Next Generation Sequencing for Cancer Research

Vikram Devgan, Ph.D., MBA
Director, R&D
vikram.devgan@qiagen.com
Welcome to the three-part webinar series

Next Generation Sequencing and its role in cancer biology

Webinar 1: Next-generation sequencing, an introduction to technology and applications
Date: March 11, 2013
Speaker: Quan Peng, Ph.D.

Webinar 2: Next-generation sequencing for cancer research
Date: March 18, 2013
Speaker: Vikram Devgan, Ph.D., MBA

Webinar 3: Next-generation sequencing data analysis for genetic profiling
Date: March 25, 2013
Speaker: Ravi Vijaya Satya, Ph.D.
Agenda

- Background of Cancer Research
- NGS in Cancer Research
- QIAGEN’s Solution for NGS-related Cancer Research
Genetic Variations

Mutations

Reference genome
AGCTCGTTGCTCAGCTC

Deletion
AGCTC----GCTCAGCTC

Insertion
AGCTCGTTGCTCAGCGTTC

Indels

Copy number variation

Structural variation

PKD
In 2008 over 12.7 millions new cases were diagnosed across the planet and approximately 7.6 million cancer death occurred

In 2030, these numbers will rise to an expected 21.4 millions new cases and 13.2 million cancer death..........  

If our ability to prevent, diagnose and treat cancer doesn't improve
Need of Cancer Research

- **Identify mutations that drive cancer progression**
  - *Lung Cancer: EGFR*

- **Identify new targets for therapies, biomarkers and diagnostic tests**
  - Cancer susceptibility genes and somatic/driver mutation

- **Identify Individuals who are at significantly increased genetic risk of cancer**
  - ~0.25% of US women (375,000) carry a mutation in BRCA1/2; high risk for breast and ovarian cancer

- **Select drugs based on the genomics of the tumor (Patient stratification)**

  - **EGFR (L858R)**
    - Response rates of >70% in patients with non-small cell lung cancer treated with either erlotinib or gefitinib
  - **KRAS (G12C)**
    - Poor response rate in patients with non-small cell lung cancer treated with either erlotinib or gefitinib
Next Generation Sequencing

Providing new opportunity for comprehensive analysis of cancer genome

- High throughput
  - Massive parallel sequencing

- Cost effective
  - Drastic decrease in cost of sequencing

- Unbiased
  - Detection of all mutations

- Quantitative
  - Digital readout of mutation frequency

Evolution of technologies for mutation detection

ARMS PCR

Liquid bead array

Sequenom MALDI-TOF

Next Generation Sequencing
Cancer Samples: Challenges

- Quantity
  - Limiting for biopsy specimens

- Quality
  - Most biopsies are formalin fixed
  - Often include necrotic, apoptotic cells

- Purity (genetic heterogeneity)
  - Cancerous cells may be a minor fraction of total sample
  - Multiple sub-clones of cancer may be present in one tumor sample

Relevant mutation may be present in low frequency
What does Cancer Researcher Want in NGS?

- Ability to work with FFPE samples
- Deep coverage
  - Detection of relevant mutations with low frequency
- Maximum clinical sensitivity
  - If mutation is there, have a possibility of finding it
- Complete and automated “sample to result” NGS workflow
- Biological interpretation of result
  - Without investing in bioinformatics
Which method to choose?

Types of next generation DNA sequencing:

- Gene Panel
- Whole Exome
- Whole Genome

Instrument throughput

Coverage per sample

Number of samples

Size of target region
Which method to choose?

Types of next generation DNA sequencing

- Gene Panel
- Whole Exome
- Whole Genome

Bench top sequencers can’t be used
Which method to choose?

Types of next generation DNA sequencing

Gene Panel

Whole Exome

Whole Genome

Medium number of samples

Low coverage per sample

Instrument throughput

Whole Exome

Bench top sequencers can’t be used
Which method to choose?

Types of next generation DNA sequencing

Gene Panel
Whole Exome
Whole Genome

Deep Coverage per sample
Instrument throughput
Medium Number of samples
Comprehensive Cancer Gene Panels

Suitable for Bench top sequencers
Which method to choose?

Types of next generation DNA sequencing:
- Gene Panel
- Whole Exome
- Whole Genome

Focused cancer-specific gene panel sequencing is most effective method.
Which is most Suitable Target Enrichment Technology?

**Hybridization**
- Microarray
- Genomic DNA
- Shearing, adapter ligation, PCR (optional)

**Hybridization, Ligation + PCR**
1. Digest genomic DNA.
2. Hybridize the HaloPlex probe library in the presence of the Barcode Primer Cassette. Hybridization results in gDNA fragment circularization and incorporation of barcodes and Ion Torrent sequencing motifs.
4. PCR amplify targeted fragments to produce a sequencing-ready, target-enriched sample.

**Multiplex PCR**
- Gene of interest

---

Hybridization

- **Workflow:**
  - Library construction (4 hrs)
  - Hybridization with probes (24-48 hrs)
  - PCR & Indexing (2 hrs)

- **DNA input:**
  - 1-3 ug

- **Time from DNA sample to NGS library:**
  - 2-3 days

Hybridization, Ligation + PCR

- **Workflow:**
  - Hybridization with probe (16 hrs)
  - Ligation (1 hr)
  - PCR & Indexing (2 hrs)

- **DNA input:**
  - 200-400 ng

- **Time from DNA sample to NGS library:**
  - 2 days

Multiplex PCR

- **Workflow:** (SIMPLE)
  - PCR amplification (3 hrs)
  - Library construction (4 hrs)

- **DNA input:** (LOW)
  - <100 ng

- **Time from DNA sample to NGS library:**
  - 1 days (RAPID TAT)
GeneRead DNAseq Gene Panel

- Multiplex PCR technology based targeted enrichment for DNA sequencing
- Cover all human exons (coding region + UTR)
- Division of gene primers sets into 4 tubes; up to 1200plex in each tube
GeneRead DNAseq Gene Panel

Focus on your Disease of Interest

- Comprehensive Cancer Panel (124 genes)
- Cancer-specific Focused Gene Panels (20 genes)
  - Breast cancer
  - Colon Cancer
  - Gastric cancer
  - Leukemia
  - Liver cancer
  - Lung Cancer
  - Ovarian Cancer
  - Prostate Cancer

How Genes on Panels Are Selected

- Biologically/Clinically relevant
  - Multiple Publicly accessible databases
- Technically relevant
  - Most frequently mutated genes
Breast Cancer Gene Panel

- Somatic: All
- Germline and Somatic: APC, BRCA1, BRAC2, TP53
- With approved FDA inhibitor: BRAF, EGFR, ERBB2
Gene Read DNASeq Gene Panel

>10 years of proven expertise in primer design: GeneRead algorithm

- Simple
  - Any gene
  - Any sequencer
  - Any sample: average amplicon size is 145 bp

- Scalable
  - High multiplexing: Up to 1200 primer pairs per well
  - High design rate: >90%

- Superior
  - High specificity: >85%
  - High uniformity (0.1X of median coverage depth): >85%
GeneRead DNAseq Custom Panel
Sample & Assay Technologies

Streamlined Sample-to-Result Workflow

- **DNA extraction**: QIAamp DNA Mini Kit or QIAamp FFPE Tissue Kit
- **Target enrichment**: GeneRead DNAseq Gene Panel
- **Library construction**: GeneRead Library Prep Kit
- **Library quantification**: GeneRead Library Quantification System
- **NGS analysis**: NGS platform
- **Sequence analysis**: GeneRead SeqVariant Analysis software
- **What’s next**: qBiomarker Somatic Mutation Assays PyroMark Assays
GeneRead DNAseq Gene Panel: Simple protocol

1. **DNA extraction**
2. **Target enrichment**
3. **Library construction**
4. **Library quantification**
5. **NGS analysis**
6. **Sequence analysis**
7. **What’s next**

- GeneRead DNAseq Gene Panel PCR primer mix + control primers
- Add GeneRead Mastermix and genomic DNA (20 ng/rxn)
- PCR amplification
- Pool reactions for each sample and purify (QIAquick PCR Purification Kit)
GeneRead DNA Library Prep Kit

DNA extraction → Target enrichment → Library construction → Library quantification → NGS analysis → Sequence analysis → What’s next

≥ 50 ng purified amplicons

End Repair
50µl
30min @ 20°C
15min @ 65°C*

A-Tailing
50µl + 10µl
30min @ 37°C
10min @ 65°C*

Adaptor Ligation
50µl + 10µl + 60µl
15min @ 20°C

Clean-up + Size Selection
30min @ RT
15µl Eluate

HiFi Library Amplification
5 cycles

One Tube
Three Steps
Total Time = 100min
Hands-on Time = 5min

Purified amplicons to ready-to-use library: 130min

*: Enzyme Inactivation
GeneRead Size Selection Kit

- DNA extraction
- Target enrichment
- Library construction
- Library quantification
- NGS analysis
- Sequence analysis
- What’s next

Mix solution with prepared library

- Incubate
- Precipitate DNA on column
- Wash & centrifuge to dry
- Elute

Before size selection

GeneRead size selection

- Novel, simple spin column based size selection
- Remove small DNA fragments & adapter-dimers
- No tedious bead handling & risk of ethanol carry-over
GeneRead DNAnseq Library Quant Array

- **DNA extraction**
- **Target enrichment**
- **Library construction**
- **Library quantification**
- **NGS analysis**
- **Sequence analysis**
- **What's next**

**Serial dilution of DNA standard with primer assays specific for library quantification**

<table>
<thead>
<tr>
<th>Well</th>
<th>Standard 1</th>
<th>Standard 1</th>
<th>Standard 1</th>
<th>Standard 1</th>
<th>Standard 2</th>
<th>Standard 2</th>
<th>Standard 2</th>
<th>Standard 2</th>
<th>Standard 3</th>
<th>Standard 3</th>
<th>Standard 3</th>
<th>Standard 3</th>
<th>Standard 4</th>
<th>Standard 4</th>
<th>Standard 4</th>
<th>Standard 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
<td>NTC</td>
</tr>
<tr>
<td>B</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
<td>Library 1</td>
</tr>
<tr>
<td>C</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
<td>Library 2</td>
</tr>
<tr>
<td>D</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
<td>Library 3</td>
</tr>
<tr>
<td>H</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
<td>dilution 1</td>
</tr>
</tbody>
</table>

**primer assays for library QC**

**primer assays for Library Quantification**
GeneRead DNASeq Library Quant Array

QC Score
1-4 4-8 >8
Results Pass Marginal Fail
Proceed Caution! Do not proceed

Dilute sample libraries (L = library)
Add diluted sample libraries (for sample wells) or H₂O (for standard wells) to GeneRead qPCR SYBR Green Mastermix
Aliquot reactions to plate and run PCR
Template Preparation and Sequencing

- DNA extraction
- Target enrichment
- Library construction
- Library quantification
- NGS analysis
- Sequence analysis
- What’s next
Data Analysis: Free, Complete and Easy to Use

QIAGEN Next Generation Sequencing Data Analysis (β)

- Add file
- Start upload
- Cancel upload

- .sff file (Ion Torrent)
- .fastq file (MiSeq/HiSeq)
- GeneRead catalog #
- Sequencing platform
  - Ion PGM
  - Illumina
- Analysis mode
  - Somatic
  - Germline
- Read mode
  - Single end
  - Pair end

- Run summary (text)
- Summary by gene (text)
- Variants report (Excel)

Note: While uploading, please don’t refresh the browser or navigate to other pages.
Genetic variant analysis in FFPE lung adenocarcinoma samples

Experimental Design

- gDNA isolated from 3 FFPE lung adenocarcinoma and one FFPE normal lung samples
- GeneRead Lung Cancer Gene Panel was used to enrich 20 genes
- Library preparation, quantification and sequencing
- QIAGEN NGS Data Analysis Web Portal
## Genetic variant analysis in FFPE lung adenocarcinoma samples

### Run Summary

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tumor Sample-1</th>
<th>Tumor Sample-1</th>
<th>Tumor Sample-1</th>
<th>Normal Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of reads</td>
<td>1,629,983</td>
<td>1,892,621</td>
<td>1,851,514</td>
<td>1,625,119</td>
</tr>
<tr>
<td>Specificity (Reads on target)</td>
<td>92%</td>
<td>92%</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Median Coverage</td>
<td>608</td>
<td>744</td>
<td>703</td>
<td>577</td>
</tr>
<tr>
<td>Uniformity (regions with &gt;10% of median coverage)</td>
<td>93%</td>
<td>92%</td>
<td>90%</td>
<td>93%</td>
</tr>
<tr>
<td>No. of SNPs</td>
<td>84</td>
<td>72</td>
<td>87</td>
<td>83</td>
</tr>
<tr>
<td>No. of Indels</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>
Genetic variant analysis in FFPE lung adenocarcinoma samples

- Snap shot of variant report

```
<table>
<thead>
<tr>
<th>Chrom</th>
<th>Pos</th>
<th>rsID</th>
<th>Gene Name</th>
<th>Mutation type</th>
<th>Codon Change</th>
<th>AA Change</th>
<th>Filtered Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr7</td>
<td>116436097</td>
<td>rs41737</td>
<td>COSM150</td>
<td>SNP</td>
<td>c.416G&gt;A</td>
<td>p.P1382</td>
<td>24</td>
</tr>
<tr>
<td>chr7</td>
<td>116437606</td>
<td>rs1621</td>
<td>A</td>
<td>MET</td>
<td>SNP</td>
<td>1,283</td>
<td></td>
</tr>
<tr>
<td>chr8</td>
<td>21968199</td>
<td>rs11515</td>
<td>COSM142C</td>
<td>SNP</td>
<td>SNP</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>25362217</td>
<td>rs9266</td>
<td>A</td>
<td>KRAS</td>
<td>SNP</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>25362526</td>
<td>rs712</td>
<td>C</td>
<td>KRAS</td>
<td>SNP</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>25362646</td>
<td>rs4362222</td>
<td>C</td>
<td>KRAS</td>
<td>SNP</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>chr12</td>
<td>105236287</td>
<td>rs35416681</td>
<td>A</td>
<td>AKT1</td>
<td>SNP</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>chr14</td>
<td>105224966</td>
<td>rs2494735</td>
<td>A</td>
<td>TP53</td>
<td>SNP</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>chr17</td>
<td>7579221</td>
<td>rs121913343</td>
<td>G</td>
<td>TP53</td>
<td>SNP</td>
<td>403</td>
<td></td>
</tr>
</tbody>
</table>
```

- dbSNP and COSMIC ID (hyperlink)
- Predicted amino acid change

```
<table>
<thead>
<tr>
<th>Variant Frequency</th>
<th>snpEff Effect</th>
<th>snpEff Impact</th>
<th>COSMIC Sample Count</th>
<th>Somatic Mutation Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.708 SYNONYMOUS CODING</td>
<td>LOW</td>
<td>MODIFIER</td>
<td>1</td>
<td>COSM150379</td>
</tr>
<tr>
<td>0.839 UTR 3 PRIME</td>
<td>MODIFIER</td>
<td>1</td>
<td>COSM14251</td>
<td></td>
</tr>
<tr>
<td>0.942 UTR 3 PRIME</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
<tr>
<td>0.935 UTR 3 PRIME</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
<tr>
<td>0.987 SYNONYMOUS CODING</td>
<td>LOW</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
</tr>
<tr>
<td>0.157 UTR 3 PRIME</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
<tr>
<td>0.791 INTRON</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
<tr>
<td>0.117 NON SYNONYMOUS CODING</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
<tr>
<td>0.956 UTR 3 PRIME</td>
<td>MODIFIER</td>
<td>10</td>
<td>COSM99933</td>
<td></td>
</tr>
</tbody>
</table>
```

- Effect of SNP
- Impact of SNP
- Link to qPCR somatic mutation assay
Genetic variant analysis in FFPE lung adenocarcinoma samples

Variant Filtering

- Remove variants also found in normal FFPE sample
- Variants impact filtering (in coding region, altering protein identity)
- High quality and impact variants, specific for cancer samples
**Genetic variant analysis in FFPE lung adenocarcinoma samples**

Detection of low frequency variant in lung adenocarcinoma

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gene</th>
<th>Codon change</th>
<th>Variant frequency (%)</th>
<th>AA change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor sample 1</td>
<td>KRAS</td>
<td>c.35G&gt;T</td>
<td>38</td>
<td>p.G12V</td>
</tr>
<tr>
<td></td>
<td>BRAF</td>
<td>c.462A&gt;C</td>
<td>15</td>
<td>p.K154N</td>
</tr>
<tr>
<td></td>
<td>PDGFRA</td>
<td>c.1432T&gt;C</td>
<td>54</td>
<td>p.S478P</td>
</tr>
<tr>
<td></td>
<td>TP53</td>
<td>c.817C&gt;T</td>
<td>24</td>
<td>p.R273C</td>
</tr>
<tr>
<td>Tumor sample 2</td>
<td>ERBB2</td>
<td>c.3590C&gt;G</td>
<td>9</td>
<td>p.P1197R</td>
</tr>
<tr>
<td></td>
<td>KRAS</td>
<td>c.35G&gt;T</td>
<td>30</td>
<td>p.G12V</td>
</tr>
<tr>
<td>Tumor sample 3</td>
<td>EGFR</td>
<td>c.1496G&gt;A</td>
<td>7</td>
<td>p.C499Y</td>
</tr>
<tr>
<td></td>
<td>KRAS</td>
<td>c.35G&gt;T</td>
<td>23</td>
<td>p.G12V</td>
</tr>
<tr>
<td></td>
<td>PDGFRA</td>
<td>c.1432T&gt;C</td>
<td>59</td>
<td>p.S478P</td>
</tr>
<tr>
<td></td>
<td>TP53</td>
<td>c.817C&gt;T</td>
<td>11</td>
<td>p.R273C</td>
</tr>
</tbody>
</table>
Validation of KRAS:G12V somatic mutation by pyrosequencing assay

Mutation analysis of codons 12 of KRAS using PyroMark Q24. Upper Pyrogram show G to T mutation in position 1 of codon 12 of KRAS in lung adenocarcinoma sample. The mutation rate (35%) is similar to the NGS results (38%) confirming the reliability of GeneRead DNAseq Gene Panel. The lower Pyrogram shows normal genotype in normal sample-1.
Streamlined Sample-to-Result Workflow

DNA extraction: QIAamp DNA Mini Kit or QIAamp FFPE Tissue Kit

Target enrichment: GeneRead DNAseq Gene Panel

Library construction: GeneRead Library Prep Kit
GeneRead Size Selection kit

Library quantification: GeneRead Library Quantification System

NGS analysis: NGS platform

Sequence analysis: GeneRead SeqVariant Analysis software
qBiomarker Somatic Mutation Assays
PyroMark Assays

What's next:

Gene Reader Eco Solution

Streamlined Standardized Automated Sample-to-Result Workflow

2013
Upcoming webinars

Next Generation Sequencing and its role in cancer biology

Webinar 3:  Next-generation sequencing data analysis for genetic profiling
Date:       March 25, 2013
Speaker:    Ravi Vijaya Satya, Ph.D.

Webinar 1:  Next-generation sequencing, an introduction to technology and applications
Date:       April 4, 2013
Speaker:    Quan Peng, Ph.D.