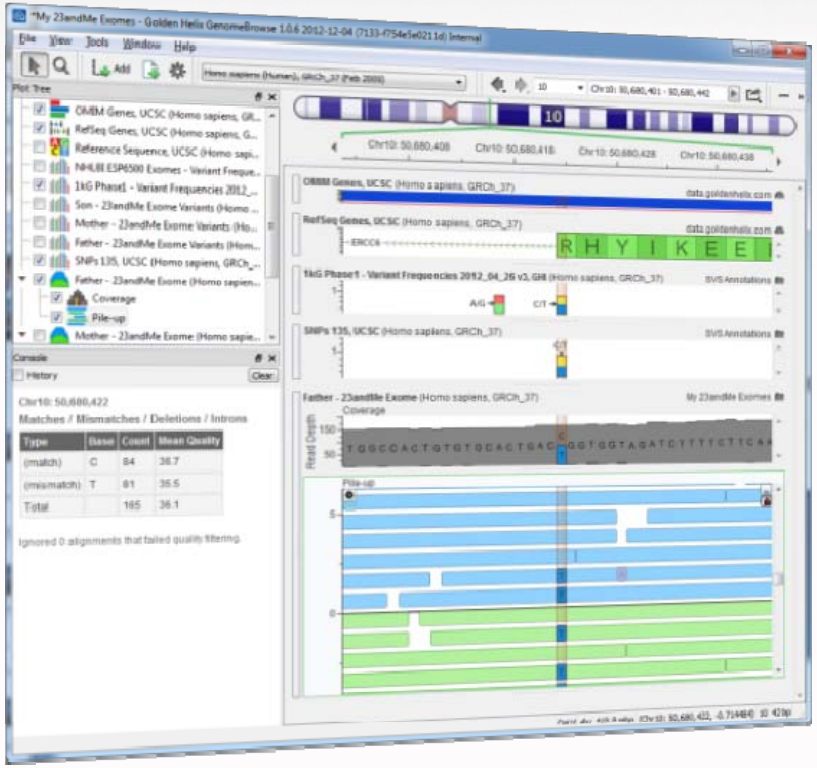
GOLDEN HELIX
Accelerating the Quest for Significance™

@gabeinformatics: 23andMe Variant Analysis of My Personal Exome



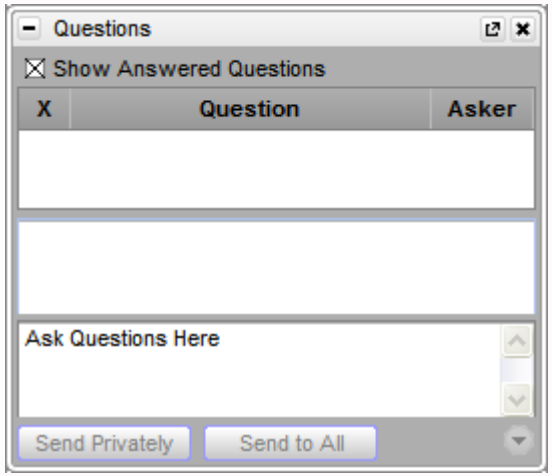
Gabe Rudy

Vice President of Product Development

December 5, 2012

Questions during the presentation

Use the Questions pane in your GoToWebinar window



GOLDEN HELIX
Accelerating the Quest for Significance™

Overview

- Exome Sequencing and Medical Tests
- Can Consumer Genomics Benefit from NGS?
- What I Hope to Do with My Exome
- My Exome: By The Numbers
- Summary of Results

Exome Sequencing: Success Leads to…

- **Wave One: Exome Sequencing as Diagnosis of Last Resort by Researchers**
  - Rare, Congenital, Highly Penetrant, Presumed Monogentic
- **Wave Two: Expanded Research and Select Clinical Use**
  - Research outside above constraints
  - Clinical Exomes as end to diagnostic odyssey or even “shortcut”
- **Wave Three: Path to Personalized Medicine?**
  - Standardized clinical use
  - Consumer driven innovation?

Journal of MEDICAL GENETICS

ORIGINAL ARTICLE

Clinical application of exome sequencing in undiagnosed genetic conditions

Anna C Need,[1] Vandana Shashi,[2] Yuki Hitomi,[1] Kelly Schoch,[2] Kevin V Shianna,[1] Marie T McDonald,[2] Miriam H Meisler,[3] David B Goldstein[1,4]

ABSTRACT

Background There is considerable interest in the use of next-generation sequencing to help diagnose unidentified genetic conditions, but it is difficult to predict the success rate in a clinical setting that includes patients with a broad range of phenotypic presentations.

Methods The authors present a pilot programme of whole-exome sequencing on 12 patients with unexplained and apparent genetic conditions, along with their unaffected parents. Unlike many previous studies, the authors did not seek patients with similar phenotypes, but rather enrolled any undiagnosed proband with an apparent genetic condition when predetermined criteria were met.

Results This undertaking resulted in a likely genetic diagnosis in 6 of the 12 probands, including the identification of apparently causal mutations in four genes known to cause Mendelian disease (*TCF4, EFTUD2, SCN2A* and *SMAD4*) and one gene related to known Mendelian disease genes (*NGLY1*). Of particular interest is that at the time of this study, *EFTUD2* was not yet known as a Mendelian disease gene but was nominated as a likely cause based on the observation of de novo mutations in two unrelated probands. In a seventh case with multiple disparate clinical features, the authors were able to identify homozygous mutations in *EFEMP1* as a likely cause for macular degeneration (though likely not for other features).

Conclusions This study provides evidence that next-generation sequencing can have high success rates in a clinical setting, but also highlights key challenges. It further suggests that the presentation of known Mendelian conditions may be considerably broader than currently recognised.

INTRODUCTION

Whole-genome and whole-exome sequencing have proven remarkably successful in identifying the causes of Mendelian diseases. These analyses have generally depended on the availability of more than one unrelated affected individual and/or linkage evidence in at least one family. However, next-generation sequencing (NGS) has also succeeded in identifying causes of genetic conditions even when they are seen in only a single patient.[1–3]

Consequently, there is growing interest in the introduction of NGS into the clinic to aid in the diagnosis of conditions for which no genetic cause can be found with targeted testing or chromosomal arrays. However, in a clinical setting, patients with undiagnosed genetic conditions tend to present with a wide range of clinical features, and it is often necessary to consider each patient's genome individually, rather than looking for common disrupted genes in multiple cases with a similar phenotype. It is not clear what success rate NGS approaches will achieve in providing genetic diagnoses in this more challenging setting. In this study, we have evaluated the use of NGS to provide genetic diagnoses using 12 parent-child trios in which the child had congenital anomalies and/or intellectual disabilities due to unexplained conditions presumed to be genetic. Importantly, the patients were chosen to be representative of a clinical sample of undiagnosed genetic conditions, in that they were not selected for genetic tractability or phenotypic homogeneity.

METHODS

Exome sequencing was performed on each patient and both parents using the Illumina HiSeq2000 platform and the Agilent SureSelect Human All Exon 50Mb Kit. Detailed methods for laboratory work can be found in the online supplementary methods.

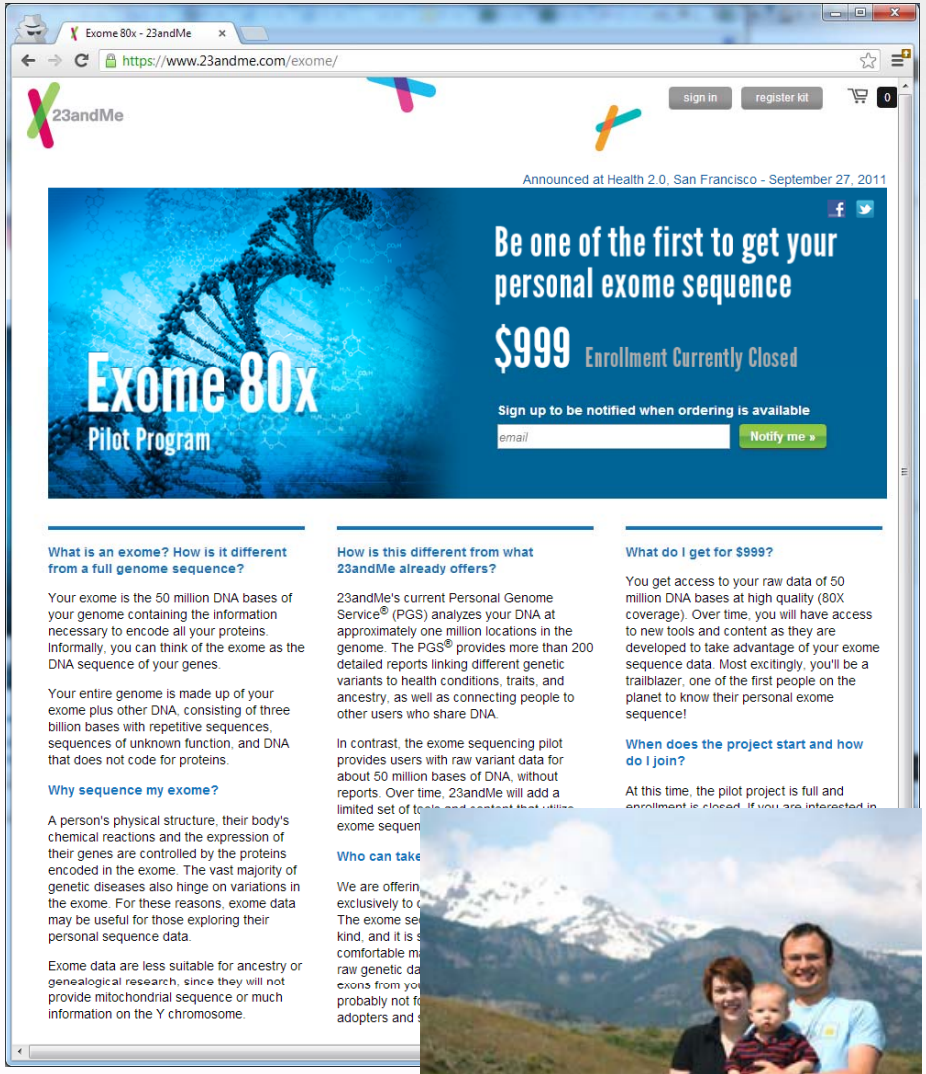
Study population

The research protocol was approved by the Duke Institutional Review Board, and all human participants or their guardians gave written informed consent. Twelve families (child, mother and father) were recruited through the genetics clinic at Duke University Medical Center based on whether their child met two or more of the following criteria: (1) unexplained intellectual disability and/or developmental delay; (2) one major congenital anomaly; (3) 2–3 minor congenital anomalies; and (4) facial dysmorphisms. In addition, the families were required to meet the following eligibility requirements: (1) both biological parents available for testing; (2) previous clinically indicated genetic testing, including a chromosomal microarray (Affymetrix 6.0, http://www.affymetrix.com), had been normal; and (3) no evidence of effects of teratogens, birth asphyxia or non-accidental trauma. Subjects were not eligible if the mother was pregnant at the time of enrolment. Finally, results were only returned to patients and/or patient families following confirmation of detected variants in a CLIA certified laboratory. Controls were subjects enrolled in Center for Human Genome Variation studies through Duke Institutional Review Board approved protocols (n=830).

GOLDEN HELIX
Accelerating the Quest for Significance™

Exome Sequencing in Consumer Genomics

- **23andMe Provides Genotyping Service**
  - ~1M SNPs genotyped
  - 48 Diseases Carrier Status
  - 57 Traits
  - 20 Drug Responses
  - 119 Diseases Risk Predictions
- **Exome done as Pilot Program**
- **80X coverage**
- **Raw Data**
- **No Interpretation**



GOLDEN HELIX
Accelerating the Quest for Significance™

Still Complex and Research Oriented

- **Found 8K phantom variants**
- **Variants Outside Exons?**
- **Exome Capture:**

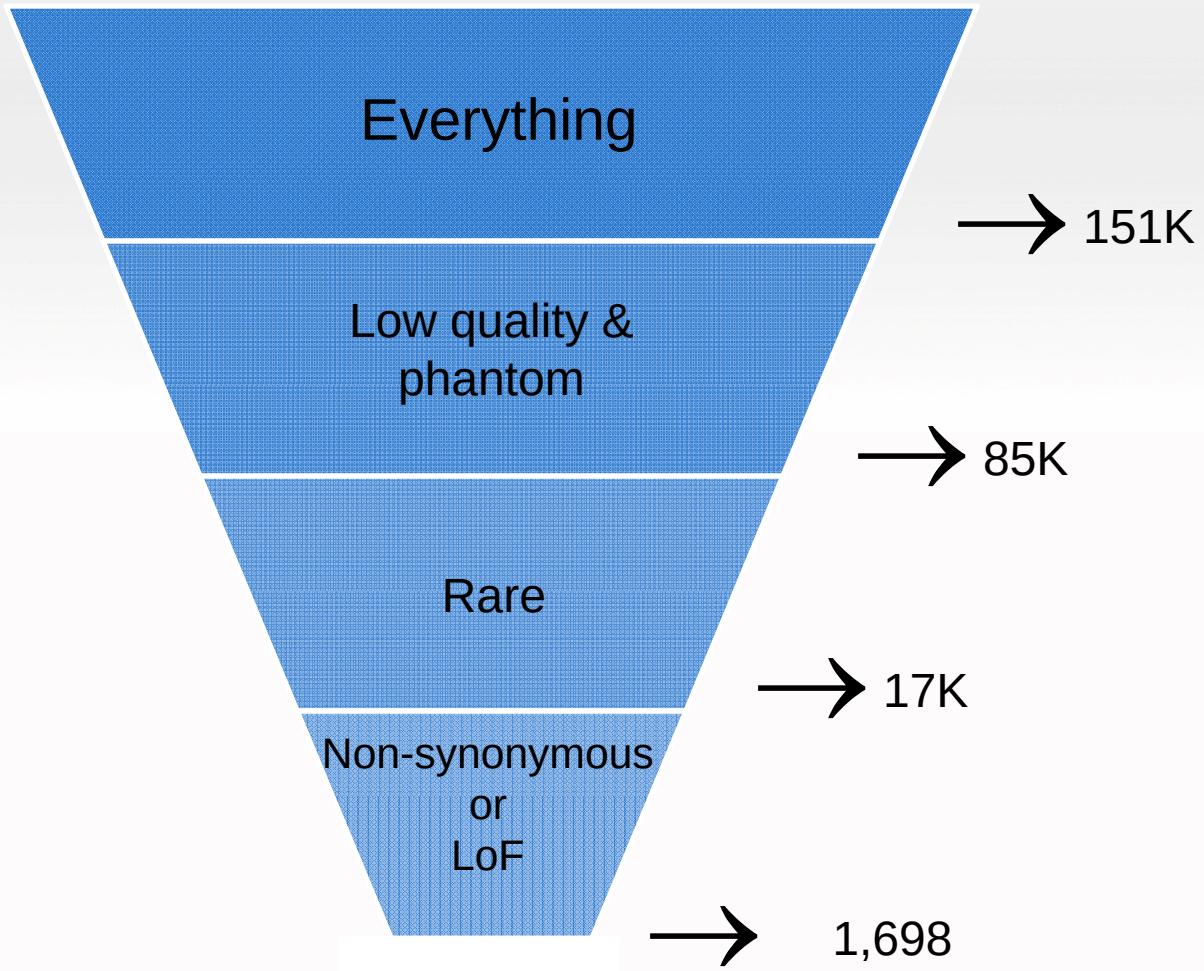
Sequencing an exome at an average depth of 30-50x typically yields 70-75% of the target region at coverage levels of at least 20x. With clinical samples sequenced at 100x, 83-94% of the target region is typically covered at 20x or greater.

~EdgeBio Exome Seq Cheat Sheet

GOLDEN HELIX
Accelerating the Quest for Significance™

Filtering Strategy

- **Follow best practice for high-impact variants**
- **Transparent and reproducible in software**
- **Interpretation more open ended**



Quality Histograms [145]

File View Help

Quality Histograms [145]

File View Help

X-Axis Zoom 0 - 30.3653

Graph Control Interface

- User Graphs
  - Graph 1
    - Read Depth
  - Graph 2
    - LK8250_GQ

Graph | Axes | Add Item

Title Legend

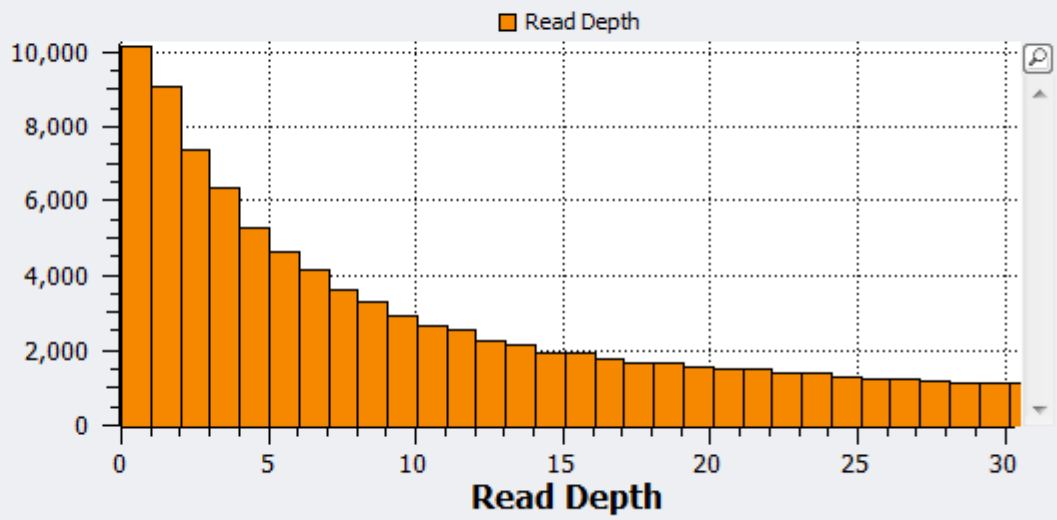
Bin Count 250

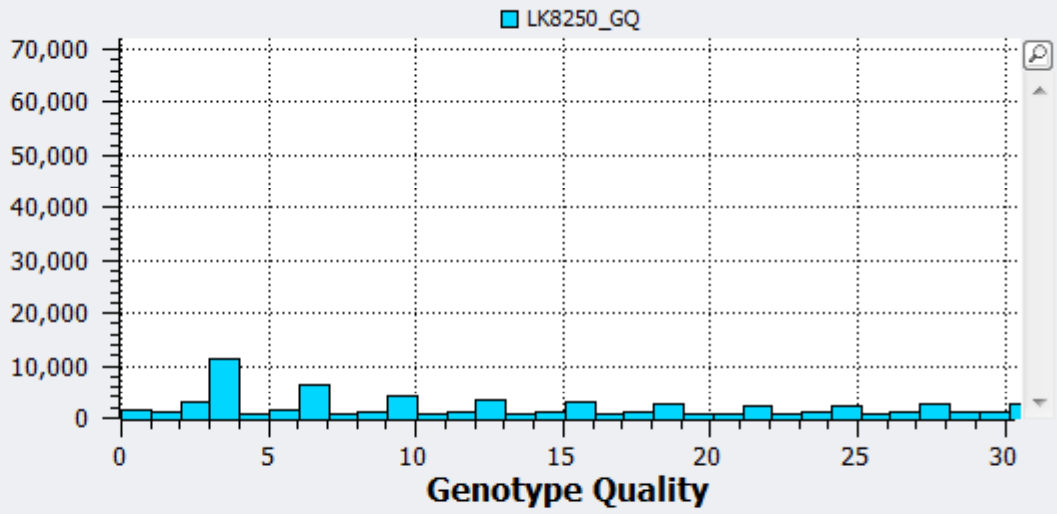
X-Min: 0 X-Max: 30.3653

Y-Min: # Y-Max: #

Data Console

Current | History | User Annotations





Example in GenomeBrowse of Quality Filter

Population Catalog and Variant Classification

Non-coding

	Variant	Rare (Novel)
Intergenic	8,462	3,609 (1,130)
Intronic	48,826	7,516 (4,418)
UTR 3/5	4,128	669 (303)
Non-coding	1,643	648 (183)

Coding

	Variant	Rare (Novel)
Splicing	79	28 (17)
Frameshift Ins/Del	196	138 (118)
Stop gain/loss	113	31 (9)
Non-synonymous	10,252	1,501 (447)
Ins/Del	256	162 (89)
Synonymous	11,080	885 (215)
Unknown	589	176 (36)

GOLDEN HELIX
Accelerating the Quest for Significance™

Loss-of-function: 197
Non-synonymous: 1,501

More Filtering Strategies

- Regions of Chromosomal Duplication (SuperDups)
- Look at genes in OMIM (most)
- Use predictions of genes as recessive/haploinsufficient to weed out low-priority genes
- For NonSynonous missense variants can use functional prediction to annotate

Review Homozygous Variants

Genes of Interest and Homozygous Variants

Loss of Function

	Rare	!Dups	OMIM	Rec Genes
Splicing	28	17	12 (2)	0 (0)
Frameshift Del	60	44	32 (4)	1 (0)
Frameshift Ins	78	66	46 (5)	3 (1)
Stop gain	31	8	7 (0)	0 (0)

Non-Synonymous

	Rare	!Dups	OMIM	Rec Genes
Damaging (3/3)	337	108	85 (0)	9 (0)
Damaging (2/3)	205	139	97 (0)	7 (0)
Damaging (1/3)	136	194	129 (1)	12 (0)
Tolerated/Unk	781	339	204 (4)	10 (1)

GOLDEN HELIX
Accelerating the Quest for Significance™

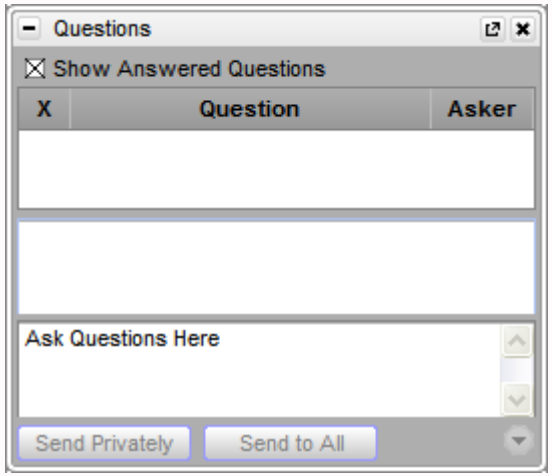
Homozygous: in OMIM: 16
Heterozygous in Rec Genes: 40

Homozygous Variants

Note	Variant	LK8250_AD	LK8250_DP	LK8250_GQ	Gene(s)	Classification	HGVS Coding 1
common	1:54605319-Ins	50,26	76	99	CDCP2	Frameshift Ins	c.1224_1225insC
reference	2:71062833-Ins	106,1	107	99	CD207	Splicing	
in-wife	5:156721864-Ins	6,91	97	99	CYFIP2	Frameshift Ins	c.279_280insC
bad-call	6:44269193-Del	120,1	121	99	AARS2	Frameshift Del	c.2607delG
bad-call?	10:46999604-SNV	21,140	161	99	GPRIN2	Nonsyn SNV	c.724A>G
common	12:26834806-Ins	95,1	96	99	ITPR2	Splicing	
in-wife	14:63784408-Ins	3,141	144	99	GPHB5	Frameshift Ins	c.156_157insC
bad-call	17:7606722-Del	161,6	167	99	WRAP53	Frameshift Del	c.1565delC
bad-call	19:54649671-Del	142,1	143	99	CNOT3	Frameshift Del	c.729delT
in-wife	22:19189004-Ins	6,183	189	99	CLTCL1	Frameshift Ins	c.3601_3602insG
VUS	X:16657321-SNV	0,54	54	99	CTPS2	Nonsyn SNV	c.1342A>C
pathogenic	X:38226614-SNV	0,29	29	84.27	OTC	Nonsyn SNV	c.148G>A
VUS	X:100496711-SNV	0,65	66	99	DRP2	Nonsyn SNV	c.380C>T
VUS/in-5-M	X:105167411-SNV	0,16	16	48.13	NRK	Nonsyn SNV	c.2912A>G
wrong-geno	X:112022302-Ins	61,1	62	99	AMOT	Frameshift Ins	c.3080_3081insCC
VUS/common	X:150349559-Del	106,4	110	96.99	GPR50	Frameshift Del	c.1504_1514delACCACTGGCCA

Do You Have Any Questions?

Use the Questions pane in your GoToWebinar window



406-585-8137
info@goldenhelix.com

GOLDEN HELIX
Accelerating the Quest for Significance™