



Ingenuity Variant Analysis, leveraging the Knowledge Base and HGMD, achieves over 30x enrichment in biologically relevant variants from whole genome and exome sequence data from patients with rare disease

S. Shah¹, A. Athvaale², K. Boycott², J. Devaney³, G. Eley⁴, R. Felciano¹, S.E. Hofherr³, A. Joecker¹, K. Kernohan², A. Krämer¹, B.W. Meltzer³, A. Muthiah¹, K. Patel¹, D. Richards¹, M.B. Seprish³, B. Solomon⁴, A. Subramanian¹, J.G. Vockley⁴, R. Yip¹

Sample to Insight

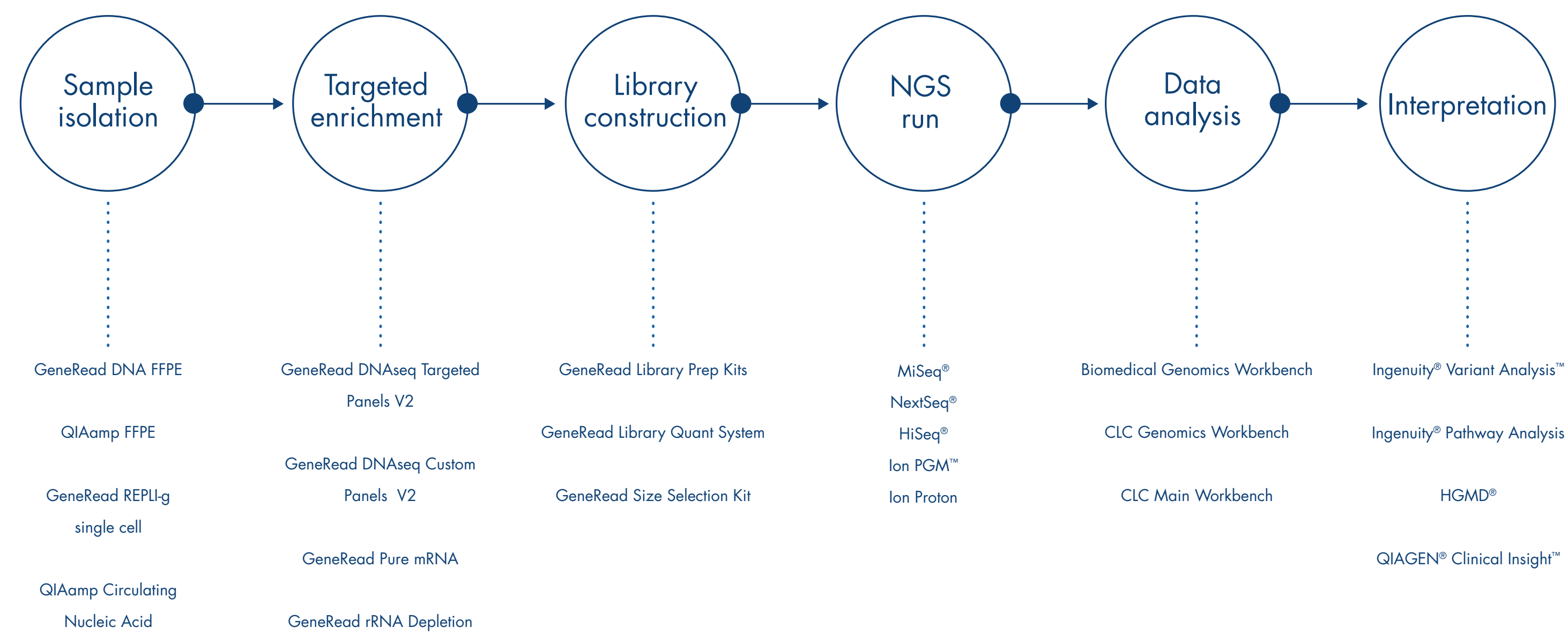
Automation solutions for seamless and cost-effective workflows, featuring:

- **Industry's most reliable sample technologies** – Top-quality assays and panels to accurately analyze and identify diseases and genetic variations.
- **Bioinformatics software and curated knowledge bases** – To transform your raw data into relevant, actionable findings.

QIAGEN Ingenuity Variant Analysis

Identify causal variants from human sequencing data in just minutes.

- **Intuitive, user-friendly interface** – No bioinformatics skills needed.
 - **Interactive Filter Cascade** – Rapidly eliminate common and non-deleterious variants with a basic set of filters.
 - **Iterative analysis** – Apply a hypothesis, visualize and evaluate results in real-time.
- **Knowledge driven algorithms and analytics** – Genetics, Statistics, Functional Prediction.
 - **Curated up to date content at your fingertips** – Utilize the Ingenuity Knowledge Base™ containing millions of biomedical findings and mutations, accurate, up-to-date and curated by experts, from the literature and public databases.
 - **Causal Network Analytics** – Identify variant in genes within 1- or 2- network “hops” of upstream or downstream mutated genes.
- **Sharing and Publication Tools** – Easily collaborate with colleagues and peers. Export data and graphics to aid in manuscript preparation and publications.
- **Scalable processing capacity** – Analyze thousands of samples simultaneously.



QIAGEN Ingenuity Variant Analysis

The interface displays a list of variants with columns for Variant ID, Gene, Chr., Position, Strand, Gene Symbol, Transcrit., Protein Variant, Case Samples, Control Samples, Translation Impact, SIFT Predicted, and Variant First. A network diagram shows the relationship between genes. Findings include citations such as "Mutations in NOTCH1 Cause Adams-Oliver Syndrome" and "Mutant human NOTCH1 gene (germline c.4487G>A) increases the risk of Adams-Oliver syndrome in human (HGMD: DM)." and "Mutant human NOTCH1 gene (c.4487G>A) increases Adams-Oliver syndrome in human."

Single Sample Analysis

Total number of samples: 80

Source	Samples
Inova Health Systems, Virginia	15
Children's National Medical Center, Washington, DC	28
Children's Hospital of Eastern Ontario	33
Legacy Ingenuity Demo cases	5

Filter Cascade:

Confidence

The Confidence filter settings include:

- Keep only variants which satisfy all of these criteria:
- Call quality is at least 20 in any case or at least 20 in any control
- Variant passed upstream pipeline filtering
- Read depth is at least 10 in any sample
- Allele fraction is at least 5 in any sample
- Outside top 5% most exonic variable 100base windows in healthy public genomes
- Outside top 1% most exonic variable genes in healthy public genomes (1000 Genomes)

Common Variants

The Common Variants filter settings include:

- Keep rare variants
- Exclude variants that are observed in any of these populations with an allele frequency of:
 - at least 0.5% in the 1000 Genomes Project
 - at least 0.5% of all NHLBI ESP exomes
 - at least 0.5% in the Allele Frequency Community (includes ExAC and CGI)
- are present in dbSNP

Predicted Deleterious

The Predicted Deleterious filter settings include:

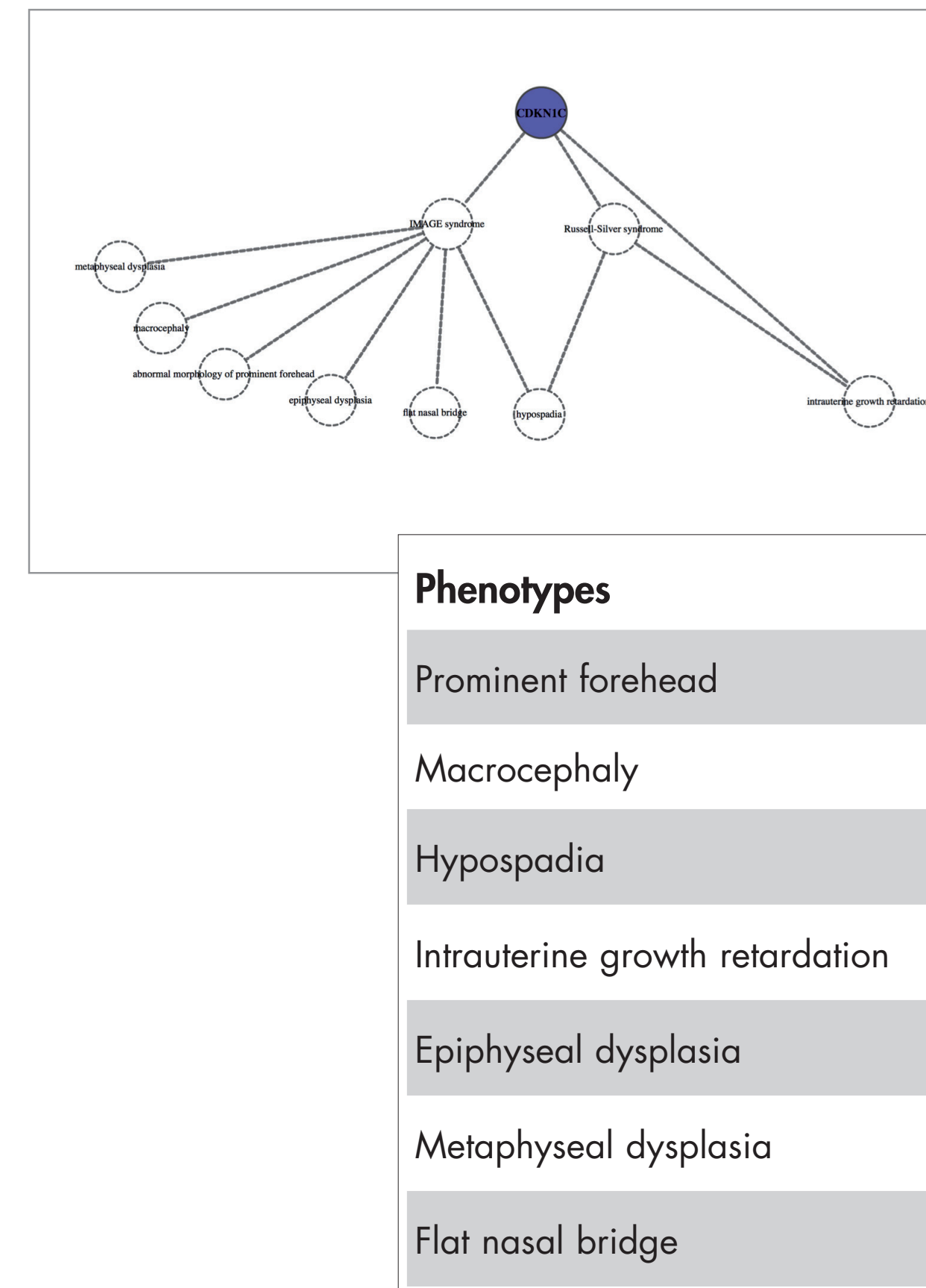
- Keep only variants that are experimentally observed to be associated with a phenotype: Disease-associated according to computed ACMG Guidelines classification
- Pathogenic
- Likely Pathogenic
- Uncertain Significance
- Likely Benign
- Benign
- Listed in HGMD®
- are associated with gain of function of a gene:
 - Established in the Literature
 - Gene Fusion
 - Inferred activating mutation by Ingenuity
 - Predicted gain of function by BSIFT
 - microRNA Binding Site
 - Copy Number Gain
- are associated with loss of function of a gene:
 - Frameshift, in-frame indel, or start/stop codon change
 - Missense unless predicted tolerated by SIFT or PolyPhen-2
 - Nullzygous
 - Splice site loss up to 2 bases into intron or as predicted by MaxEntScan
 - Deleterious to a microRNA
 - Structural Variant
 - Promoter Loss with ENCODE TFBS
 - Enhancer

Biological Context (Example)

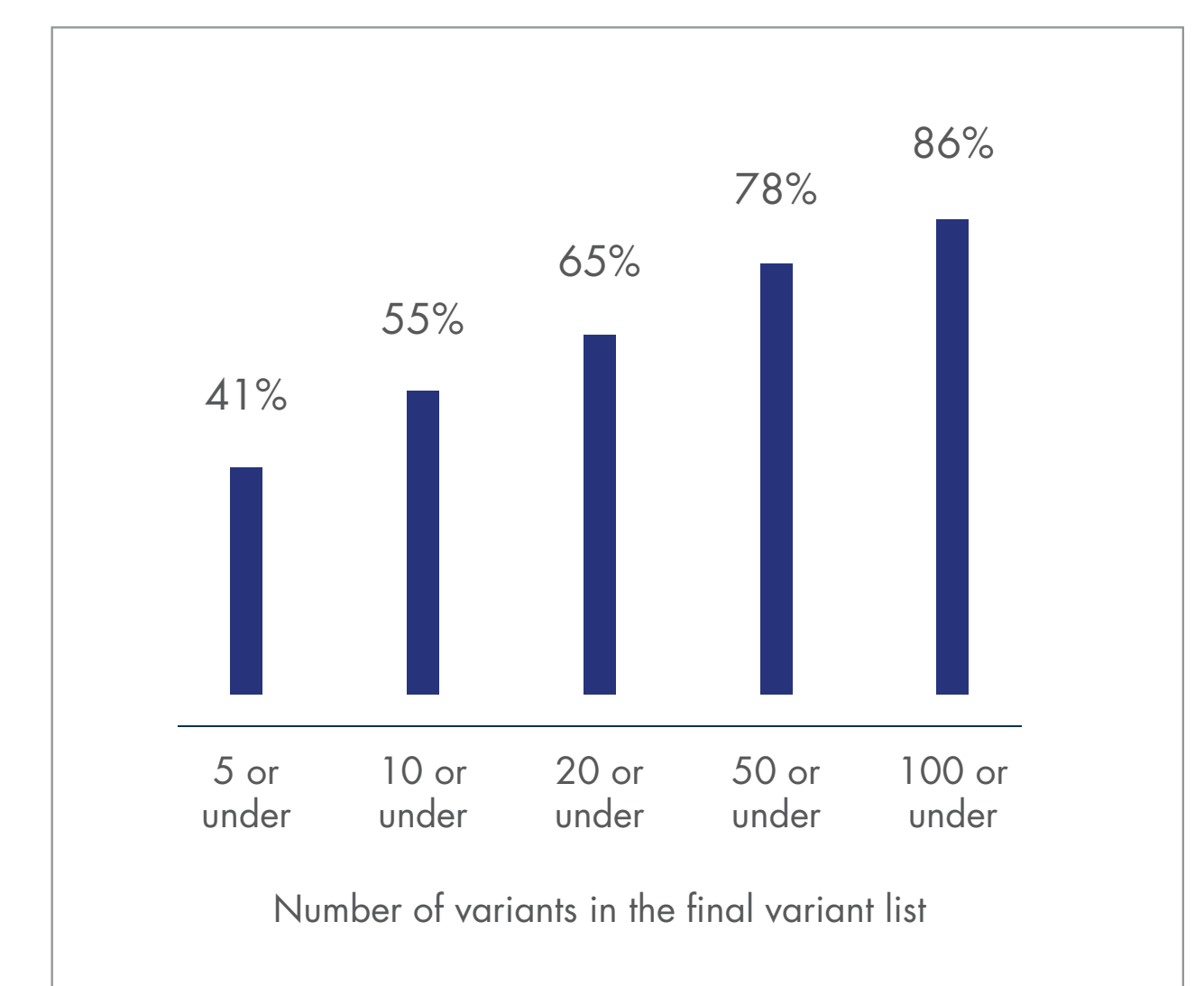
The Biological Context filter settings include:

- Keep only variants
- within 1 hop upstream
- that are known or predicted to affect:
 - macrocephaly [disease]
 - hypospadias [disease]
 - intrauterine growth retardation [disease]
 - epiphyseal dysplasia [disease]
 - metaphyseal dysplasia [disease]
- and genes within 1 hop downstream of above
- include diseases consistent with the phenotypes above

Path-to-Phenotype



Number of cases solved:



Summary:

- All 80 cases were solved using Ingenuity Variant Analysis, by re-adjusting the pre-configured filter settings. The causal variant was selected based on the number of phenotypes linked to the gene and/or consistent with known syndromes.
- In 86% cases, we were able to narrow the list of variants to <100, with a 30x enrichment in variants linked to one or more phenotypes.
- In 5% cases, the reported causal variant was >0.5% MAF in the population.

Author affiliation: 1) QIAGEN Bioinformatics, 1700 Seaport Blvd, Third Floor, Redwood City, CA, 94063, USA; 2) Children's Hospital of Eastern Ontario (CHEO), 401 Smyth Road, Ottawa, Ontario K1H 8L1; 3) Children's National Medical Center, 111 Michigan Ave NW #800, Washington, DC 20010; 4) Inova Translational Medicine Institute, 8110 Gatehouse Road, Falls Church, VA, 22042, USA