types and family structures, based on an independent review of 171 publications and more than 25,000 complex rare disease cases analyzed with Congenica.[2,3]



The average diagnostic yield across disease samples is 33%, whereas in a study of 25,000 complex rare disease cases Congenica software enabled teams to achieve an average diagnostic yield of 43% across all sample types, representing, on average, a 30% increase in complex rare disease cases that can be solved.[2,3]

In Congenica, all rare disease samples including panels, exomes and genomes, go through the Exomiser automated variant prioritization framework. Exomiser annotates the consequence of variants (based on Ensembl transcripts) and then filters and prioritizes them for how likely they are to be causative of the proband's disease based on user-defined criteria:

- the predicted pathogenicity and allele frequency of the variant in reference databases
- how closely the proband's phenotypes match the known phenotypes of diseases and model organisms associated with the gene.[4,5]

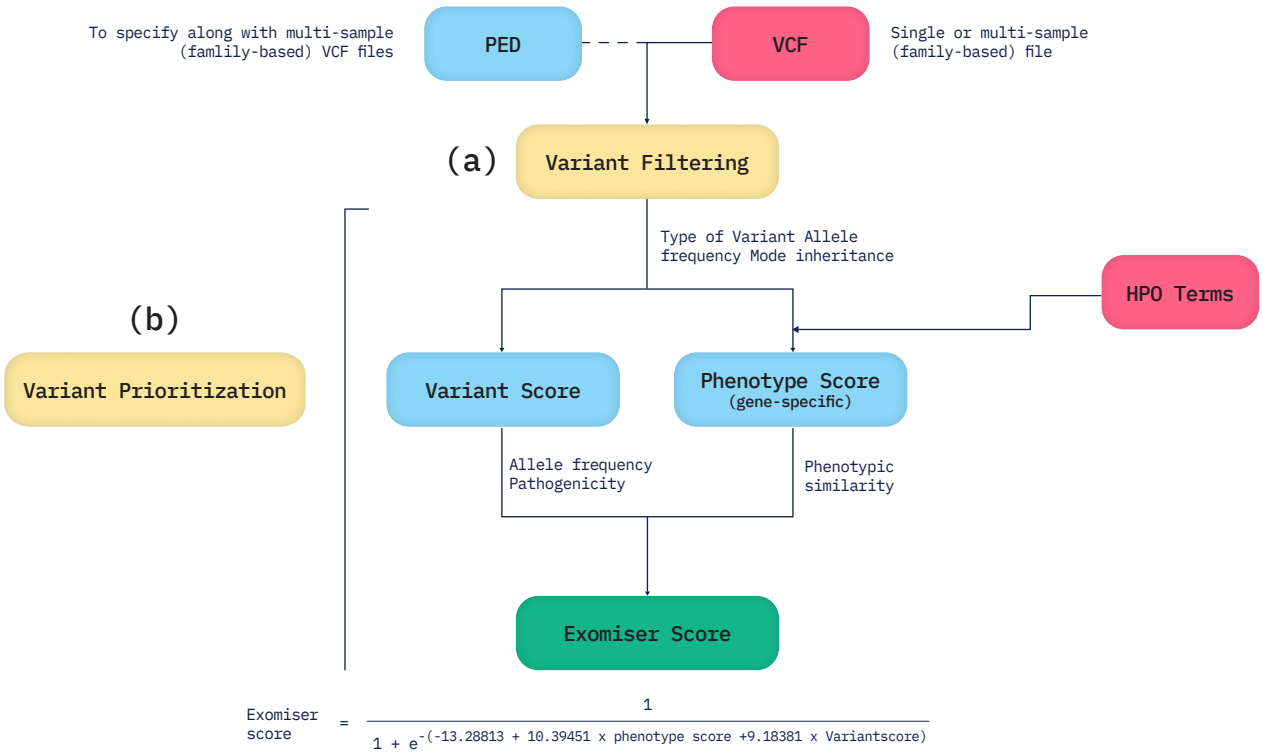


Figure 1: Overview of Exomiser workflow analysis. [1]

Exomiser performance: clinical evidence and independent validation

Numerous peer reviewed articles highlighting clinical evidence provide independent validation of the benefits of using Exomiser.[1,5,6] Early published releases of Exomiser were able to prioritize disease-causative variants as top candidates in up to 97% of simulated whole-exomes.

Cipriani et al. assessed the performance using a set of 134 whole-exomes from cases with a range of rare retinal diseases and known molecular diagnosis. Using default settings, Exomiser correctly ranked the causal variants as the top candidate in 74% of the dataset and top five in 94%.[1]

A study by Thompson et al. compared methods for incorporating phenotype into the interpretation process and assessed the extent to which phenotypic annotation aids prioritization of the correct variant. Using a cohort of 29 patients with congenital myasthenic syndromes with causative variants in known or newly discovered disease genes, exome data and the Human Phenotype Ontology (HPO) coded phenotypic profiles, it was shown that gene list filters created from phenotypic annotations perform similarly to curated disease gene virtual panels. Candidate variants were ranked with Exomiser while varying phenotypic annotation. Analyzing 3,712 combinations, the study showed that increasing phenotypic annotation improved prioritization of the causative variant, from 62% ranked first on variant alone to 90% with seven HPO annotations.[5]

Ji et al. implemented a semi-automated and phenotype-driven WES diagnostic workflow, incorporating both the DRAGEN pipeline and the Exomiser variant prioritization tool, at an academic children's hospital with an ethnically diverse pediatric patient population. A 41% molecular diagnostic rate for 66 duo-, quad-, or trio-WES cases, and 28% for 40 singleton-WES cases was achieved.[6]

Exomiser's code is extensively tested and built automatically from the GitHub source using a public continuous integration server. Code quality is checked using static code analysis tools and the release candidates undergo full system tests against a large number of gold-standard rare disease cases where the output is checked against previous release versions and diagnoses.

Preliminary validation

The primary diagnostic analysis for the Genomics England 100,000 Genomes Project aims to report back to participants variants with sufficient evidence for diagnostic reporting related to their primary condition. The Genomics England pipeline aims to facilitate this by annotating a shortlist of variants that are likely or plausibly disease causing for assessment by NHS staff.[8]

Genomic England's Exomiser pipeline has been validated on 62, randomly selected, 100,000 Genomes Project cases with a positive diagnosis from the NHS (50 GrCh37 and 12 GrCh38). The variant(s) reported as diagnostic by the NHS were correctly returned as the top ranked candidate(s) in 44/62 (71%) of cases (sensitivity = 0.71, precision = 0.71) and in the top five for 57/62 (92%) of cases (sensitivity=0.92, precision=0.18). [8]

Exomiser offers a complementary approach to the panel-based, tiering pipeline as shown by an analysis of ~200 clinically solved cases. 72% of the diagnoses were identified in the applied gene panels by the tiering pipeline with high precision (one to two candidates per case). Exomiser identified 81% of the causal variants in its top five ranked results. Combining the tiering and Exomiser results leads to an increased recall of 90% of the causal variants compared to using either approach alone, with a precision of 0.17, meaning an average of five to six variants are presented for consideration.[8]

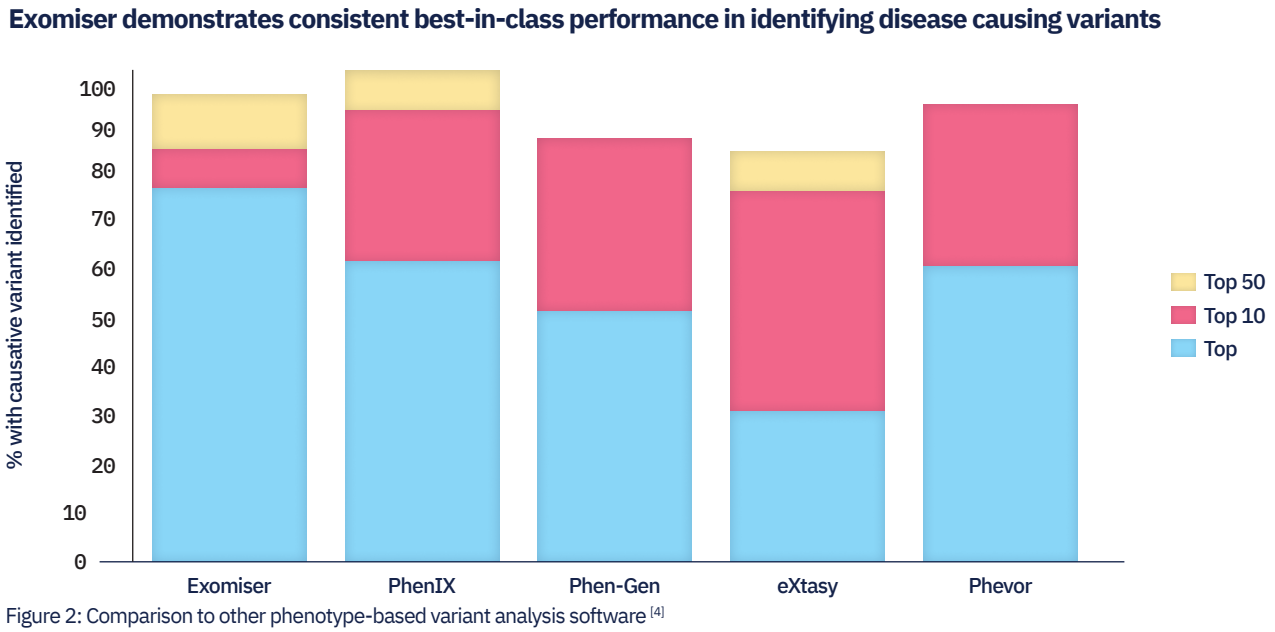


Figure 2: Comparison to other phenotype-based variant analysis software [4]

A number of other exome analysis tools that utilize HPO annotations of the patient have been produced. A review of these from 2015[4] showed that Exomiser performed significantly better than all other tested phenotype-based variant analysis software. In this benchmarking the authors spiked in known disease variants to exomes and recorded how often each software identified this variant as the top hit (blue) or in the top 10 or 50 (red, yellow), as illustrated in Figure 2. To simulate the typical scenario in clinical sequencing projects where it is a struggle to collect phenotype data, the researchers only used three randomly selected disease phenotypes as input and added imprecision by making two of them more general terms and added noises with two completely random terms.

Unfortunately, running a truly informative comparison on the most recent software, that would help users come to decisions, is extremely challenging. The majority of the academic solutions are either not maintained post publication, suffer from broken installation documentation, or are complicated to use so others end up running them with sub-optimal settings. Commercial solutions such as Moon and Fabric have licensing terms that prohibit publishing comparisons without their permission. However, the biggest impediment is the lack of access to a large and diverse set of unbiased, blinded real-life samples and an independent review mechanism. The community is well aware of these issues and are coming together to generate a DREAM challenge that will hopefully resolve this.

The South West Thames Centre for Genomics has reviewed whole genome sequencing data for over 1,500 patients with rare disease recruited to the 100,000 genomes project. We have extensively used the Exomiser tool to prioritize variant review. This tool has assisted us to identify disease causing variants in genes not covered by the disease gene panel approach. Further to this, follow up of Exomiser prioritized variants has also assisted us in new disease gene discovery by highlighting candidate variants on the basis of disease in model organisms or gene-gene interactions.

Dr. Meriel McEntagart
Consultant Clinical Geneticist and Joint Care Group Lead

NHS
St George's University Hospitals
NHS Foundation Trust

Impact on genetic data analysis and interpretation

Exomiser enables positive variant identification that otherwise would have been missed.

The Exomiser demo web-application receives approximately 3000 hits a year. It is used as a teaching aid for bioinformatics students at the Queen Mary University of London in the United Kingdom and an internal version is hosted for the clinicians at the Charité Hospital in Berlin. The command-line application is used by hospitals and large-scale national rare-disease projects such as the NIH UDP, NIH UDN and is a component of the European Joint Programme for Rare Disease (EJPRD). It is acknowledged as a 'Recognized Resource' by the International Rare Diseases Research Consortium (IRDiRC) and is also a component of two other IRDiRC recognized resources.[7]

The core library has been used to incorporate the phenotype matching capabilities into other rare-disease resources such as Phenopolis and Matchbox as part of the Matchmaker Exchange network. It has also been incorporated as a service within the UK 100,000 Genomes Project where it is run on all the enrolled rare-disease patients and the results incorporated into the clinical diagnosis tool.[8]

Exomiser's role within the Congenica platform

Exomiser is seamlessly integrated into the Congenica platform, using optimal settings for diagnostic and discovery performance based on the advice of the Exomiser development team. Exomiser complements the variety of other secondary and tertiary analysis tools and databases incorporated into Congenica's platform to enable end-to-end genomic data interpretation and analysis from sequencer to report. Among these are Mastermind® – a key reference data source with approximately 6 million variants published to instantly identify relevant clinical evidence in published literature – and DECIPHER, which enhances genomic interpretation by retrieving information from a variety of bioinformatics resources relevant to the variant found in the patient. Data on DECIPHER includes over 37,000 curated SNVs (single nucleotide variants) and CNVs (copy number variants). All data is displayed in Congenica' s own genome browser.

Congenica's unique integration of multiple tools and datasets into a single scalable solution means that you no longer need to use multiple resources to perform sequence alignment, variant calling, prioritization, interpretation, classification and reporting, saving you time and ensuring that no important information is missed to maximize efficiency and diagnostic yield.

Conclusion

Achieving efficient interpretation of NGS data with higher diagnostic yields is critical to relieve backlogs of cases, increase sample throughput and improve outcomes to make a difference to peoples' lives. The original benchmarking of Exomiser performed on 10,000 simulated rare disease whole-exomes showed that up to 97% of the data received the top scoring hit for the correct causative variants. Clinical evidence and independent validation of Exomiser in numerous peer-review publications demonstrate Exomiser's ability to enable higher throughput and increased diagnostic yield.

Exomiser is a class-leading tool to guide and accelerate the typically complex and time-consuming, multidisciplinary work required to prioritize causal variants, increasing diagnostic yield even in the most complex cases. Its integration into the Congenica platform helps clinical geneticists to increase data analysis and interpretation efficiency while enhancing diagnostic yield and case throughput. In turn, the ability to perform faster and more insightful analysis at scale will ultimately support health providers to help more people receive life-changing answers to previously unsolvable complex genetic questions.

Demo Congenica software to see how class-leading variant prioritization will increase your diagnostic yield, confidence and efficiency.

Demo online at www.congenica.com/demo

References

[1]
Cipriani et al., An Improved Phenotype-Driven Tool for Rare Mendelian Variant Prioritization: Benchmarking Exomiser on Real Patient Whole-Exome Data. Genes 2020, 11, 460; doi:10.3390/genes11040460

[2]
Yang et al., Genetic aetiology of early infant deaths in a neonatal intensive care unit. Journal of Medical Genetics (2019) 0, 1 – 9; dx.doi.org/10.1136/jmedgenet-2019-106221
Smedley D, Jacobsen JO, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004W2015. doi:10.1038/nprot.2015.124

[3]
Smith et al., Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. Genet Med. Genetics in Medicine (2019) 21: 3–16. doi:10.1038/s41436-018-0024-6.

[4]
Smedley D, Robinson PN. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. Genome Med. 2015;7(1):81. doi:10.1186/s13073-015-0199-2

[5]
Thompson et al., Increasing phenotypic annotation improves the diagnostic rate of exome sequencing in a rare neuromuscular disorder. Human Mutation, Volume40, Issue10, October 2019, Pages 1797-1812 doi.org/10.1002/humu.23792

[6]
Ji et al., A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants. Cold Spring Harb Mol Case Stud 5: a003756 Published by Cold Spring Harbor Laboratory Press. Published in Advance February 12, 2019, doi:10.1101/mcs.a003756

[7]
https://phenomecentral.org/

[8]
https://cnfl.extge.co.uk/pages/viewpage.action?pageId=128355594

[9]
Shefchek et al., The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species . Nucleic Acids Res. 2020 Jan 8;48(D1):D704-D715. doi:10.1093/nar/gkz997.

[10]
Köhler et al., Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources, Nucleic Acids Res. 2019 Jan 8;47(D1):D1018-D1027. doi:10.1093/nar/gky1105.