Opening the floodgates to get a sip of water: Challenges of whole exome sequencing analysis as a diagnostic tool

N.T. Strande[1,2], C. Bizon[3], J.K. Booker[1,2], A. Brandt[1], A.K.M. Foreman[1], I. King[2], K. Lee[1], M. Li[2], L. Milko[1], J.M. O'Daniel[1], P. Owen[3], B.C. Powell[1], B.A. Seifert[1], D. Young[3], K.C. Wilhelmsen[3], J.P. Evans[1], J.S. Berg[1], K.E. Weck[1,2]

[1]*Dept. Genetics, UNC-Chapel Hill,* [2]*Dept. Pathology & Laboratory Medicine, UNC-Chapel Hill,* [3]*Renaissance Computing Institute, Chapel Hill, NC*

Introduction

North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing

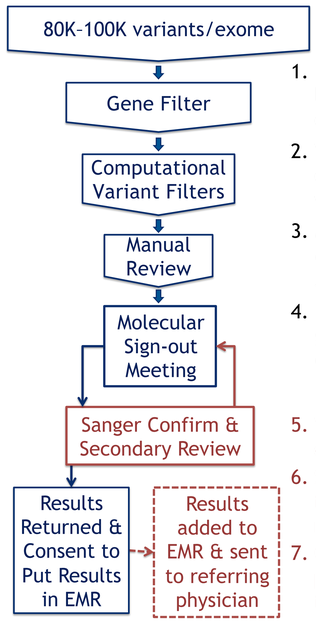
NCGENES is a research study evaluating whole exome sequencing (WES) as a diagnostic tool in a diverse group of patients with conditions likely to have a genetic etiology, but have evaded diagnosis by traditional methods.

The study aims to answer the following questions:

- Who is the appropriate patient population for (WES)?
- What conditions should be considered for WES analysis?
- What is the most efficient & accurate WES analysis?
- How should incidental or secondary findings be managed?
- What is an acceptable level of uncertainty in the results for patients/clinicians?

Methods

NCGENES Workflow For Exome Analysis



Methods to minimize analysis time

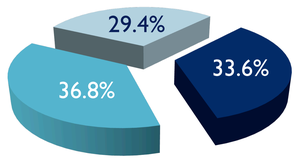
1. *A priori* Diagnostic Gene Lists: Analysis is limited to broadly designed gene lists consistent with the participant's disorder.
2. Variants are prioritized by type & effect on protein. Population variants & those with poor quality are filtered out.
3. Manual review of literature, variant databases, allele frequency databases, *in silico* pathogenicity predictors, etc.
4. Results are discussed weekly by diverse group of clinical geneticists, genetic counselors, clinicians, fellows, etc.

Workflow in CLIA Laboratory

5. Variants meeting our reporting criteria are confirmed via Sanger sequencing.
6. Secondary Variant Review (more thorough): If review alters interpretation, results are discussed at group meeting.
7. Clinical geneticists return results to participants & obtain consent for results to go in EMR (optional).

Participant Demographics

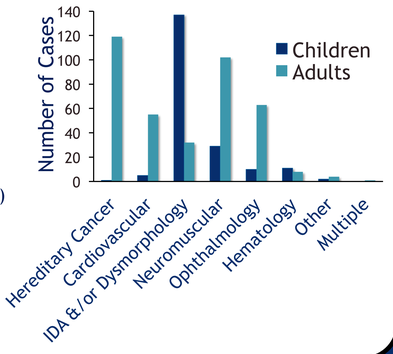
Age of Study Population



Total # of Participants: ~575

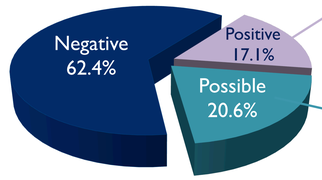
Participant age range: 0-84 years

Participant average age: 32.8 years



Results: First 575 Cases

Overall Diagnostic Yield



Independent of age & disease, our overall diagnostic yield is 17.1% (after follow-up), similar to published clinical exome sequencing results (24-26%).

Positive
- **Diagnostic:** known pathogenic variant, consistent with diagnosis
- **Probable:** likely pathogenic variant in a gene that fits phenotype

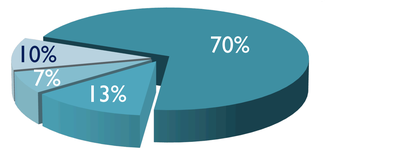
Possible/Uncertain
- **VUS:** variant of uncertain significance in a gene that is consistent with phenotype
- **Contributory:** variant may contribute to but NOT completely explain phenotype
- **Autosomal Recessive:** only 1 pathogenic variant or 2 variants of unknown phase

What category of uncertainty is most common?

Examples of uncertainty

- VUS: *SCN8A* p.E415G (VUS) in participant with seizures → *de novo* per family testing
- 1 hit in AR: Mitochondrial disorder suspected in individual heterozygous for *CPT2* p.S113L; no 2nd variant found
- Contributory: *BARD1* p.E652fs – unclear risk for breast cancer
- Other: Two variants in *MCOLN1*, p.R322* (LP) & p.D471A (VUS) with unknown phase → *In trans* per family testing & gastrin levels confirmed mucolipidosis IV

Types of Possible/Uncertain Results



Family Testing in "Possible" Cases

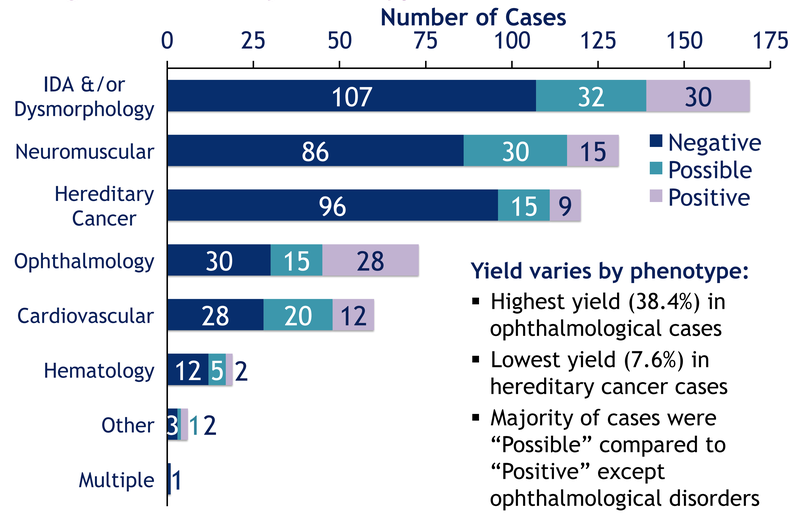
- # of cases: 34
- Uncertain cases → positive from family testing: 29.4%
- Average age: 14.9 yrs
- Estimated diagnostic yield if all cases were trios: 21.8% positive

Diagnostic Yield by Age

Stage of Life		Positive	Possible	Negative
Childhood		22.8% (44)	18.1% (35)	59.1% (114)
Adulthood	18-50yr	18.0% (38)	21.8% (46)	60.2% (127)
	>50yr	9.5% (16)	21.9% (37)	68.6% (116)
	Total (>18yr)	14.2% (54)	21.8% (83)	63.9% (243)

Yield varies by age:
- 22.8% of pediatric cases were positive compared to 14.2% of all adult cases
- In general diagnostic yield decreased with increasing participant age
- This result is likely related to the clinical phenotypes observed most often in children vs. adults.

Diagnostic Yield By Phenotype



Yield varies by phenotype:
- Highest yield (38.4%) in ophthalmological cases
- Lowest yield (7.6%) in hereditary cancer cases
- Majority of cases were "Possible" compared to "Positive" except ophthalmological disorders

Conclusions

How effective is WES as a diagnostic tool?

Depends on the condition
- Low yield in hereditary cancer
  - Most patients had extensive prior testing for known genes
  - Most often multifactorial, even with a family history
- Low yield in neuromuscular disorders
  - Conditions in this category tend to overlap with others
  - Often many genes associated with each condition

Depends on the age of participant
- Lowest yield observed in our cohort over the age of 50
  - Many conditions are more likely to manifest later in life (e.g. cancer, many neuropathies, etc.)
  - Our cancer and neuromuscularcohorts (lowest yields) are mostly comprised of adult participants
  - Verifying variant phase is difficult in elder participants
- Yield is better in pediatric cohort – where family segregation analysis is most practical

Remaining Challenges

Variant interpretation is a bottleneck
- Genome is big & all variation has not been discovered
- Large majority of variants will be VUSs
- Rare variants are frequent & difficult to assess
- Use of appropriate filters can help reduce the number of variants requiring analysis, thus limiting VUSs

Limited phenotypic information in the clinic
- Directly impacts interpretation of results
  - Difficult to narrow the list of variants with limited clinical information
  - Difficult to differentiate between diagnostic & incidental findings
- Propose that clinical labs work closely with clinicians

Can we successfully balance benefit vs. harm?

Acknowledgements

NCGENES is part of the Clinical Sequencing Exploratory Research (CSER) program supported by the National Human Genome Research Institute (NHGRI) & National Cancer Institute (NCI).
U01 HG006487 (J.P.E., PI)

References

- Lee, H. *et al.* Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* **312**, 1880-7 (2014).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-24 (2015).
- Yang, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* **312**, 1870-9 (2014).
- Zhu, X. *et al.* Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. *Genet Med* (2015).