Exploring the diagnostic yield of whole exome sequencing in a broad range of genetic conditions: The first 200 cases in the NCGENES study

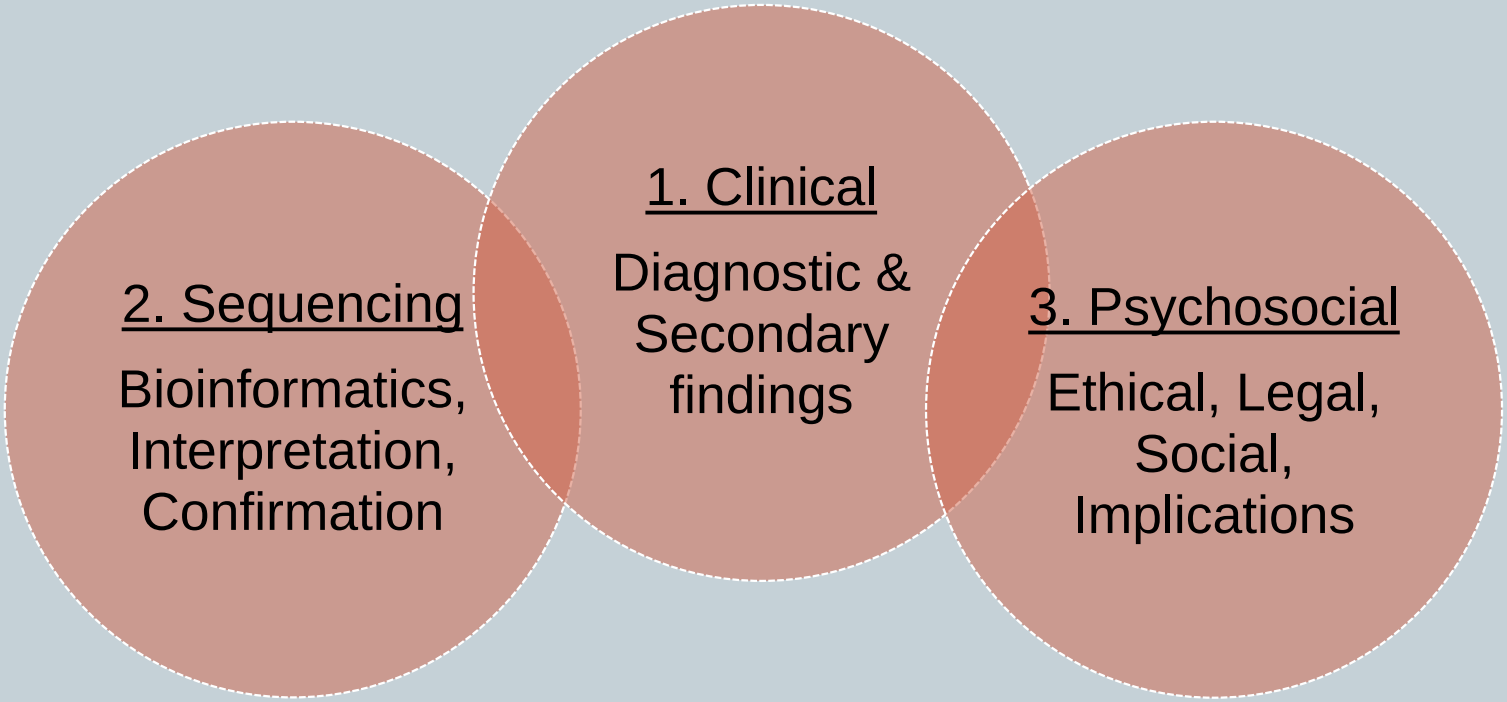
NATASHA STRANDE PH.D.
DEPARTMENT OF GENETICS
UNC CHAPEL HILL

strande@email.unc.edu

NGS as a Diagnostic Tool in the Clinic

- Unanswered questions
  - Cost effectiveness per condition?
  - Diagnostic yield per genetic condition?
  - Off-target results?
- NCGENES (North Carolina Genomic Evaluation of Next-generation Exome Sequencing)
  - 3 Overlapping research projects
  - Determine diagnostic yield of WES for a broad range of genetic conditions

NCGENES Overview



NCGENES Candidate Selection

- Participant Diversity
  - Children and adults with disease
  - Target underrepresented communities
- Disease Criteria
  - Suspected undiagnosed genetic condition
  - Range of disorders: Hereditary cancer, Cardio, Neuro, Retinal, Pediatric Genetics/Dysmorphology

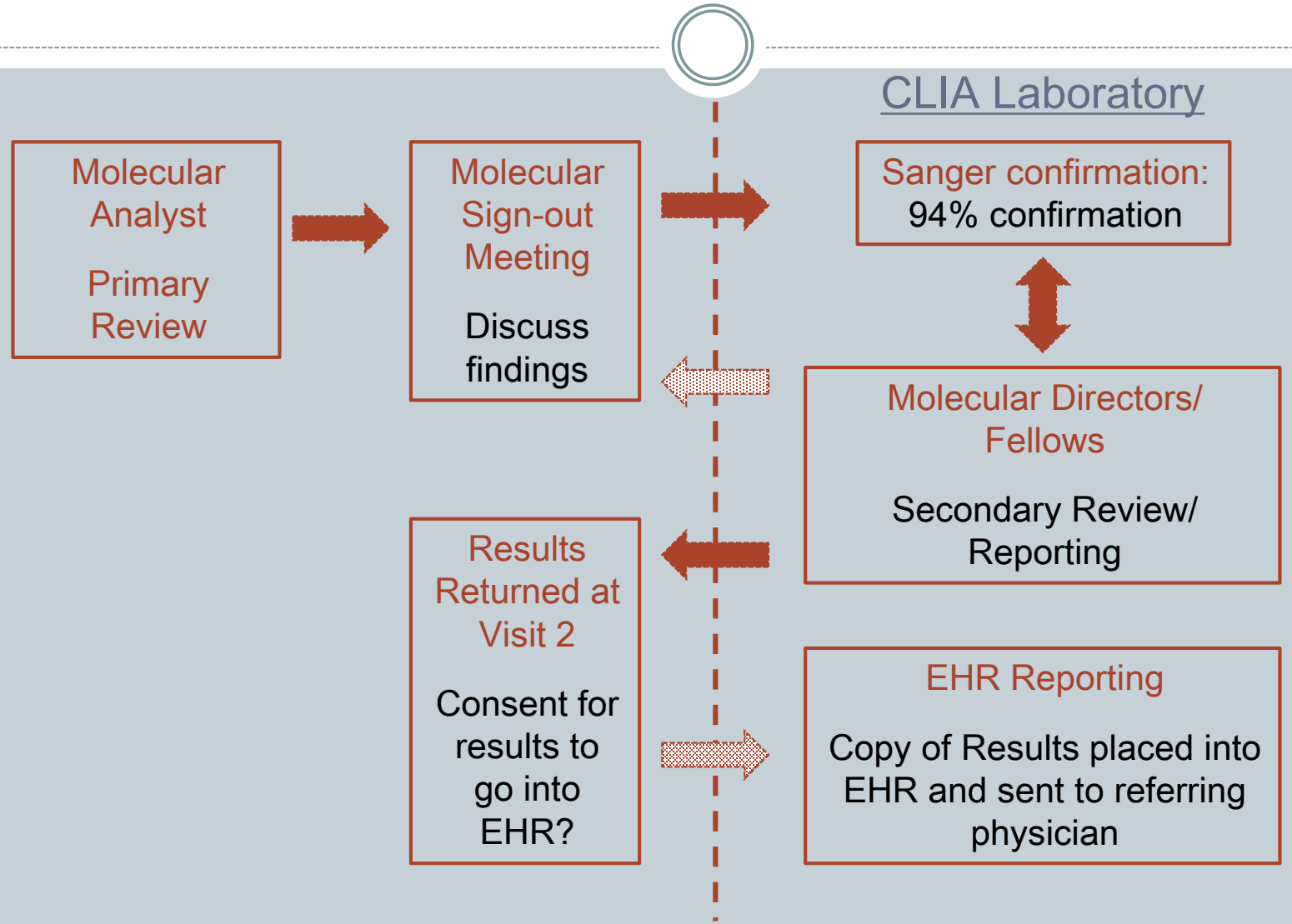
All Those Variants!

- Approach: *A priori* diagnostic gene lists
  - Broad dx gene lists – B. Powell, Platform: 372 @ 05:45PM Tue
  - Ex: Hereditary cancer, seizures, ID & Autism, etc.
  - One or more dx gene list/participant
- Computational Variant Analysis
  - Quality of the data
  - Type of variant
  - Allele frequencies
- Manual Variant Annotation
  - Does the variant make biological sense?
  - Does the gene make phenotypic sense?

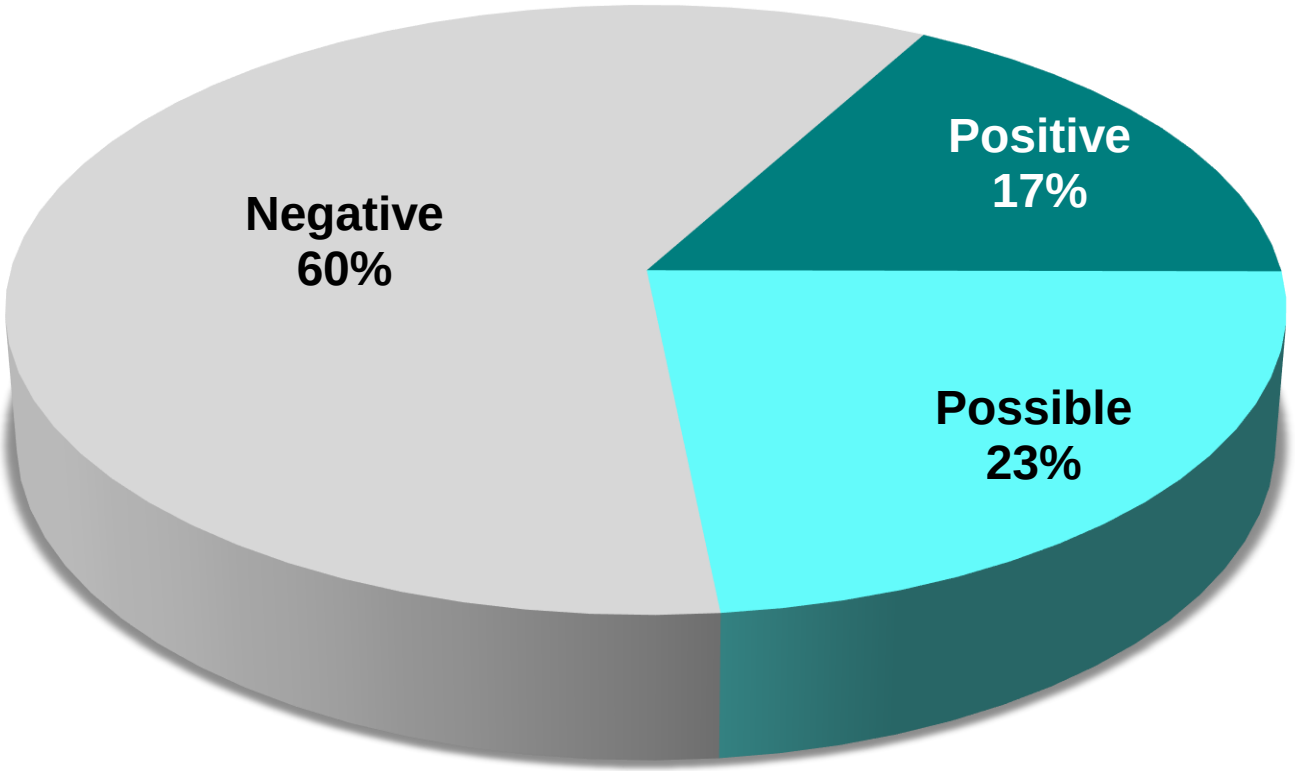
80K – 100K variants/exome

Pic via: Kardia lab at UM

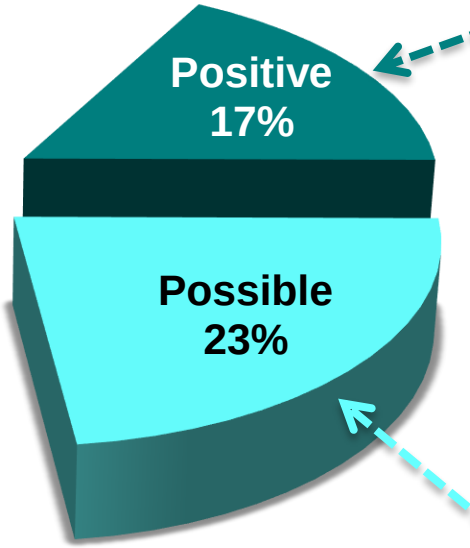
NCGENES Validation Process



Diagnostic Yield for First 300 Cases



Breaking Down the Results: Confidence Matters



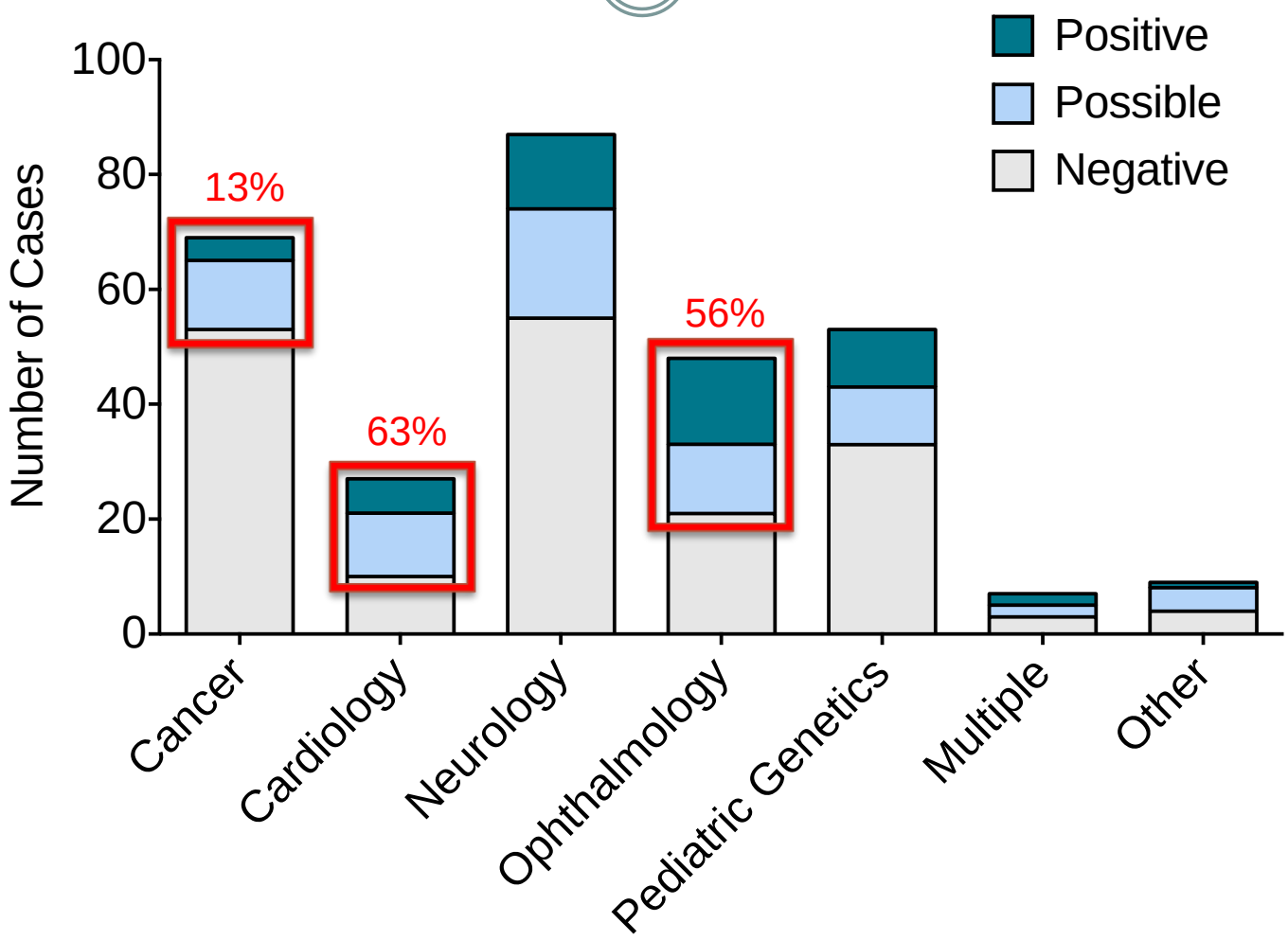
POSITIVE RESULT:

1. **Definitive:** known pathogenic variant in a gene consistent with phenotype
2. **Probable:** likely pathogenic variant in a gene consistent with phenotype

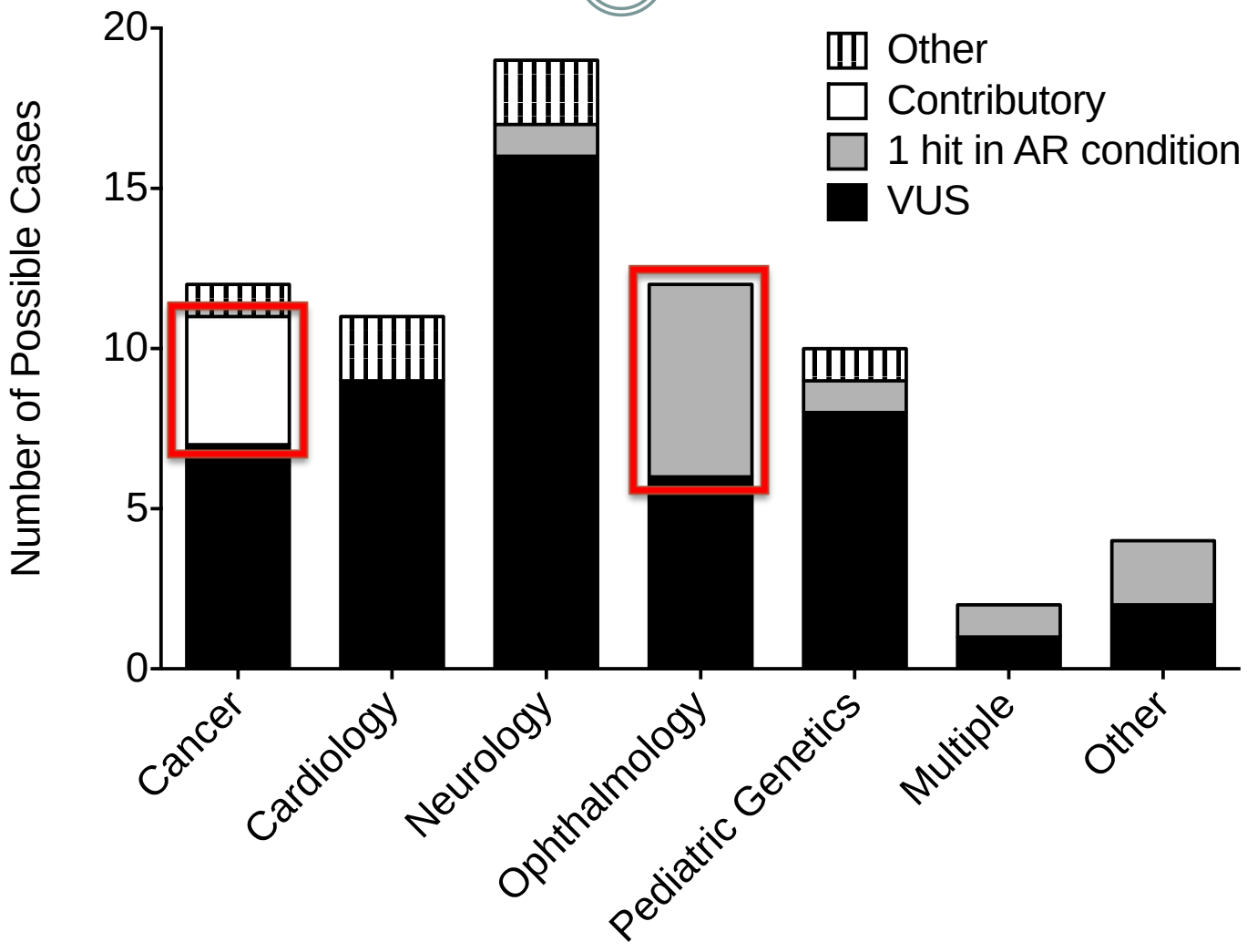
POSSIBLE RESULT:

1. **VUS:** (variant of uncertain significance)
    2. Unclear if the variant is indeed pathogenic
    3. Novel/rare missene in gene consistent with phenotype
2. **1 hit in AR condition:** single probable pathogenic variant in a gene consistent with AR phenotype
3. **Contributory:** gene that may contribute to but cannot completely explain phenotype
4. **Other:** e.g. two variants unknown phase, etc.

Does Diagnostic Yield Vary by Condition?

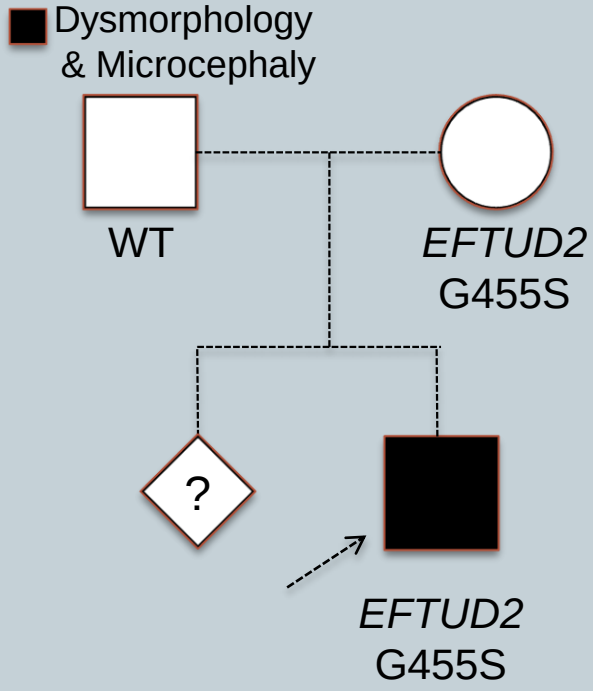


Is All Uncertainty Created Equal?



How useful is family testing?

- ~ 20% of cases possible
  - Little information on variant
  - Will segregation data change this category?
- Segregation analysis
  - 15 families with follow up
  - 12 of these were VUS



Segregation analysis allowed us to go from Possible to Negative

Diagnostic Yield Greatly Varies: Caution is Key

- Why does diagnostic yield vary by phenotype?
  - Was prior genetic testing done?
  - Abundance/lack of evidence for genes on a dx list
  - How frequently is a condition monogenic?
- Family testing is helpful to work up possible results
- NGS in the clinic: a balancing act
  - Harm Vs. benefit to participant re: unclear results?
  - How well does the provider understand the results?
  - The genome is big!!! Coincidences are inevitable!

Thanks from the NCGENES Team!

NCGenes

North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing

Funding Support for NCGENES:

- NHGRI CSER Consortium (U01HG006487)
- University Cancer Research Fund (UCRF)

UNC LINEBERGER

NCGenes

- **Project 1**
  - Jim Evans
  - Myra Roche
  - Kate Foreman
  - Kristy Lee
  - Cécile Skrzynia
  - Julianne O'Daniel
  - Art Aylsworth
  - Cindy Powell
  - Jane Fan
  - Yael Shiloh-Malawsky
  - Bob Greenwood
  - Muge Calikoglu
  - Mike Tennison
  - Tim Carey
  - Brian Jensen
  - Jennifer Brennan
  - Sonia Guarda
  - Hassan Alhosaini

Project 2

- Kirk Wilhelmsen
- Karen Weck
- Phil Owen
- Chris Bizon
- Bradford Powell
- Jessica Booker
- Kristy Crooks
- Ian King
- Dan Duncan
- Laura Milko
- Dylan Young
- Gloria Haskell
- Daniel Marchuk
- Christian Tilley
- Kristen Dougherty
- Mei Lu
- Manyu Li
- Piotr Mieczkowski
- Michael Adams
- Janae Simons
- Sai Balu
- Glenda Stone

Project 3

-Gail Henderson
-Chris Rini
-Debra Skinner
-Cynthia Khan
-Dan Nelson
-Sonia Guarda
-Elizabeth Moore
-Eric Juengst
-Martha King
-Kriste Kuczynski
-Gabe Lazaro
-Lacy Skinner
-Kelly Rasberry
-Michelle Brown
-Christina Leos
-Jenny Morgan
-Christian Tilley
-Sam Cykert

NCGenes

North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing

Acknowledgements

Co-PIs:

Jim Evans

Jonathan Berg

Kirk Wilhemsen

Karen Weck

Gail Henderson

Bioinformatics:

Renaissance computing Institute (RENCI)

Chris Bizon

Phillips Owen

Dylan Young

CLIA lab:

Karen Weck

Jessica Booker

Kristy Crooks

Mei Lu

Manyu Li

Molecular Fellows

Sample Preparation

Laura Milko

(Poster: 2586M)

Christian Tilley

Kristen Dougherty

Molecular Analysts

Kate Foreman

(Poster: 2595M)

Gloria Haskell

(Poster: 2126M)

Bradford Powell

(Platform: 372)

Kristy Lee

(Poster: 1550T)

Julianne O'Daniel

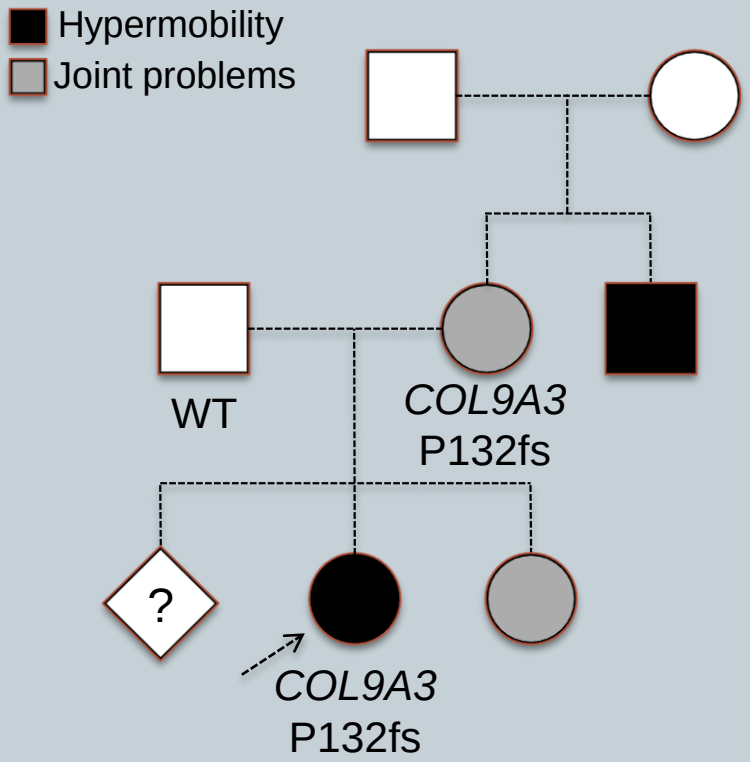
(Poster: 2353T)

Cecile Skrzynia

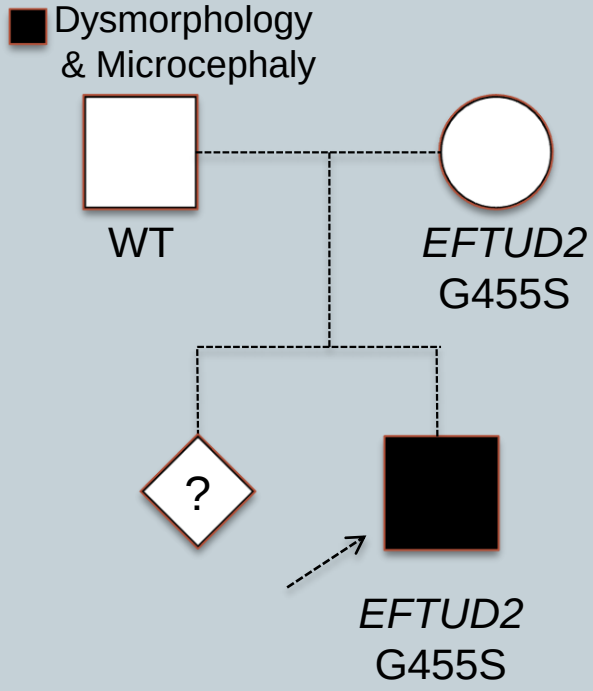
Family Testing Results

Case	Before Testing	After Testing
1	Probable	Definitive
2	VUS	Other
3	VUS	Probable
4	VUS	Negative
5	VUS	VUS
6	VUS	Negative
7	VUS	Other
8	VUS	VUS
9	VUS	Negative
10	Probable	Probable
11	VUS	VUS
12	VUS	VUS
13	1 Variant in AR gene	Negative
14	VUS	Probable

Examples of Segregation Analysis



Segregation analysis allowed us to go from Probable to Definitive



Segregation analysis allowed us to go from Uncertain to Negative

Standard Data Flow

Raw Data
- Raw Reads from the Illumina machines
- Pooled across samples

Processing
- MAPSEQ
- Coordinates programs that must run together to turn raw reads into data products

Data Management
- iRODS
- Store, backup, and tag data

Variant Analysis
- VarDB
- Stores sample-level variant information and associates variants with annotation
- Performs diagnostic and incidental sweeps

Data use
- NCGenes and derived applications
- Return data to users and captures their responses

MapSeq (Grid Computing Engine)

- Raw data needs to be:
  - Demultiplexed
  - Aligned
  - QC checked
  - Variant Called
- This basic set of jobs actually requires about 20 steps using about 8 different tools, all of which must coordinate where and when they will run.
- MapSeq is a tool that manages running these jobs in a standard way (a pipeline) that provides
  - Flexibility in where jobs run
  - Error handling
  - Auditing capability
- End results of these steps are the basic data products:
  - BAM files (aligned reads)
  - VCF files (called variants)
  - Auxiliary files (coverage, QC, etc.)