Analytical Validation and IDE submission – a researcher's perspective

Jonathan S. Berg, MD/PhD
Associate Professor
Department of Genetics
UNC Chapel Hill

NC NEXUS

- Exploratory project examining exome sequencing in the context of newborn screening
  - Assessing performance of sequencing as a screening test
    - 200 “known” affected infants and children
    - 200 “unknown” healthy newborns
  - Studying parental decision-making about whether or not to have their child undergo exome sequencing, and their decisions about whether to learn about non-medically actionable information

This is what a “significant risk” determination results in…

NC NEXUS IDE submission

- Analytic validation
  - Difficult to know how to respond
    - Commercial saliva collection kits
    - Automated DNA extraction in core facility
    - Library preparation using commercial exome kits
    - High-throughput sequencing in core facility
    - Standard bioinformatics pipelines
  - We did not independently “validate” the kit components; we did specify the QC steps that would be followed

NC NEXUS IDE submission

- Sections included (among others)
  - Report of Prior Investigations
    - Description of exome sequencing pipeline
    - Pilot study of exome preparation from saliva DNA
  - Investigational Plan
    - Brief reiteration of wet lab and bioinformatics
    - Detailed information about variant analysis and categories of results to be returned
  - Appendices with detailed laboratory SOP

NC NEXUS IDE submission

- Validation of sequencing
  - Referred to publications on NGS technique
  - Mentioned our previous experience in exome sequencing of ~600 individuals
    - Did not have extensive validation of “knowns”
    - Sanger confirmation of any variants to be reported, with >99% confirmation rate
- FDA had questions about False Positive and False Negative results…

Analytic Validity

- Measures the ability of an assay to accurately detect an analyte
  - Sensitivity: "How often is the test positive when a mutation is present?"
  - Specificity: "How often is the test negative when a mutation is not present?"
  - Also concerned with reproducibility and robustness of the assay

http://www.cdc.gov/genomics/gtesting/acce/acce_proj.htm

Variant

Test Result

	Present	Absent
Positive	**TP**	**FP**
Negative	**FN**	**TN**

The classic 2x2 table

Test Result \ Genotype	Alt/Alt	Ref/Alt	Ref/Ref
Alt/Alt	**TP**	**(TP)**	**FP**
Ref/Alt	**(TP)**	**TP**	**(FP)**
Ref/Ref	**FN**	**(FN)**	**TN**

(It's actually kind of a 3x3 table)

Variant Calling

- Short reads with individual base quality scores
- Reads aligned to a reference sequence
  - Affected by base quality, reference completeness, genomic architecture, genetic variation
- Variant calling as a statistical inference based on observed bases
  - Tunable algorithms can adjust sensitivity/specificity
  - Allele fraction thresholds or Bayesian inference for determining heterozygosity/homozygosity

Variant

Test Result	Present	Absent
Positive	TP	FP
Negative	FN	TN

- Region absent from library
- Low coverage region
- Incomplete reference genome
- Type of variant not accurately "called" (eg. triplet repeat, CNV)

Variant

Test Result

	Present	Absent
Positive	TP	FP
Negative	FN	TN

- Sequencing artifact
- Type of variant not accurately detected by platform (eg. small indel)
- Genomic architecture (homopolymer region, pseudogene)

- Region absent from library
- Low coverage region
- Incomplete reference genome
- Type of variant not accurately "called" (eg. triplet repeat, CNV)

Variant

		Present	Absent
Test Result	Positive	TP	FP
	Negative	FN	TN

- Sequencing artifact
- Type of variant not accurately detected by platform (eg. small indel)
- Genomic architecture (homopolymer region, pseudogene)

- Region absent from library
- Low coverage region
- Incomplete reference genome
- Type of variant not accurately "called" (eg. triplet repeat, CNV)

Technical FN and FP of NGS are somewhat of a "blind spot" without a gold-standard "truth" set

Variant

Test Result		Present	Absent
	Positive	TP	FP
	Negative	FN	TN

- Sequencing artifact
- Type of variant not accurately detected by platform (eg. small indel)
- Genomic architecture (homopolymer region, pseudogene)

- Region absent from library
- Low coverage region
- Incomplete reference genome
- Type of variant not accurately "called" (eg. triplet repeat, CNV)

Technical FP can be minimized by orthogonal confirmation, in which case the rate of technical FP of NGS is less important (except in cost of orthogonal testing)

Variant

Present | Absent

Test Result

Positive

Negative

	Present	Absent
Positive	TP	FP
Negative	FN	TN

- Sequencing artifact
- Type of variant not accurately detected by platform (eg. small indel)
- Genomic architecture (homopolymer region, pseudogene)

- Region absent from library
- Low coverage region
- Incomplete reference genome
- Type of variant not accurately "called" (eg. triplet repeat, CNV)

If the orthogonal method is considered to be the "truth" then the technical FN will include the biases of the orthogonal test

Test Result \ Genotype	Alt/Alt	Ref/Alt	Ref/Ref
Alt/Alt	TP	TP	FP
Ref/Alt	TP	TP	TN
Ref/Ref	FN	(FN)	TN

The orthogonal confirmation method can rescue some of the potential confusion regarding zygosity of the called variants

Variant

	Present	Absent
Test Result: Positive	TP	FP
Variant calling thresholds		
Test Result: Negative	FN	TN

The reality is that test "positives" and "negatives" depend on thresholds set at the level of the variant calling algorithm (quality, depth, allelic ratio, posterior probability)

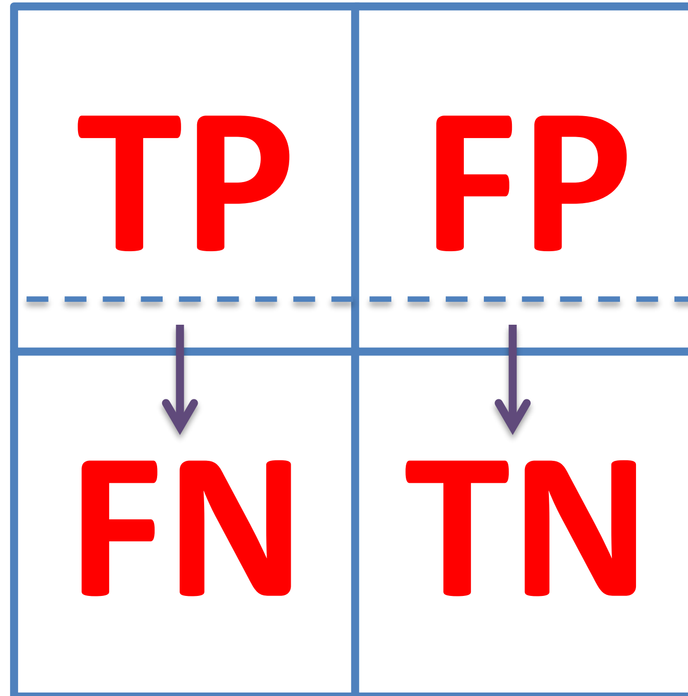
Variant

Present Absent

Test Result

Positive

Negative



Stringent threshold =

- More FN
- Fewer FP

Variant

		Present	Absent
Test Result	Positive	**TP**	**FP**
	Negative	**FN**	**TN**

Relaxed threshold =

- Fewer FN
- More FP

Variant

Test Result		Present	Absent
	Positive	**TP**	**FP**
		Variant calling thresholds	
	Negative	**FN**	**TN**

Variant calling threshold becomes a pragmatic decision – the "confirmation rate" (eg. by Sanger sequencing) is correlated with the statistical probability that a variant is present.

Variant

Test Result

	Present	Absent
Positive	TP	FP
Variant calling thresholds		
Negative	FN	TN

One could empirically determine the “optimal” threshold based on rate of conversion between TP/FN and FP/TN.

But that costs $$$

Variant

Test Result		Present	Absent
	Positive	TP	FP
		Variant calling thresholds	
	Negative	FN	TN

Should a researcher be responsible for quantifying variant calling accuracy before engaging in research? Or is it enough to understand that choices made in the informatics pipeline will affect these parameters? Does it depend on the research question?

What do we know about the accuracy of NGS variant calling?

- A great deal of work has been done:
  - Comparing different sequencing platforms
  - Comparing different variant calling tools
  - Comparing multiple combinations of sequencing and variant calling tools
- My take-home:
  - Nothing is perfect
  - There is room for improvement
  - Things are constantly getting better

O'Rawe *et al. Genome Medicine* 2013, **5**:28
http://genomemedicine.com/content/5/3/28

Genome **Medicine**

RESEARCH **Open Access**

Low concordance of multip
pipelines: practical implicati
genome sequencing

Jason O'Rawe[1,2], Tao Jiang[3], Guangqing Sun[3], Yiyang Wu[1,2], W
Hakon Hakonarson[6], W Evan Johnson[7], Zhi Wei[4], Kai Wang[8,9*]

ANALYSIS _computational BIOLOGY

Integrating human sequence data sets provides a
resource of benchmark SNP and indel genotype calls

[4], Oliver Hofmann[2], Winston Hide[2] &

NUMBER 3 MARCH 2014 NATURE BIOTECHNOLOGY

SCIENTIFIC REPORTS

OPEN

Systematic compari
calling pipelines usi
standard personal e

Received: 18 March 2015
Accepted: 06 November 2015
Published: 07 December 2015

Sohyun Hwang[1,2,*], Eiru Kim[2,*], Insuk Lee[2] & Edward

The success of clinical genomics using next generation se
consistent identification of personal genome variants. As
developed, which show low concordance between their ca
variant callers could give important guidance to NGS-base
confident variant calls for one individual (NA12878) has be
(GIAB) consortium, enabling performance benchmarking
on the gold standard reference variant calls from GIAB, w
variant calling pipelines, testing combinations of three re
Novoalign—and four variant callers—Genome Analysis To
mpileup, Freebayes and Ion Proton Variant Caller (TVC), f
sequenced by different platforms including Illumina2000,

CSH Cold Spring Harbor Laboratory

bioRχiv beta
THE PREPRINT SERVER FOR BIOLOGY

HOME |

Search

New Results

Extensive sequencing of seven human genomes to characterize benchmark reference materials

Justin M Zook, David Catoe, Jennifer McDaniel, Lindsay Vang, Noah Spies, Arend Sidow, Ziming Weng, Yuling Liu, Chris Mason, Noah Alexander, Dhruva Chandramohan, Elizabeth Henaff, Feng Chen, Erich Jaeger, Ali Moshrefi, Khoa Pham, William Stedman, Tiffany Liang, Michael Saghbini, Zeljko Dzakula, Alex Hastie, Han Cao, Gintaras Deikus, Eric Schadt, Robert Sebra, Ali Bashir, Rebecca M Truty, Christopher C Chang, Natali Gulbahce, Keyan Zhao, Srinka Ghosh, Fiona Hyland, Yutao Fu, Mark Chaisson, Jonathan Trow, Chunlin Xiao, Stephen T Sherry, Alexander W Zaranek, Madeleine Ball, Jason Bobe, Preston Estep, George M Church, Patrick Marks, Sofia Kyriazopoulou-Panagiotopoulou, Grace Zheng, Michael Schnall-Levin, Heather S Ordonez, Patrice A Mudivarti, Kristina Giorda, Ying Sheng, Karoline Bjarnesdatter Rypdal, Marc Salit, Genome in a Bottle Consortium

FDA's own effort

Validation on gold-standard materials

- Genome-in-a-bottle consortium is working with NIST to provide reference materials that can be used to validate sequencing platforms and variant calling procedures
- Extremely useful for clinical deployment of NGS technologies
- Is it necessary to use this in research?
- Should researchers re-validate with every change in their platform/pipeline?

Clinical Validity

- Understanding whether a finding is “real” or not is important, but determining what it “means” is critical
  - Is the variant a pathogenic disease-causing variant, or a normal polymorphism?
  - Is the gene truly associated with disease?
  - How well does the case-level data (phenotypic and genotypic) provide an “answer”?

Test Result		Disease: Present	Disease: Absent
	Positive	**TP**	**FP**
	Negative	**FN**	**TN**

Again, the “ideal” test performance 2x2 table

Disease

Test Result

	Present	Absent
Positive	TP	FP
Uncertain	?	
Negative	FN	TN

… but genetic test results are not "ideal"

Variant pathogenicity

- Assessment is based on review of multiple heterogeneous data types
  - prior literature
  - allele frequency
  - protein effect, computational predictions
  - functional assays (when available)
  - family segregation / allelic data

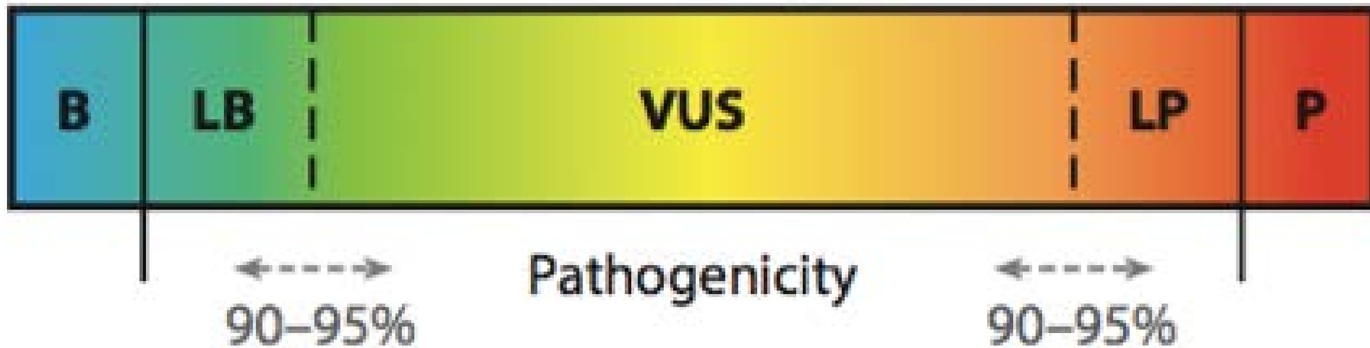
© American College of Medical Genetics and Genomics ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

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

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Disclaimer: These ACMG Standards and Guidelines were developed primarily as an educational resource for clinical laboratory geneticists to help them provide quality clinical laboratory services. Adherence to these standards and guidelines is voluntary and does not necessarily assure a successful medical outcome. These Standards and Guidelines should not be considered inclusive of all proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, the clinical laboratory geneticist should apply his or her own professional judgment to the specific circumstances presented by the individual patient or specimen. Clinical laboratory geneticists are encouraged to document in the patient's record the rationale for the use of a particular procedure or test, whether or not it is in conformance with these Standards and Guidelines. They also are advised to take notice of the date any particular guideline was adopted and to consider other relevant medical and scientific information that becomes available after that date. It also would be prudent to consider whether intellectual property interests may restrict the performance of certain tests and other procedures.

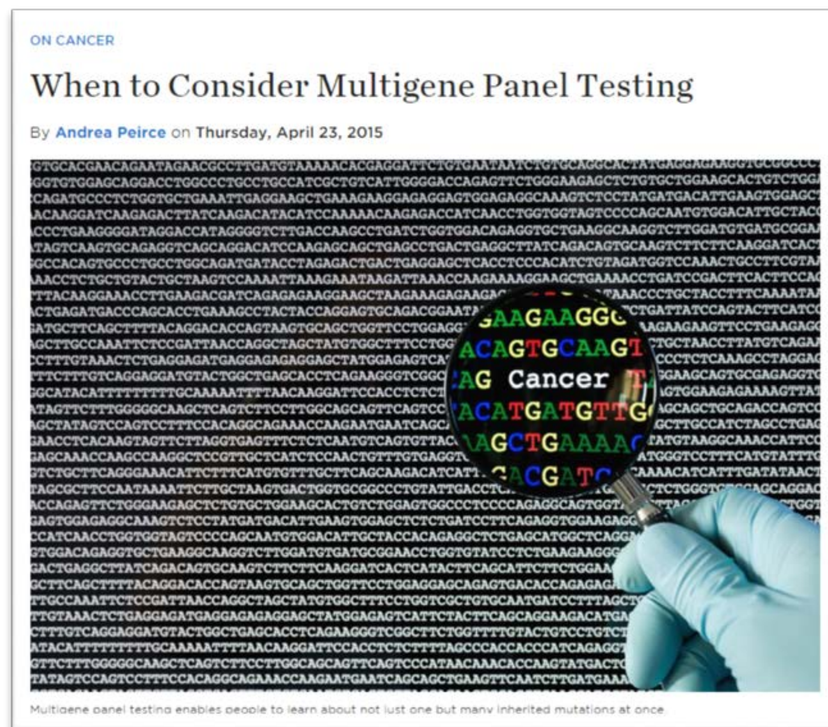
	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			



- Five accepted categories of classification
- “Known” pathogenic and benign variants have >99.9% certainty
- Thresholds for “likely pathogenic” and “likely benign” variants differ
  - IARC = 95%; ACMG = 90%; individual lab rubrics
  - No generalizable methods for quantifying likelihood
- VUS spans a wide range of probability

Gene-disease association

- How strong is the evidence that variation in a given gene causes the disease in question?
- What genes should be included in a multiplex test?
- What genes should be analyzed in a genome-scale test?



ClinGen Clinical Validity Framework

Classification	Description
Definitive	Repeatedly demonstrated in research & clinical settings
Strong	Excess of pathogenic variants in cases vs. controls & supporting experimental data
Moderate	≥3 unrelated probands with pathogenic variants & supporting experimental data
Limited	<3 unrelated probands w/ pathogenic variants
No Evidence Reported	"Candidate" genes based on animal models or disease pathways, but no pathogenic variants reported

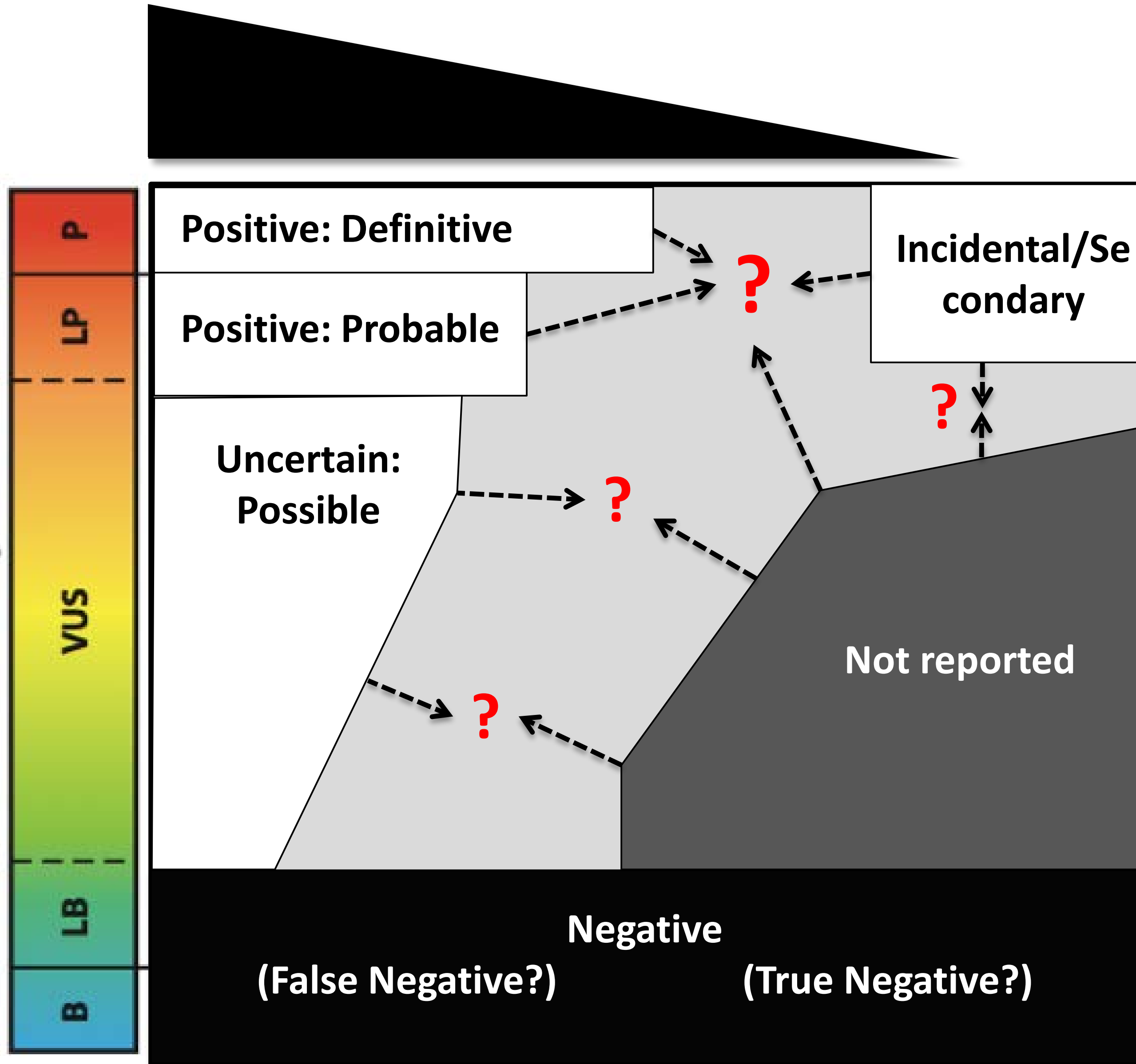
Conflicting Evidence Reported

Classification	Description
Disputed	Convincing evidence disputing a role for this gene in this disease has arisen
Refuted	Evidence refuting the role of the gene in the specified disease significantly outweighs any evidence supporting the role

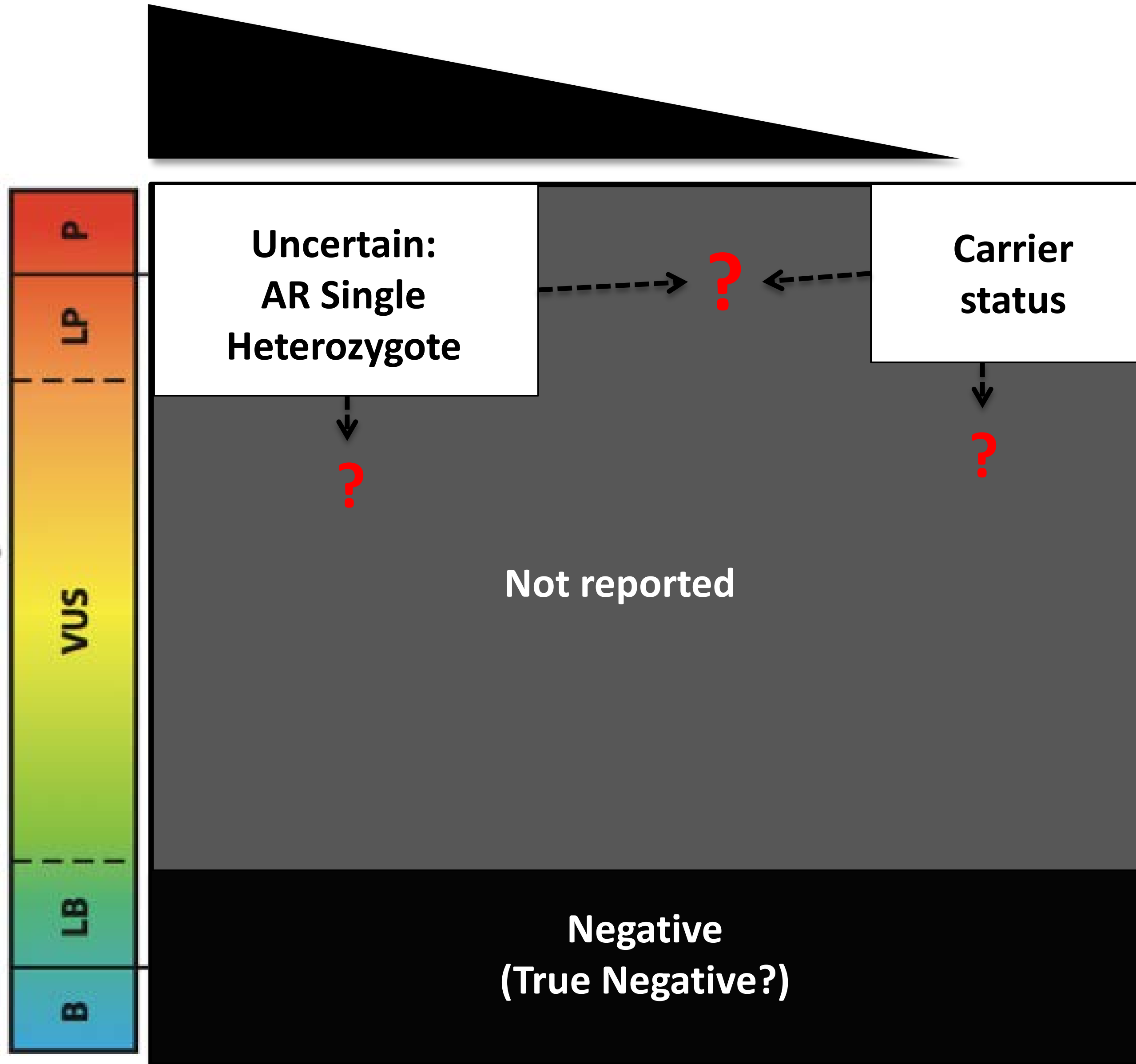
Case level data – phenotypic "fit"

- When reviewing variant data, the analyst also needs to consider whether the phenotype is consistent with the condition of interest
  - If so, the finding is a "diagnostic" finding
  - If not, the finding is a "secondary" finding
- How much phenotype data is needed? How should genes be prioritized for analysis?
- How are the "results" categorized?

Degree of phenotypic match



Heterozygous variant, AD condition; OR Homozygous/biallelic variant, AR condition


Degree of phenotypic match
P
LP
VUS
LB
B
Uncertain:
AR Single
Heterozygote
?
?
Carrier
status
?
Not reported
Negative
(True Negative?)
Heterozygous variant, AR condition

Disease

Test Result		Present	Absent
	Positive	**TP**	**FP**
	Uncertain	**?**	
	Negative	**FN**	**TN**

How does one validate the clinical sensitivity and specificity of a genetic sequencing test?

Simplest example - HbS

- Sickle cell disease can be identified clinically by pathognomonic red blood cell shape
- The condition is caused by homozygosity for a single pathogenic variant – HBB p.Glu7Val
- Analytic performance thus directly determines Clinical sensitivity and clinical specificity:
  - Can NGS accurately detect the nucleotide substitution?

More complicated example - CF

- Cystic fibrosis is clinically recognizable by early failure to thrive chronic bronchiectasis, abnormal sweat chloride level
- The condition is caused by biallelic variants in the *CFTR* gene
- ClinVar has ~250 high confidence pathogenic variants (reviewed by Expert Panel or Practice Guideline)

GeneReviews® [Internet].

Show details

GeneReviews by Title

Search GeneReviews

GeneReviews Advanced Search Help

< Prev Next >

CFTR-Related Disorders

Includes: Congenital Absence of the Vas Deferens, Cystic Fibrosis

Samuel M Moskowitz, MD, James F Chmiel, MD, Darci L Sternen, MS, CGC, Edith Cheng, MS, MD, and Garry R Cutting, MD.

Author Information

Initial Posting: March 26, 2001; Last Update: February 19, 2008.

Table 1.

Summary of Molecular Genetic Testing Used in *CFTR*-Related Disorders

Test Method	Mutations Detected	Mutation Detection Frequency by Population Group	
Targeted mutation analysis	*CFTR* mutations using the original 25-mutation panel [1]	Ashkenazi Jewish	97% [2]
		Non-Hispanic white	88.3% [3]
		African American	69% [2]
		Hispanic American	57% [2]
		Asian American	Unknown
Deletion analysis	*CFTR* exonic and whole-gene deletions	All populations	Unknown
Sequence analysis	*CFTR* sequence variants [4]	All populations	98.7% [5]

See Table A. Genes and Databases for chromosome locus and protein. See Molecular Genetics for information on allelic variants.

1. The original 25-mutation panel recommended by the American College of Medical Genetics [Grody et al 2001] included the 23 mutations listed in Table 8, 1078delT, and I148T. The 23-mutation panel recommended in 2004 is expected to have a similar mutation detection rate [Watson et al 2004]. Other panels may have significantly different mutation detection rates.

2. Grody et al [2001]

3. Palomaki et al [2002]

4. Examples of mutations detected by sequence analysis may include small intragenic deletions/insertions and missense, nonsense, and splice site mutations.

5. Using an assay to sequence all the coding sequences, splice donor and acceptance sites, the promotor region, and two intronic sequences [Strom et al 2003]

Table 2.

Expected Percentage of Abnormal Alleles Detected in Individuals with CF Based on the Detection Frequency of the Test Method Used

Mutation Detection Frequency of Test Method	Percentage of Individuals with CF for which a Given Number of Abnormal Alleles is Identified		
	Two Abnormal Alleles	One Abnormal Allele	No Abnormal Allele
98%	96%	4%	0%
95%	90%	10%	0%
90%	81%	18%	1%
85%	72%	26%	2%
80%	64%	32%	4%
75%	56%	38%	6%
70%	49%	42%	9%
60%	36%	48%	16%
50%	25%	50%	25%
40%	16%	48%	36%
30%	9%	42%	49%

Calculated using Hardy-Weinberg Rule

- CFTR sequencing expected to have ~96% clinical sensitivity for biallelic mutations, and 100% sensitivity to detect at least 1 mutation (either alone or with second VUS?)

Even more complicated example – Hereditary ovarian cancer

- 10-15% of ovarian cancer is associated with rare hereditary cancer syndromes
- Moderate genetic heterogeneity (~10 genes with strong disease association)
- Variable data on proportion of cases accounted for by different types of variants
- Difficult to assess false negatives because most ovarian cancer cases are multifactorial

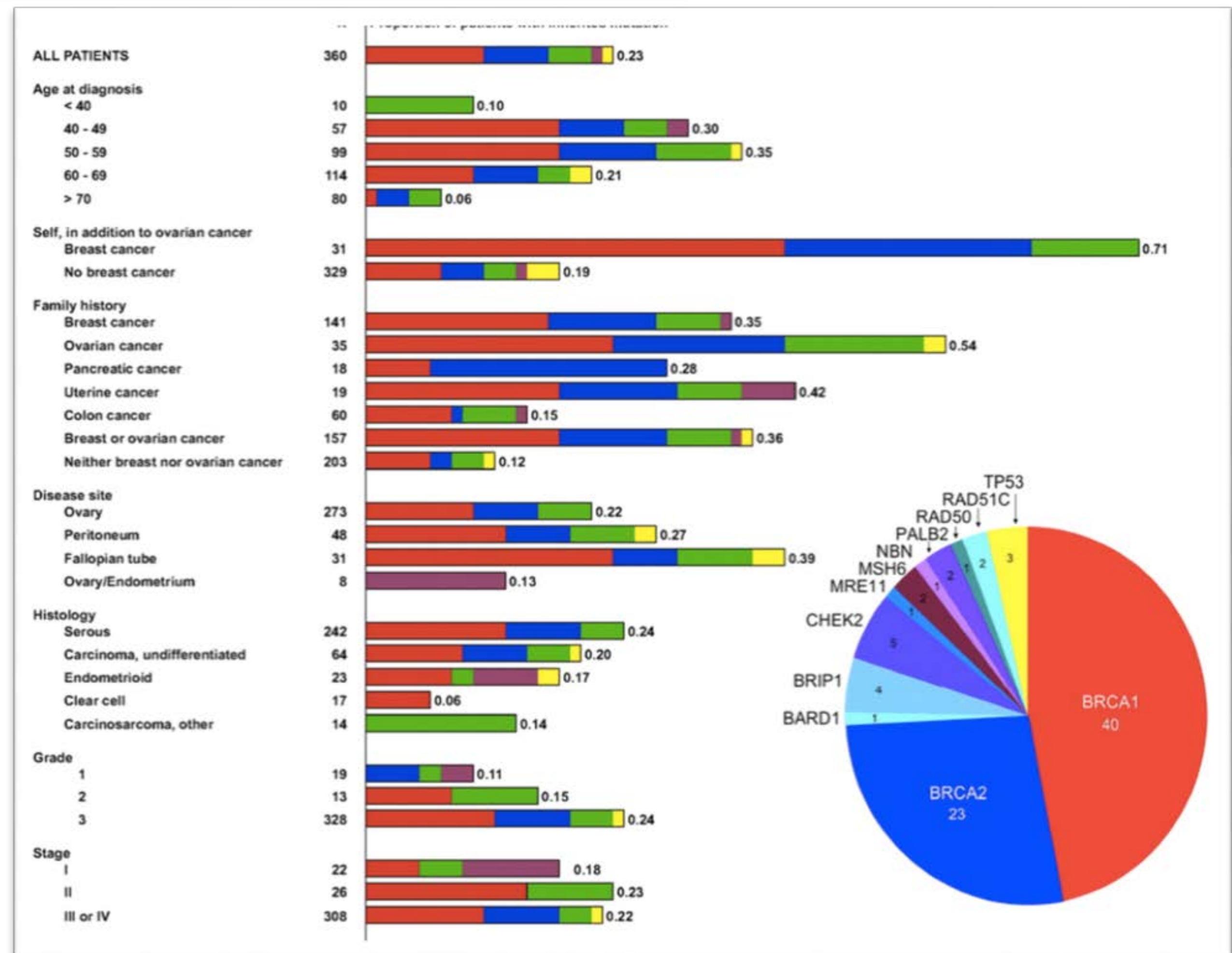
Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing

PNAS

Tom Walsh[a], Silvia Casadei[a], Ming K. Lee[a], Christopher C. Pennil[b], Alex S. Nord[a], Anne M. Thornton[a], Wendy Roeb[a], Kathy J. Agnew[b], Sunday M. Stray[a], Anneka Wickramanayake[b], Barbara Norquist[b], Kathryn P. Pennington[b], Rochelle L. Garcia[c], Mary-Claire King[a,1], and Elizabeth M. Swisher[a,b,1]

[a]Division of Medical Genetics, Department of Medicine, [b]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, and [c]Department of Pathology, University of Washington School of Medicine, Seattle, WA 98195

Contributed by Mary-Claire King, September 19, 2011 (sent for review August 25, 2011)



Ridiculously complicated example – Syndromic Intellectual Disability

- Intellectual disability is relatively common, highly heterogeneous
  - Can be genetic, non-genetic, or multifactorial
  - Molecular etiologies include chromosomal, single gene (recessive, X-linked, *de novo*), epigenetic
- >800 genes have been reported as causing intellectual disability
  - With varying degrees of evidence
  - Virtually none of them have systematic data about the proportion of cases caused, or the contributions of different types of variants

Disease

Test Result		Present	Absent
	Positive	**TP**	**FP**
	Uncertain	**?**	
	Negative	**FN**	**TN**

How does one validate the clinical sensitivity and specificity of a genetic sequencing test?

The good news

- FDA accepted our proposal without excessive requirements for prior validation
  - With the use of CLIA Sanger sequencing as confirmation for all variants returned
  - Understanding that the goal of research was not to commercialize
- Genome-scale sequencing vastly out-performs traditional testing in terms of diagnostic yield
  - Ability to interrogate hundreds of genes simultaneously enhances diagnostic efficiency
  - Practitioners need to understand potential reasons for false negatives (even if they cannot be quantitated)