Introduction to Next-Gen Sequencing

Christopher Corless, MD, PhD
Professor of Pathology, Oregon Health & Science University
Chief Medical Officer, Knight Diagnostic Laboratories
## Disclosures

<table>
<thead>
<tr>
<th>Company</th>
<th>Honoraria</th>
<th>Consulting fee(s)</th>
<th>Contract or research support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astex</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Asuragen</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayer</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ImClone</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Ion Torrent</strong></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Leica Microsystems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MolecularMD</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Novartis</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pfizer</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Synta</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Tactical Therap.</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Ventana</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Topics

• Brief overview of next-gen sequencing applications
• Preparing DNA ‘libraries’ for sequencing
• Steps involved in data analysis
• CLIA/CAP requirements for NGS
• Variant/mutation assessment and reporting
• Influences of tumor quality and amount
Deploying Next-Gen Sequencing

• Broad-based approach
  – Whole genome
    • All 3.2 billion base pairs
  – Whole exome
    • ~1.6% of genome that encodes proteins
  – Whole transcriptome
    • All mRNAs in a sample
  – Discovery emphasized over cost & throughput

• Focused approach
  – Panels of genes (20-400)
  – Limited to targets regarded as actionable
  – Cost & throughput emphasized over discovery
Sanger Sequencing

Next-gen Sequencing (Massively Parallel Sequencing)

ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG
ACTGGTCC\textcolor{red}{T}GCTGGTTAG

Whole genome
Whole exome
Whole transcriptome

Multi-gene panels

Illumina HiSeq
Ion Torrent PGM
Illumina MiSeq
Preparing a Sequencing ‘Library’

- Hybridization-Capture Approach -

~50-500 ng Genomic DNA
Hybridize to biotinylated probes for desired genes/exons
Purify hybridized RNA probes with magnetic beads
Treat with Rnase and add adapters

Sequence
Preparing a Sequencing ‘Library’

- Amplicon-Based Approach -

20-50 ng Genomic DNA
Add PCR primers to genes/exons of interest
Amplify by PCR
Add adapters and barcodes

Sequence
## DNA Library Preparation

### Hybrid-Capture vs Amplicons

<table>
<thead>
<tr>
<th></th>
<th>Hybrid-Capture</th>
<th>Amplicon-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input DNA</td>
<td>500 – 50 ng</td>
<td>20 ng</td>
</tr>
<tr>
<td>Overnight hybridization</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>On-target reads</td>
<td>40-70%</td>
<td>95%</td>
</tr>
<tr>
<td>Gene copy number</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fusion gene detection</td>
<td>Yes#</td>
<td>No*</td>
</tr>
</tbody>
</table>

#Not all fusions are detected  
*Can be done from RNA instead of DNA

- Whole exomes and large panels (200+ genes): Hybrid-capture  
- Smaller panels: (<200 genes): Amplicon-based
The Sequencing Pipeline

Sequence Generation
- Normal tissue
- Tumour tissue 1
- Tumour tissue N
  - DNA isolation and fragmentation
  - Target sequence capture and amplification
  - Library preparation
  - Single/pair-end sequencing
  - Raw reads passing QC

Initial Data Processing
- Raw reads
  - Mapping reads to human genome (De novo cancer genome assembly)
  - Local realignment around indels and duplicate removal
  - Analysis-ready reads

Identifying Variants/Mutations
- Normal reads
- Tumour 1 reads
- Tumour N reads
  - Ploidy and tumour heterogeneity assessment
  - Ploidy-aware identification of single-nucleotide variation, indels and large genomic rearrangements (GRs)
  - Supporting data for example SNP arrays
  - Analysis-ready variants

Analysis and Reporting
- Raw SNVs
- Raw indels
- Raw genomic rearrangements
  - Identification and removal of sequencing artefacts
  - Identification and annotation of somatic mutations
  - Phase 5. Identification of clinically relevant changes (drug targets, biomarkers, etc.)

## Sequencing Pipeline

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base calling</td>
<td>Reading the actual sequences</td>
</tr>
<tr>
<td>Sequence alignment</td>
<td>Matching sequences to normal genome</td>
</tr>
<tr>
<td>Identifying single nucleotide substitutions</td>
<td>Could be mutations or polymorphisms</td>
</tr>
<tr>
<td>Identifying in/dels</td>
<td>Insertions and deletions</td>
</tr>
<tr>
<td>Translating alterations at the protein level</td>
<td>Requires assigning a transcript</td>
</tr>
<tr>
<td>Variant annotation</td>
<td>Determining what is known from public databases</td>
</tr>
<tr>
<td>Copy number estimation</td>
<td>Looking for copy alterations</td>
</tr>
<tr>
<td>Report generation</td>
<td>Assigning clinical significance</td>
</tr>
</tbody>
</table>
Normal FFPE Breast Tissue

Log2(CN Ratio)

NRAS DDR2 AKT3 ALK VHL PIK3CA FGFR3 KIT KDR PIK3R1 RAF1 EGFR MET BRAF FGFR1 CDKN2A NRAS HRAS KRAF CDK4 PTEN RB1 AKT1 MAP2K1 NTRK3 TSC2 TP53 NF1 ERBB2 STK11 GNA11 AKT2 GNAS
FFPE Ductal Carcinoma of the Breast

Log2(CN Ratio)

EGFR
EGFR Amplification in Breast Carcinoma

68 year old F

Tumor metastatic to the liver

HER2-negative

[6% of breast carcinomas have EGFR amplif]

Bhargava et al. Mod Pathol. 2005 Aug;18(8):1027-33
Recommendations for NGS Validation in CLIA Labs

- **Accuracy:**
  - Comparing assay output with known reference DNAs

- **Precision:**
  - Repeatability (within a run)
  - Reproducibility (between runs; between technicians)

- **Analytical sensitivity:**
  - The threshold for variant detection

- **Analytical specificity:**
  - The false positive rate

- **Reportable range:**
  - The sequences covered by the assay

- **Reference range:**
  - Expected variants within normal reference DNA
Recommendations for NGS Validation in CLIA Labs

- **Accuracy:**
  - Used samples previously genotyped by another platform (Sequenom or Sanger sequencing)

- **Precision:**
  - Repeatability: Duplicate samples on the same run
  - Reproducibility:
    - Same sample run on different chips / by different techs
    - Amplicon read depths

- **Analytical sensitivity:**
  - Comparison of allele ratios measured by mass-spec
  - Dilution series of cell line DNA

- **Analytical specificity:**
  - Sanger sequencing to assess unexpected/suspicious variants

Accuracy and Sensitivity
Sequenom MassArray vs Next-Gen Allele Ratios

$R^2 = 0.92$

53 Known mutations

Precision: Single vs Multiplexed Sequencing Runs

R² = 0.88

43 Known mutations

## Example of Variant List For a 37-Gene Panel

| Chrom | Position_Start | Position_End | Ref | Variant | Type     | Consequence   | Zygosity | Var_Freq | Gene     | p_AA_change |
|-------|----------------|--------------|-----|---------|----------|---------------|-----------|----------|----------|------------|-------------|
| chr4  | 55968089       | 55968089     | T   | G       | SNP      | nonsynonymous | Het       | 61.29    | KDR      | p.K747N    |             |
| chr9  | 21971096       | 21971096     | C   | A       | SNP      | nonsynonymous | Het       | 83.33    | CDKN2A   | p.E88*     |             |
| chr17 | 7577548*       | 7577548      | CGCCCA | C | DEL      | stop gain   | Het       | 30.52    | TP53     | p.G245G    |             |
| chr17 | 7577557*       | 7577557      | A   | C       | SNP      | stop gain    | Het       | 31.28    | TP53     | p.C242G    |             |

### Amino acid

<table>
<thead>
<tr>
<th>Gene</th>
<th>Change</th>
<th>% Allele</th>
<th>Interp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDR</td>
<td>K747N</td>
<td>28%</td>
<td>Het</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>E88*(stop)</td>
<td>48%</td>
<td>Hom</td>
</tr>
<tr>
<td>TP53</td>
<td>G245G</td>
<td>23%</td>
<td>Het</td>
</tr>
<tr>
<td>TP53</td>
<td>C242G</td>
<td>27%</td>
<td>Het</td>
</tr>
</tbody>
</table>

Tumor in Sample = 50% of cells
Specimen submitted: Left cervical node
Reported diagnosis: Moderately differentiated keratinizing squamous cell ca

**PTEN Immunohistochemistry**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Result</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTEN</td>
<td>Expression intact (IHC)</td>
<td>No indication of increased PI3 kinase signaling.</td>
</tr>
</tbody>
</table>

**GeneTrails™ Panel: Mutation Screening by Next-Generation Sequencing**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation Status</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDR</td>
<td>Positive for K747N</td>
<td>This mutation has not been reported in the COSMIC database. However, KDR encodes VEGFR2, which is potentially sensitive to TKI258 and other VEGFR inhibitors.</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Positive for E88*</td>
<td>Genomic alterations affecting the CDKN2A gene, which encodes the p16 tumor suppressor protein, are common in carcinomas and contribute to dysregulation of the cell cycle. Pre-clinical studies suggest that tumors with loss of CDKN2A may be sensitive to CDK4/6 inhibitors.</td>
</tr>
<tr>
<td>TP53</td>
<td>Positive for M234fs*2</td>
<td>There are currently no approved therapeutics that specifically target TP53 mutations.</td>
</tr>
</tbody>
</table>
Head & Neck SQCC

Before Treatment

Treatment Day 15
TKI With VEGFR2 Activity
Metastatic Melanoma in Lung
Tumor purity: ~90%
Selecting Tumor-Rich Material for DNA Extraction

Coring a block

Scraping slides
Metastatic Colorectal Carcinoma in Mesentery

Tumor purity: ~10-20%
Transbronchial Biopsy
Adenocarcinoma of the Lung

Please, get another sample!
Sample Integrity

- **Decalcification** of a bone biopsy or bone marrow biopsy
  - Acid is often used to soften the bone for sectioning and this degrades nucleic acids
  - EDTA-based solutions are much better

- **DNA deamination** is common in older paraffin blocks

![Chemical structures showing DNA deamination](image)

**Result**

Mimics a C>T mutation

**Solution:** amplify both strands and require that mutations be present on both reads
Summary

- NGS has major advantages over traditional assays
  - Good sensitivity (~5% mutant allele)
  - Less input DNA required
  - Can detect gene copy number alterations

- However, there is no such thing as a perfect test. One must keep in mind:
  - Sample quality
    - Tumor content; fixation
  - Assay limitations
    - False positives
    - Challenge of detecting longer in/dels