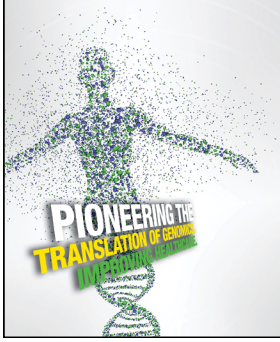
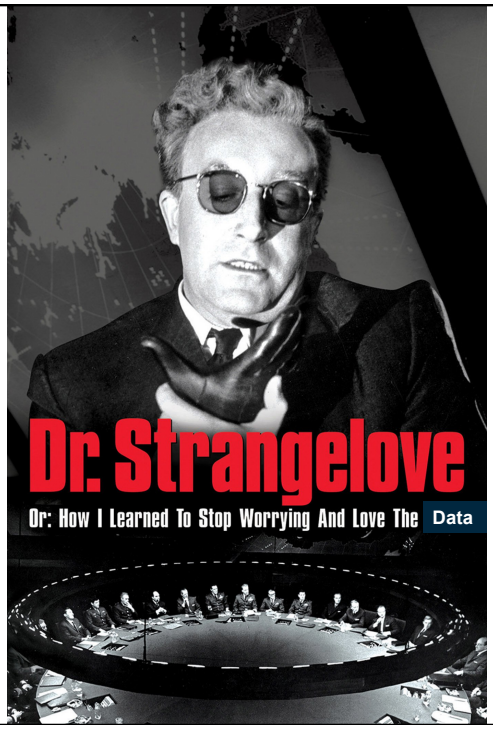
National Society of Genetic Counselors

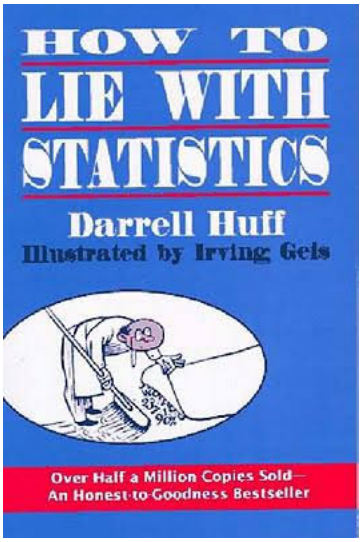
Validation of Clinical Genetic Tests
- or –
How to stop worrying and love the data!

Stephen E. Lincoln
Scientific Affairs, Invitae



34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA





Look out for

Red Flags!

Disclosure

I work for Invitae, a diagnostic laboratory providing clinical genetic testing services.

Opinions I express here are my own and are not necessarily those of my employer.

34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

All NGS is not created equal, even when it uses the same core platform (e.g. Illumina)

There's a reason it's called a "Laboratory Developed Test"

Bioinformatics, up front Biochemistry and Interpretation can vary considerably

The ACC/ACCE/ACCEE Validation Framework

Analytic Validation 

Clinical Validation

Clinical Utility

(**E**thical, Legal and Social Issues)

(**E**conomics)

www.egappreviews.org

34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

As the number of genes tested increases, so does the risk of error

- Assume a hypothetical test:
  - » 1 error per 200kb tested (kb = 1,000 base pairs)
  - » Sequencing target: 2kb per gene
- Single gene test: 1 error per 100 patients
- 25 gene panel: 1 error every 4th patient
- 3,000 gene medical exome: 30 errors per patient!
- Whole exome or genome: lots of errors per patient!

Analytic Validation of NGS based tests

Some Published Examples, Good and Less Good

Lincoln, et al. J Mol Dx 2015. A systematic comparison of traditional and multi-gene panel testing for hereditary breast and ovarian cancer in more than 1000 patients

Judkins, et al. BMC Cancer 2015. Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk

Chong et al. PLOS One 2014. The Validation and Clinical Implementation of BRCAplus

34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

What Questions Should I be Asking?

Documented and Clear Test Specifications and Limitations
- Validation data to back that up (e.g. Indel sizes)

Sanger Confirmation Used always/sometimes/never:
- Exact Criteria?
- If not always, what data backs up that decision?

Interpretation Criteria
- Exact role of prediction algorithms
- Interpretations deposited into ClinVar for Peer Review
- ACMG 2015 Guidelines Compliant?

Analytic Validation Approach

Addresses sensitivity, possibly specificity

1. Start with specimens for which **high quality "gold standard"**[1] data exist
   - » Data must be from a different **"orthogonal"** platform, (e.g. Sanger), not the one you are validating. **Do not confuse reproducibility with accuracy**!
2. Test patients with the new test (blindly)
3. Compare Results

➢ Also do run clinical specimens and confirm variants found (e.g. by Sanger), but…

Addresses specificity, *but not sensitivity*

1. Rehm *et al.*, ACMG NGS Guidelines, *Genet. Med.* 2013

The Analytic Validation Challenge: Where to Get Good Positive Control Specimens?

This is a known hard problem. Options:

- Test patients who had been previously tested another way
  - » Beg other labs to send you DNA
- Test patients who meet the indication for testing
  - » Who are thus enriched for positives (but how many?)
  - » And then confirm… but this generally only tells you about specificity
- Comb public biobanks for relevant samples
- Engineer synthetic positive controls

NIST Genome in a Bottle Consortium (GIAB)

- Collaboration working on a set of reference DNA samples with very high accuracy whole genome sequence data
  - » Known variants in most genes in the genome
  - » Accurate assessment of invariant (negative) regions
  - » First GIAB data set (for sample **NA12878**) is widely used

But…

- » 23% of the genome is not "high confidence", and that includes about ¼ of the known medically relevant genes!
- » Almost all of the alterations in genes are "easy" SNVs with a very few very small indels; no big indels, no complex variants, no hard stuff!
- » No CNV (del/dup) data yet

Genome in a Bottle Consortium

GenomeInABottle.org

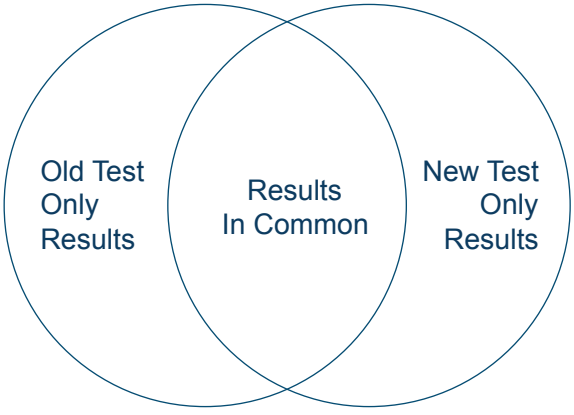
34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

There's Lies, Damned Lies and Statistics*

- Imagine this hypothetical validation study.
  - » Test genes/exons of medical relevance in GIAB sample NA12878
  - » Compare your test results to GIAB reference data
  - » Count concordance
- Imagine an assay which fails to detect all "hard" variants, but which works highly accurately on the easy variants
- For the total spectrum of variants, both sensitivity and specificity will be over 99%, perhaps 99.9% or 99.99%
- **But on a patient sample there is a >10% chance of a false negative!**

*Mark Twain

Not An OK Way to Report a Validation Study
The Infamous Venn Diagram



What does this tell you? Not so much…

The right way: 2x2 Concordance Matrix a.k.a. Confusion Matrix, a.k.a. Contingency Table

1. Fill This In

Gold Standard

New Test	Variant Present	Variant Not Present
Variant Detected	True Positives (TP)	False Positives (FP)
Variant Not Detected	False Negatives (FN)	True Negatives (TN)

2. Do The Math

3. Get The Answer

Accuracy Measures (the Math)

		Condition (as determined by "Gold standard")			
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive ☆	False discovery rate (FDR) = Σ False positive / Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative	Negative predictive value (NPV) = Σ True negative / Σ Test outcome negative
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = Σ True positive / Σ Condition positive ☆	False positive rate (FPR, Fall-out) = Σ False positive / Σ Condition negative	Accuracy (ACC) = Σ True positive + Σ True negative / Σ Total population	
	Negative likelihood ratio (**LR–**) = FNR/TNR	False negative rate (FNR) = Σ False negative / Σ Condition positive	True negative rate (TNR, Specificity, SPC) = Σ True negative / Σ Condition negative ☆		
	Diagnostic odds ratio (**DOR**) = LR+/LR–				

Wikipedia: Sensitivity and specificity (Accessed 9/2015)

Hypothetical Example

	Variant Present	Variant Not Present
Variant Detected	10 TP	0 FP
Variant Not Detected	0 FN	90 TN

100 patients were tested with a new technology and previously by an existing **gold standard method**. The results were:

Sensitivity = TP / (TP + FN) = 10/10 = 100%
False Negative Rate (FNR) = 1 – Sensitivity = 0%

Specificity = TN / (FP + TN) = 90/90 = 100%
False Positive Rate (FPR) = 1 – Specificity = 0%

Positive Predictive Value (PPV) = 10/10 = 100%
False Discovery Rate (FDR) = 1 – PPV = 0%

But There's a Problem… No CIs!

	Variant Present	Variant Not Present
Variant Detected	10 TP	0 FP
Variant Not Detected	0 FN	90 TN

100 patients were tested with a new technology and pr g
gold standard method. The results were:

Sensitivity = TP / (TP + FN) = 10/10 = 100% CI 78% - 100%
False Negative Rate (FNR) = 1 – Sensitivity = 0%

Specificity = TN / (FP + TN) = 90/90 = 100% CI 97% - 100%
False Positive Rate (FPR) = 1 – Specificity = 0%

Positive predictive value (PPV) = 10/10 = 100% CI 78% - 100%
False Discovery Rate (FDR) = 1 – PPV = 0%

95% CIs calculated using the Jeffries method

34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

Analytic validity of NGS in N=1105 individuals

NGS vs. Traditional Methods In 1105 Individuals	
Sensitivity	**100.0%**
Specificity	100.0%

To achieve this, **specialized NGS methods,** biochemical and bioinformatics, are required.

The most challenging classes of variation tend to be **not well represented** in other validation studies.

750 Comparable Variants (Pathogenic or Otherwise)	
Sequence Changes	721
Del/dups (CNV)	**29**

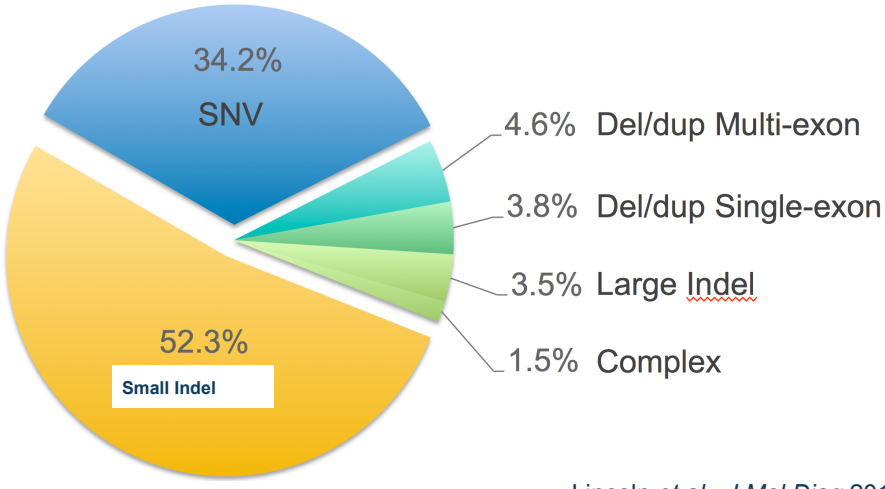
Single Nucleotide	549
Small Indel	156
Large Indel*	**13**
Complex	**6**

Large Indel is deletion≥10bp, insertion≥5bp

"Complex" includes homo-polymer associated variants, indels in low-complexity sequence, short range haplotypes, etc.

Lincoln *et al., J Mol Diag* 2015

A significant fraction of the pathogenic variants in clinical cases are technically challenging



Lincoln *et al., J Mol Diag* 2015

Analytic validity of NGS in another study

NGS vs. Traditional Methods in 250 Patients	
Sensitivity	100.0%
Specificity	99.99%

Note: These numbers assume that Table 2 includes all of the indels and del/dups, and that the remaining variants in the are benign SNPs. This could have been worded more clearly in the paper.

3025 Variants Appropriate to Measure Sensitivity	
Sequence Changes	3021
Del/dups (CNV)	4?

Single Nucleotide	3010
Small Indel	11?
Large Indel	0
Complex	0

Large Indel is deletion≥10bp, insertion≥5bp

"Complex" includes homo-polymer associated variants, indels in low-complexity sequence, short range haplotypes, etc.

Chong *et al.*, *PLOS One* 2014

99.99% Specificity Sounds Pretty Good

One standard measure of specificity is the fraction of false positives per base pair tested

Chong et al. reports from the very same data 30 false positives in 250 patients, concluding (correctly) that Sanger confirmation is required in their lab.

Conclusion: *Per base pair specificity can be a very misleading metric, but many labs still use it…*

Better: False Discovery Rate 30/3025 variant calls or ~1% (this includes benign SNPs - the FDR for pathogenic variants is not determinable from this paper but is higher)

Red Flags to Watch Out For

Over-reliance on the Genome in a Bottle (NA12878) data

No confidence intervals (CIs) given

Unclear technical specifications or assay limitations

Inadequate data to back up those specifications
- Few large indels, for example

Per base-pair specificity only (hard to know what it means)

Presumption that published/reviewed studies are good

Good News

NGS can be highly accurate
- With specialized biochemistry and bioinformatics
- Not so much with “off the shelf” methods, currently

There are good validation studies out there

Genetic counselors can evaluate these studies and ask the right questions!

34th Annual Education Conference
October 21-24, 2015 | David L. Lawrence Convention Center | Pittsburgh, PA
National Society of Genetic Counselors

Questions?

steve.lincoln@invitae.com
steve.lincoln@me.com