Increased Yield of Clinically Relevant Candidates in the UK 100,000 Genomes Project Using Opal™ Clinical for Hereditary Disease

Melanie Babcock[1], Charlene Son Rigby[1], Marco Falcioni[1], Andrew Guo[1], Anna Lewis[1], Eric Kofman[1], Anthony P. Fejes[1], Josh Grigonis[1], Edward S. Kiruluta[1], Mark Yandell[2], and Martin G. Reese[1]
1. mbabcock@fabricgenomics.com, Fabric Genomics™ Inc., Oakland, CA; 2. Department of Human Genetics, University of Utah, Salt Lake City, UT

ABSTRACT

The 100,000 Genomes Project, spearheaded by Genomics England (GeL), is a United Kingdom National Health Service sponsored study aimed at identifying disease-causing genetic variants in patients and families with rare genetic diseases and cancer using a whole genome sequencing (WGS) approach. For this study, clinical history was used to recruit patients into specific disease categories, each of which were associated with gene panels curated in the GeL PanelApp tool.

Fabric Genomics™, a clinical interpretation partner for the 100,000 Genomes Project, has analyzed over 600 clinical cases using Opal™ Clinical. The variant filtering and prioritization protocols utilized for case analysis include GeL's variant tiering methodology, ClinVar, and Fabric Genomics' proprietary variant and gene ranking algorithms VAAST (Variant Annotation, Analysis and Selection Tool) and Phevor (Phenotype Driven Variant Ontological Re-ranking Tool). We report results showing that by applying VAAST and Phevor we increase the clinical candidate yield compared to using the GeL tiering system alone. We identified candidate causal genes/variants in 44.7% of the cases. In 19.6% of these cases (8.7% overall) candidates were only obtained by using the VAAST/Phevor top 20 ranked genes/variants.

Interpretation Methodology

We have implemented a whole genome interpretation methodology within Opal™ Clinical that is comprehensive and rapid.

The workflow developed for GeL includes consideration of the following variants:

- Variants scored under different inheritance modes in the top 20 genes ranked by VAAST and Phevor, two sequentially applied proprietary algorithms: VAAST integrates sequence conservation, genetic consequence, and allele frequency in a probabilistic framework to identify disease-causing alleles, Phevor then combines HPO-based patient phenotype descriptions with the VAAST results to re-rank the variants
- Variants in phenotype matched panels curated in the GeL PanelApp tool and categorized using GeL's tiering methodology (https://bioinfo.extge.co.uk/crowdsourcing/PanelApp/): known pathogenic, protein truncating, and de novo protein alternating variants (Tier 1), and other protein altering variants (Tier 2)
- Variants with at least one ClinVar Pathogenic or Likely Pathogenic submission

The interpretation methodology concludes with an expert review by Fabric Genomics' Clinical Services team. Each case was interpreted by two primary reviewers. Identified candidates were scored using the ACMG variant interpretation guidelines.

The candidate genes/variants are further evaluated by the NHS Genomic Medicine Centre laboratories and these labs will determine which of those candidates are reported back to the patients.

GeL workflow for receiving interpretation requests, processing samples and returning clinical reports to Genomic Medicine Centres

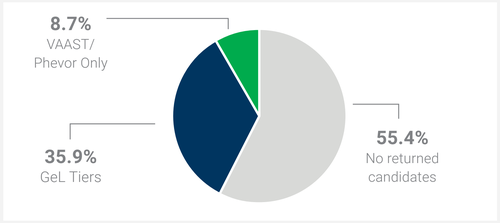
Fabric Genomics' Gene Ranking Algorithms: VAAST[1] and Phevor[2]

VAAST is an algorithm that was developed in collaboration with the University of Utah. Using VAAST for gene prioritization speeds diagnosis and improves diagnostic yield by providing a ranking of genes based on their likelihood to cause disease. Every variant is assessed for comparative functional impact on the protein product, conservation of the position across species, and the allele frequency.

Phevor re-ranks genes that have already been prioritized by VAAST by using the Human Phenotype Ontology (HPO) terms provided for the proband. Phevor starts by mapping phenotype terms to the Human Phenotype Ontology, Gene Ontology and other ontologies, then uses a unique network propagation approach to identify additional gene candidates. This process creates a ranked list of genes ordered by the specific phenotype provided. Phevor then combines this prioritized list of genes with the VAAST analysis to produce a combined ranking of candidate genes based on deleteriousness and the specific phenotype or phenotypes in question. This re-ranking allows rapid discovery of clinically relevant variants in genes related to the proband's phenotype, including genes not directly annotated to that phenotype.

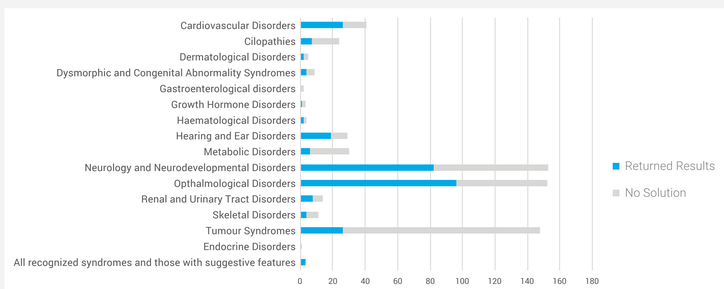
Return of Results in 44.7% of 609 cases

Of the 609 cases, we returned 44.7% with casual candidate genes/variants, 35.9% of those cases returned with GeL tiering and 8.7% cases exclusively from VAAST/Phevor top 20 (19.6% additional yield).



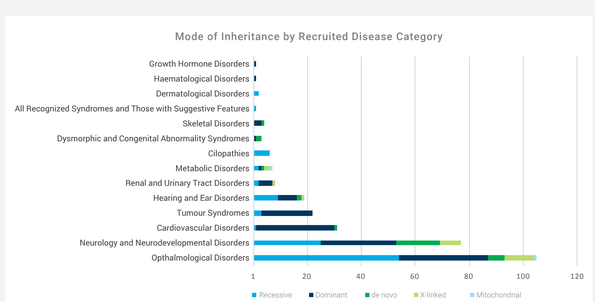
Return of Results by Recruited Disease Category

Recruited Disease Categories are defined by the recruiting Genomic Medicine Centres (GMC). The rate of return of results was highest for Cardiovascular Disorders, Opthalmological Disorders, and Hearing and Ear Disorders. Of those recruited disease categories with >5 cases, Tumor Syndromes had the lowest return of results.



Variants Returned Across all Major Modes of Inheritance

Mode of Inheritance for returned results. The workflow identified variants across all major modes of inheritance, with Dominant and Recessive modes being the most common



Example cases where clinically relevant candidates were only found using VAAST/Phevor

Main Recruited Disease Category	HPO/Phenotype	Panel applied	Gene	Tier status	VAAST/ Phevor rank	ACMG class
Haematological Disorders	Recurrent respiratory infections	A - or hypo-gammaglobulinaemia	CTPS1	Not in the panel applied (present in another panel)	4/4	Pathogenic
Neurology and Neurodevelopmental Disorders	Delayed speech and language development, café-au-lait spots, neurofibromas, ID, seizures, generalized hypotonia, amelogenesis imperfecta, delayed gross motor development, delayed fine motor development	Intellectual Disability	NF1	In the panel but was inherited from "unaffected" mother so not tiered	8/1	Likely Pathogenic
Cardiovascular Disorders	Abnormality of the nervous system, autistic behavior, joint hypermobility, abnormal autonomic nervous system physiology	Ehlers-Danlos Syndrome type 3	SETBP1	Not in the panel applied (present in another panel)	1/1	Pathogenic
Neurology and Neurodevelopmental Disorders	Seizures, generalized hypotonia, absence seizures, EEG abnormality, limb hypertonia, hypsarrhythmia, infantile encephalopathy, intellectual disability, severe, epileptic spasms, infantile spasms, abnormality of movement, subdural hemorrhage	Epileptic Encephalopathy	STXBP1	Not tiered; Not in any panel	1/1	Pathogenic

CONCLUSIONS

Opal™ Clinical has provided GeL with potential causative candidates in 44.7% of cases. In the 609 cases we present here, GeL's tiering system achieves a results return rate of 35.9%. By using our VAAST/Phevor algorithms we were able to increase the yield of candidate genes/variants by 19.6%, thus highlighting the complementary utility of the VAAST and Phevor algorithms and GeL's tier filtering methodology.

Fabric Genomics™ supports labs to maximize their diagnostic yield. The VAAST and Phevor ranking algorithms accelerate identification of disease-causing candidates. Opal™ Clinical is optimized to provide efficiency for variant scientist's time: the cases reported here each took less than three hours; however there are cases that took significantly less time. This time included both identification of candidate variants, and scoring using the ACMG 2015 Variant Interpretation Guidelines.

This makes Opal™ Clinical ideal for hard to solve rare genetic disease cases, and for large scale country projects.

1. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. Rope et al. Am J Hum Genet. 2011 Jul 15;89(1):28-43.

2. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. Singleton et al. Am J Hum Genet. 2014 Apr 3;94(4):599-610

FABRIC GENOMICS™

FORMERLY omicia

Learn more at

www.fabricgenomics.com ■ info@fabricgenomics.com ■ 510.595.0800