Development & Implementation of NG Multi-gene Sequencing Panels: A step toward WES/WGS

Sherri J Bale, PhD, FACMG
Managing Director
GeneDx
Conflict of Interest Statement and Disclaimer

- I work for GeneDx, a wholly-owned subsidiary of BioReference Laboratories, Inc., a public company.
- GeneDx provides genetic testing for hereditary disorders, and develops and applies new methods for molecular diagnosis.
- I own stock in BioReference Laboratories.
- The discussion pertaining to how to properly validate an NGS test is my opinion, and does not necessarily represent that of CAP, ACMG, CDC or any other professional body.
Learning Objectives

- Understand the differences between gene panels, whole exome, and whole genome sequencing
- Know when each test is most appropriate
- Understand the approach to validation for NextGen sequencing tests
Evolution of gene tests

- Single Variant
- Full sequence of individual gene
- Selected gene panel
- Whole Exome sequence
- Whole Genome sequence
When to develop a gene panel

- Multiple genes, each of which could cause the same or related phenotype
  - Ex: Hypertrophic Cardiomyopathy
    - 23 genes, mostly coding for sarcomeric proteins; specific mt genes, GLA

- Fewer than 100 genes

- Strong evidence for inclusion (clinical utility)
Development of a gene panel

- Target enrichment
  - Individual PCR designed for each amplicon
  - Microfluidic chips (Raindance, Fluidigm)
  - Microarray (solution based capture)
Target Enrichment: PCR

- Standard approach, PCR primers designed to flank all exons, placed to capture the exon/intron boundaries
- Fragment size not critical
- Allows optimization to
  - Avoid pseudogenes
  - GC rich regions
  - Homologous genes
Microfluidic chip (Raindance Technology)
In-solution hybridization (e.g. Agilent SureSelect)
Targeted panels: Issues

- Coverage depth – define
- What is a complete/finished test?
  - Pre-emptive amplicons
  - LC/LQ amplicons
  - Sensitivity
    - CNVs, repeat regions, pseudogenes, in/dels
- Reporting
Development of a gene panel
Ensuring “even” coverage for all amplicons in the panel, reducing the number of LC amps

LQT after re-balancing; 231 amplicon test
WES

- Sequencing of the exome (all coding exons of all genes)
  - ~1% of the genome (30Mb)
  - ~20,500 genes
  - 180,000 exons

- Capture of the exons using liquid-phase hybridization (with DNA or RNA biotinylated “bait”).
When to use Whole Exome Sequencing

- Patients who have undergone an extensive diagnostic odyssey, with no molecular basis identified

- Patients with a clinical phenotype that could be explained by one of many, many genes (MR/cognitive disability/developmental delay) where sequencing each individual gene is prohibitive
Capturing the Exome

- Several companies sell in-solution “capture kits” (Nimblegen, Illumina, Agilent)

WES: Issues

- Percent of target captured (90-95%)
- Depth of coverage (mean, 100X)
- Sensitivity
  - CNVs, indels, repeats, pseudogenes
- Who in the family is part of the test?
- Data analysis – filtering for variants
- Reporting
  - “secondary findings”
Whole Genome Sequencing

- Includes everything in WES plus:
  - Regulatory regions
  - Introns
  - 3’ and 5’ UTRs

- Will result in ~3 million variants called, ~2/3 in non-coding regions. Several hundred thousand will be novel

Fragments generated by random shearing, joined to a pair of oligos in forked adapter configuration. Ligated products amp’d using 2 primers -> dbl stranded, blunt ended product, diff adapters at each end.

Fragments denatured, single strands are annealed to complementary oligo on flowcell. Clusters of fragments formed by bridge synthesis.

DNA in each cluster is linearised by cleavage within 1 adapter and denatured -> single stranded template for sequencing by synthesis.

Long range paired end sample preparation. To sequence the ends of long (e.g. >1 kb) fragments, the ends are tagged by incorporation of biotinylated (B) nucleotide, circularized, forming a junction between the two ends. Circularized DNA is randomly fragmented and the biotinylated junction fragments are recovered and used as starting material.
When to use Whole Genome Sequencing

- Price comes down
- WES is unrevealing
  - One mutation identified in a gene in a recessive disorder; where is the other?
  - Want to investigate:
    - Regulatory regions
    - Complete introns
WGS: Issues

- Sensitivity
  - CNVs, indels, repeats, pseudogenes
- Reporting
  - “secondary findings”
  - VUS
All current NGS tests are LDT

- LDT (Laboratory Developed Test, aka “home brew”) vs. FDA-approved tests
  - FDA-approved tests are “kits” that are developed and sold by a manufacturer to a clinical laboratory to use, following specific protocols. The clinical laboratory does a limited validation prior to use.
  - LDT tests are developed by the clinical laboratory and require a full validation prior to use.
Components of a test validation

- Sensitivity
- Specificity
  - False negative
  - False positive
- Robustness
- Reproducibility

Relatively straight-forward concepts for a simple genetic test
- Presence/absence of a specific DNA variant identified by sequencing or other assays
- For full gene/gene panel/WES/WGS sequencing, validation may be better done by assessing the METHOD
Three components of an NGS test that require validation – involves the use of an “alternative assessment” approach, as defined by CAP

* Alternative Assessment – COM.01500 (2011) “Alternative assessment may be performed by method and specimen type, rather than for each tested abnormality”
Validation of NGS Tests

- Three components:
  - Platform
  - Test/Panel
  - Analysis pipeline
Platform Validation

- Maybe better called “Optimization” and includes:
  - Library prep method (PCR component for targeted panels/capture by hybridization to targets for exomes/shotgun for genomes)
  - Instrument validation (first by manufacturer, then by clinical lab)

*DOES THE METHOD OF DATA GENERATION BEING USED GENERATE SEQUENCE THAT REPRESENTS THE PATIENT’S GENOMIC COMPLEMENT?*
Analysis Pipeline in the “Old Days”

- Patient has clinical diagnosis of Gorlin Syndrome
Platform Validation

- Define the performance metrics of the system
  - Depth of coverage required to “pass”
  - Quality of data required to “pass”
  - For WES, percent of target required to “pass”
  - Sensitivity of platform to generate detectable sequence variation

- While I view the analysis pipeline as a separate component of NGS validation, it is necessary to use a pipeline to assess the sensitivity metric
Platform Validation

- What types of variation are possible?
- Detection of *which* is being validated?
  - Single base changes
  - Small deletions of one or a few bases
  - Small insertions of one or a few bases
  - Small insertion/deletion changes of a few bases
  - Large deletions
  - Large insertions
  - Gross rearrangements
Sample Selection

- Reference DNAs where the entire genome has previously been sequenced either by:
  - Another lab using the same/similar method (Illumina/SOLiD); some are being developed by CDC consortium
  - Venter DNA, previously done by Sanger (available from Coriell)

- Patient samples with previously-detected variants in genes that are intended to be included in one or several of the laboratory’s gene test panels
Sample Selection

- “Control” DNAs from individuals not expected to have disease-associated mutation(s) in the genes included on one or several of the laboratory’s gene test panels.

- “Patient” DNAs, not previously analyzed on another platform, from individuals with various phenotypes that could be associated with variants in the genes represented on the gene panels.
Sample Selection

- The samples for platform validation must provide a high likelihood that all types of above variants are represented at a significant enough frequency that the ability of the platform to detect each type can be assessed.
- The number of samples should be large enough to provide confidence in the sensitivity of the platform. Performance across all tests developed by the lab can be combined to establish a cumulative platform performance.
Test/Panel Validation

- Addresses unique gene-specific aspects of the panel
- Samples harboring any *common mutation* in a gene on a panel *must* be included among the validation samples
- It is not necessary to have validation samples with mutation(s) in every gene/exon
- Pseudogene issues must be addressed; including genes with homology on the same panel must be addressed. Are they resolvable?
- Sample number should be sufficient to test particularly robustness and reproducibility
- If Sanger confirmation of reportable variants is being performed, concerns about false-positive rates are minimized
- Inter- and intra-run validation is required
- Inter-instrument validation is required (if you have more than one NGS instrument)
Analysis Pipeline Validation

- Both the platform and test/panel validation requires an analysis pipeline to “see” the results
  - Aligns patient sequence to the reference sequence
  - Calls the variants
  - Identifies the zygosity (het, homoz, hemi)
NGS Analysis
Pipeline, not your mother’s pipeline
The parameters of the analysis pipeline can be "tweaked" so that variants that you know were present in the sample are called appropriately. Optimize the pipeline so it functions robustly.

Once the pipeline is established, it can be used in the platform validation and test/panel validations.

Every subsequent change to the pipeline must be validated by running previous samples which include all the types of variants expected to occur in the test.
It’s an interlocking, interdependent system...
Acknowledgements

- Dr. Madhuri Hegde, Emory Molecular Genetics Lab
- Various instrument/reagent websites
- Dr. Yuriy Shevchenko, CTO, GeneDx
Self-Assessment Questions

- What are the 3 components of a NG validation?
- How do gene panels and WES/WGS differ with respect to sensitivity?
- List the types of genetic variation that are problematic for NextGen sequencing approaches to identify with high sensitivity.