Overview of Molecular Diagnostic Principles and Technologies: Part II
Laura J. Tafe, MD
Assistant Professor of Pathology
Assistant Director, Molecular Pathology
Dartmouth-Hitchcock Medical Center
Lebanon, NH

Overview

- Quantitative PCR (qPCR)
  - Real-time PCR (analog)
  - Digital PCR
- Chromosomal microarrays (CMAs)
  - Constitutional
  - Oncology
- Next generation sequencing (NGS)
  - NGS process stepwise
  - Illumina
  - Ion Torrent

Quantitative PCR

PCR applications

- Endpoint PCR
  - PCR + electrophoresis
  - PCR + sequencing
  - PCR + probe hybridization
  - Variable efficiency – inconsistent, not precise
- Real-time PCR
  - Quantitative (viral load)
  - Presence/absence (pertussis)
  - Genotyping (allelic discrimination)
  - Melt curve analysis (specificity)

Real-time PCR

- Target sequence is amplified in the presence of a reporter (probe)
- Instrument excites & detects reporter
- Signal intensity is directly proportional to the amount of amplified DNA
- Threshold cycle (Ct) determination; the cycle at which target is first detected

Goal:
To combine and automate these steps
PCR phases

- **Exponential**: exact doubling of product is accumulating at every cycle
- **Linear**: reagents are being consumed and reaction is slowing
- **Plateau**: reaction has stopped; no more products are being made

### CT: Primary signal analysis

- Threshold cycle
- Threshold value
- Threshold line
- Threshold penetration

Real-time PCR

- Permits rapid target Id; <30 min
- Eliminate post-PCR processing
- Highly specific;
  - Hybridization probes
- Allows multiplexing
- Permits quantification

Detection Chemistries

- Intercalating/binding dyes
  - SYBR® green I
  - Ethidium bromide
- Dual label probes/Quenched probes
  - TaqMan® (5’ hydrolysis probes)
  - Molecular beacons
  - FRET probes

SYBR Green I

- Believed to be a minor groove binding dye.
- No probe; lower reagent cost than TaqMan.
- No more than 0.5μg of DNA in sample.
- Cannot multiplex.

### Specificity

**Melt (dissociation) curves**

(SYBR® Green I)

- Following amplification, temperature is slowly increased (60°C to 95°C, 0.2°C/sec)
- Strands denature
- Dye released; fluorescence decreases
- Melt curve: temperature (X-axis) vs. fluorescence (Y-axis)
Melt curves

- Sharp downward turn denotes melting temperature (T_m)
- T_m is a function of %GC content and amplicon length

TaqMan Probes

- A real-time PCR chemistry
- Based on 5’>3’ exonuclease activity of Taq polymerase
- Also called 5’ hydrolysis probes
- Dual labeled oligonucleotide probe
  - Fluorescent dye (5’) and quencher (3’)

TaqMan uses a dual-labeled probe

Qualitative vs. Quantitative Assays

- Qualitative (yes or no)
  - Diagnostic
  - Screening
- Quantitative (how much)
  - Usually not diagnostic
  - Prognostic
  - Monitoring Progression
  - Response to Therapy
Two basic quantification methods

• Absolute Quantification
  – Standard curve

• Relative Quantification
  – Compare to a reference/internal control
  – ΔΔCt method

Issues to be Considered

• Specimen type
• Extraction Method
• Inhibitors
• Target sequences
• Controls (type and number)

Efficiency of PCR

• 100% - perfect doubling with each cycle
• Elements that effect efficiency:
  – Inhibitors
  – Primer set
  – PCR conditions
• Ideal 90-110%

Amplification curve

• Baseline correction
• Set threshold
• Ct – cycle at which crosses threshold

Quantitative Assay Performance

Log ΔRn

Linear Range

Set Threshold

Ct

Quantitative Assay Performance

Log (Δq)

Linear Range

LOQ

LOD

Analyte Concentration (U/mL)
**Absolute Quantification - Standard Curve**

- $C_T$ is inversely proportionate to amount of starting material
- $3.3$ $C_T$s = 1 log

**Standard Curve – linear regression plot $\log [ ]$ vs. $C_T$**

**Relative Quantification - ΔΔCt method (Comparative Ct)**

- Assumes efficiency = 100%
- Assumes efficiency consistent from run to run
- Reference/internal control
- Doesn’t require a standard curve

**Performance Characteristics That Must Be Determined**

- **Accuracy**
- **Precision**
  - Same results with same sample?
  - Repeatability: With-in run
  - Reproducibility: Between run, inter-operator
- **Sensitivity or Analytical Sensitivity**
  - Limit of detection: lowest amount of analyte that is statistically distinguishable from background

**Trueness (Accuracy)**

**Precision (Reproducibility)**

<table>
<thead>
<tr>
<th>Panel 1 (50)</th>
<th>Panel 2 (500)</th>
<th>Panel 3 (5000)</th>
<th>Panel 4 (50000)</th>
<th>Panel 5 (500,000)</th>
<th>Panel 6 (5,000,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log10 Observed Viral Load</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Roche TaqMan 48 Sensitivity

$y = 0.9533x + 0.2239$

$R^2 = 0.9952$
Applications of Quantitative Molecular Testing

**Absolute quantification**
- Oncology – BCR-ABL
- Minimal residual disease
- Methylation
- Infectious disease – Viral load – HIV, HCV, BKV

**Relative quantification**
- Fold changes – miRNAs, mRNAs
- Copy number – SNPs, validate CMAs

Digital PCR

Applications of digital PCR
- Copy number variation (CNV)
- Rare sequence detection
- Gene expression and miRNA analysis
- Single-cell analysis; plasma
- Pathogen detection
- Next-generation sequencing (NGS) library analysis
Example qPCR

Detection of BCR-ABL

Molecular response
Complete: no detectable Ph+  
Major molecular response: defined as a BCR-ABL RNA level ≤ 0.1% on the International Scale

Chromosomal microarray

Chromosomal microarray analysis (CMA)
• First line test for:
  — Multiple congenital anomalies
  — Delayed intellectual development
  — Autism spectrum disorder

Oligo Arrays
• Oligonucleotides replaced BACs (bacterial artificial chromosome) on arrays
• Detection of smaller alterations
• Greater control over design

Classic Comparative Genomic Hybridization
SNP arrays

- Instead of “comparative genomic hybridization” SNP arrays produce allelic plots and copy number plots
- SNPs limited to specific locations in genome – SNP only arrays = positional restrictions
- Non-polymorphic (copy number) probes fill gaps to allow broad coverage

Copy number + SNP array

- Improved detection of mosaicism
- LCSH: long-contiguous stretch of homozygosity (copy number of 2); UPD, LOH, identity by descent (consanguinity)

Affymetrix

- CytoScan arrays
  - 2.6 million total probes
    - 750 SNPs and 1.9 million copy number only
- ChAS (Chromosome Analysis Software)

Copy number + SNP array

ChAS Analysis Software: Chromosome 4

Copy number + SNP array

Allele Peaks (SNP data)

Each allele has a value of 0.5; difference of A-B
Hemizygous loss: 11 Mb loss (20q11.22)

Allele Peaks - loss

AA
AB
BB

Allele Peaks - gain

AA
AB
BB

Allele Peaks - mosaicism

AA
AB
BB

CNV Filtering

- Based on probes, not size
  - 25 probes for losses
  - 50 probes for gains
- ~15 CNVs per sample
- Many CNVs quickly eliminated
  - Common benign variants (DGV, internal database)
  - Intronic
  - No genes
CNV reporting
ACMG guidelines (postnatal)
• Pathogenic
• Uncertain clinical significance, likely pathogenic
• Uncertain clinical significance
• Uncertain clinical significance, likely benign
• Benign

Example - CMA

Case example - 3 mo female
• Developmental delay, abnormal brain MRI and congenital heart defects (ASD/VSD)

750 kb deletion from 19p13.3

Confirmation by qPCR

Literature search...
Diagnosis
• **19p13.3 contiguous deletion syndrome**
  – Loss of approximately 35 genes, including STK11, within cytoband 19p13.3
  – STK11 - Peutz-Jeghers syndrome
• Phenotypes expand beyond classical features of PJS
  – borderline intellectual disabilities, hypotonia, seizures, mildly dysmorphic facial features, weight and height at or below 10th percentile, as well as variably expressed cleft palate and congenital heart defects

CMAs in oncology
• Copy number changes (deletions/duplications)
  – 1p 19q deletions in gliomas
  – Melanoma vs. atypical spitz nevi
• CN-LOH - UPD
• Amplification
  – HER2 in Breast Cancer
• Research
  Unable to detect balanced rearrangements

Next Generation Sequencing

Adoption of NGS

<table>
<thead>
<tr>
<th>NGS</th>
<th>DNA SEQ</th>
<th>RNA SEQ</th>
<th>Other SEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted gene panels</td>
<td>Exome</td>
<td>Genome</td>
<td>miRNA</td>
</tr>
<tr>
<td>Transcriptome/Gene expression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clinical implementation
Sample preparation
Genomic DNA or Enriched Target Genes

↓
End repair and ligate synthetic adaptors +/- barcodes

↓
Fragment library

Target enrichment
Multiplex PCR
Hybrid Capture

Sequencing modes

- Single end reads
- Paired end reads
Illumina Sequencing by Synthesis with reversible dye terminators

Illumina instruments

Ion Torrent (Life Technologies) pH-based Sequencing

<table>
<thead>
<tr>
<th>Ion Torrent PGM</th>
<th>Ion Torrent Proton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip Type</td>
<td>PGM 314</td>
</tr>
<tr>
<td># of sensors</td>
<td>1.3M</td>
</tr>
<tr>
<td>Total output</td>
<td>10-40Mb</td>
</tr>
<tr>
<td>Run time</td>
<td>1-2 hrs</td>
</tr>
<tr>
<td>Read length</td>
<td>up to 400bp</td>
</tr>
<tr>
<td>Total reads</td>
<td>up to 0.6M</td>
</tr>
</tbody>
</table>

Data Analysis Pipeline
Primary Analysis

- NGS output – FASTQ file
- Base Calling Accuracy – Phred score – probability of incorrect call

Secondary Analysis

- Alignment to reference genome – .bam file
- Integrative Genomics Viewer (IGV, Broad Institute) to view alignment
- Variant calling – VCF file

Alignment to hg19

Tertiary Analysis

- Filter variants
- Annotation
- Generate report

Variant filters:

- Coverage (read depth)
  - Need ~20x coverage to be 99% confident to detect variant allele (germline)
- Mapping quality (confidence in location of read)

Variant filters:

- Allelic fraction (% reads containing variant)
  - Define LODs (eg. <5% SNVs, <20% INDELS)
- Strand bias (preferential presence of variant in one direction)
- Non-coding and Synonymous variants
Annotation

- Databases (OMIM, COSMIC, dbSNP, ClinVar, etc.)
- In-silico prediction software (SIFT, Polyphen2, MutationTaster, etc.)
- Internal curated database

Variant Calls – Constitutional

Coming soon:

- **Standards and Guidelines for the Interpretation of Sequence Variants:**
  A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Association of Molecular Pathology (AMP)

Evidence based classification of variants; standardize terminology (draft)

- Pathogenic
- Likely pathogenic
- Benign
- Likely benign
- Uncertain significance

Variant Calls – Oncology

No guidelines yet, several proposed classifications

<table>
<thead>
<tr>
<th>Level</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Mutation is predictive or prognostic in patient’s tumor type</td>
</tr>
<tr>
<td>Level 2</td>
<td>Mutation is predictive or prognostic in another tumor type</td>
</tr>
<tr>
<td>Level 3</td>
<td>Mutation has been reported as a somatic mutation in cancer</td>
</tr>
<tr>
<td>Level 4</td>
<td>Novel variant of unknown significance (VUS)</td>
</tr>
<tr>
<td>Level 5</td>
<td>Known polymorphism</td>
</tr>
</tbody>
</table>

Example - NGS

**AmpliSeq Cancer Hotspot Panel v2**

- Single pool of primers
  - 207 Primer Pairs
  - 50 Genes
- Targets genomic “hot spots”
- Requires only 10ng DNA

Hagemann, I. et al. Cancer Genetics 2014
Special considerations

- Somatic disease (cancer)
  - FFPE - DNA quality
  - Tumor tissue not 100% tumor/ heterogeneity (Reduced allelic fraction)
  - Need high coverage to distinguish from background error (~1000x coverage)

- Type of mutation
  - SNV, Indel
  - CNV, structural variant (SV)

NGS References

- https://www.youtube.com/watch?v=6Is3W7jFpd (Elaine Mardis 2014)
- Mamanova et al. Nat Methods 2010
Overview

• Quantitative PCR (qPCR)
  – Real-time PCR (analog)
  – Digital PCR
• Chromosomal microarrays (CMAs)
  – Constitutional
  – Oncology
• Next generation sequencing (NGS)
  – NGS process stepwise
  – Illumina
  – Ion Torrent

AMP’s NGS 101 Webinar series

Training and Education Committee

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birgit Funke</td>
<td>NGS 101 for the Clinic</td>
</tr>
<tr>
<td>Patrik Vitazka</td>
<td>Viewing and interpreting sequencing data: An introduction to genome browsers</td>
</tr>
<tr>
<td>Bob Daber</td>
<td>What you need to know about structural variants and NGS nomenclature and Manual over-riding of the nomenclature.</td>
</tr>
<tr>
<td>Monica Basehore</td>
<td>An integrated approach to assay selection and validation</td>
</tr>
<tr>
<td>Colin Pritchard</td>
<td>The role of the pathologist in reporting and communicating accurate and succinct results in the genomic era</td>
</tr>
</tbody>
</table>

DHMC Molecular Pathology Laboratory and Translational Research Program

Francine de Abreu, Ph.D.
Samantha Allen
Leanne Cook
Betty Dokus
Torrey Gallagher
Diane Green
Arnold Hawk
Joel Lefferts, Ph.D.
Emmeline Liu
Rebecca O’Meara
Jason Peterson
Elizabeth Reader
Heather Steinmetz
Laura Tafe, M.D.
Gregory Tsongalis, PhD
Terri Wilson
Eric York
Wendy Wells, MD

Thank you

Laura.J.Tafe@hitchcock.org