Discover More, Sequence Less with SeqCap Products

John C. Tan, Ph.D.
Roche NimbleGen, Research & Development
Roche NimbleGen’s Target Enrichment Solutions

Overview

Why Targeted Sequencing

Technology Overview

• Workflow Summary
• NimbleGen Probe Library
• SeqCap Designs

SeqCap EZ Data

SeqCap EPI Overview

SeqCap EPI Data
Why Targeted Sequencing?
Faster discovery, more relevant content

**Discovery**

Discover more, quickly
- Higher Speed to discovery
- Lower Cost per study
- Larger Sample Size leading to more certain conclusions

**Clinical Research**

Focus on relevant and actionable targets
- Content specific to certain conditions
- Timely results
- Easy data interpretation

*Example: Exome (1% of genome) harbors >85% of known disease causing mutations.*

*Example: Well defined gene panels for cancer and neurology or your regions/ diseases of interest.*
## Different Enrichment Approaches

Solutions for both exome and gene panel sequencing

<table>
<thead>
<tr>
<th></th>
<th><strong>Whole-genome sequencing</strong></th>
<th><strong>Exome sequencing</strong></th>
<th><strong>Gene panels</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Applications</strong></td>
<td><strong>Discovery</strong>-discover novel variants for diseases or traits</td>
<td><strong>Discovery/Clinical research</strong>-discover variants in coding regions</td>
<td><strong>Clinical research</strong>-identify variants in disease related genes</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Unbiased, comprehensive coverage, easy sample prep</td>
<td>Cost effective, fast analysis</td>
<td>Most relevant content, quick to results, easy data interpretation</td>
</tr>
<tr>
<td><strong>Drawbacks</strong></td>
<td>Expensive, slow in sequencing, difficult to analyze</td>
<td>Lack coverage of non-coding regions, extra sample prep steps</td>
<td>Limited to knowledge at time of design, extra sample prep steps</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>6 Gb =1 sample</td>
<td>60 Mb =1 sample</td>
<td>5 Kb -5 Mb (depending on # of genes)</td>
</tr>
<tr>
<td></td>
<td>6 Gb = 100 samples</td>
<td>6 Gb/1,000-10,000 samples</td>
<td></td>
</tr>
<tr>
<td><strong>Cost per sample (approx)</strong></td>
<td>$10,000</td>
<td>&lt;$1,000</td>
<td>$50-$500</td>
</tr>
<tr>
<td><strong>Time from sample prep to interpreted data</strong></td>
<td>10 days</td>
<td>3-5 days</td>
<td>2-5 days</td>
</tr>
</tbody>
</table>
Roche NimbleGen’s Target Enrichment Solutions

Overview

Why Targeted Sequencing

Technology Overview

• Workflow
• NimbleGen Probe Library
• SeqCap Designs

SeqCap EZ Data

SeqCap EPI Overview

SeqCap EPI Data
SeqCap EZ Library Workflow

Streamlined workflow with KAPA Library Preparation

<table>
<thead>
<tr>
<th>Target Regions</th>
<th>Prepare KAPA Library with Next-Gen Sequencing Adaptors</th>
<th>SeqCap EZ Probes (Solution Capture)</th>
<th>Amplify DNA and Enrichment QC</th>
<th>Sequence DNA on a Next-Gen Sequencer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Genomic DNA**

- **Library Preparation**
  - KAPA Library Preparation Kits
  - SeqCap EZ Adapter kits: 24 indices + LM-PCR primers
  - SeqCap EZ Accessory kit v2: KAPA HiFi Hot Start Ready Mix, Cot-1 DNA & PCR reagents

- **Hybridization**

- **Capture and Washing**

- **Amplification and QC**

- **Sequencing**

**Virtual Reagent Kit v2**: Accessory kit + Hyb/Wash kit + HE Oligo kit A

**Virtual Reagent Kit Plus v2**: Virtual Reagent Kit + Bead kit

**Fully validated protocol to cover the whole workflow with maximum flexibility**
Newest NimbleGen products

- December 2013: KAPA DNA library prep kits and SeqCap Adapter kits

- January 2014: SeqCap Epi Enrichment System
  - Targeted enrichment platform for DNA methylation research
  - More on this platform later after SeqCap EZ
NimbleGen’s Probe Design

More probes, superior design, uniform coverage, better capture

Simple Tiling Design

Target Region

120mer probes/baits tiled across the region.

NimbleGen’s Sequence Capture Design

Target Region

Up to 2,100,000 (50-105mer) probes selected for the region using special algorithm.

Significantly more probes.

Benefits:

- Higher density tiling
- Redundancy to reduce risk of unbalanced coverage or missed regions
- Ability to move probes for better uniformity
**SeqCap Probe Design**

*More probes reduce risk of missing regions*

---

**Simple Tiling Design**

Small insertion

- Target Region

  - Missing Region

---

**NimbleGen Sequence Capture Design**

Small insertion

- Target Region

  - Flanking probes can capture novel variants

- Target Region

  - Redundant probes reduce risks

---

*Roche*
Third party studies comparing target enrichment platforms

  - “[T]he densely packed, overlapping baits of the NimbleGen SeqCap EZ Exome demonstrate the highest efficiency target enrichment, able to adequately cover the largest proportion of its targeted bases with the least amount of sequencing.”

  - “[T]he NimbleGen platform had better capture efficiency and could provide a higher proportion of CDs with high quality genotype assignments (thus higher completeness of SNP detection), and required lower sequence coverage because of its greater evenness.”

Discover More, Sequence Less
Comprehensive solutions with exome & custom designs

<table>
<thead>
<tr>
<th>Products</th>
<th>Design Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeqCap EZ Exome Library v2.0</td>
<td>44 Mb</td>
</tr>
<tr>
<td>SeqCap EZ Exome Library v3.0</td>
<td>64 Mb</td>
</tr>
<tr>
<td>SeqCap EZ Exome +UTR Library</td>
<td>96 Mb</td>
</tr>
<tr>
<td>SeqCap EZ Exome Plus Library</td>
<td>Up to 114 Mb</td>
</tr>
<tr>
<td>Human 50 Mb UTR Design</td>
<td>50 Mb</td>
</tr>
<tr>
<td>HGSC VCRome</td>
<td>45.2 Mb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Products</th>
<th>Design Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeqCap EZ Choice Library</td>
<td>Up to 7Mb</td>
</tr>
<tr>
<td>SeqCap EZ Choice XL Library</td>
<td>Over 7Mb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>Total Genes/ Design Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive Cancer Design</td>
<td>578 genes</td>
</tr>
<tr>
<td>Neurology Panel Design</td>
<td>256 genes</td>
</tr>
<tr>
<td>Human MHC Design</td>
<td>~5 Mb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Products</th>
<th>Design Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeqCap Epi CpGiant Enrichment Kit</td>
<td>84 Mb</td>
</tr>
<tr>
<td>SeqCap Epi Choice Enrichment Kit</td>
<td>Up to 90 Mb</td>
</tr>
<tr>
<td>SeqCap Epi Developer Enrichment Kit</td>
<td>Up to 210 Mb</td>
</tr>
</tbody>
</table>

Create your own design at: http://www.nimblegen.com/nimbledesign
Our Technology Enables Novel Research

Sample publications using SeqCap EZ products

Whole-exome sequencing identifies mutations in the nucleoside transporter gene SLC29A3 in dysostosclerosis, a form of osteopetrosis

Mutations of ANK3 Identified by Exome Sequencing Associated with Autism Susceptibility

Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia

Noninvasive Prenatal Measurement of the Fetal Genome

Oncogenic CSF3R Mutations in Chronic Neutrophilic Leukemia and Atypical CML

Exome Capture Reveals ZNF423 and CEP164 Mutations, Linking Renal Ciliopathies to DNA Damage Response Signaling
Roche NimbleGen’s Target Enrichment Solutions

Overview

Why Targeted Sequencing

Technology Overview

• Workflow
• NimbleGen Probe Library
• SeqCap Designs

SeqCap EZ Data

SeqCap EPI Overview

SeqCap EPI Data
SeqCap EZ Data

- SeqCap Workflow performance
  - High/low GC targets
  - Reduced DNA input amounts
- Circulating tumor DNA detection
High and Low–G/C targets

Uniform performance across a wider range of GC contents

Experiment:

- 2 replicate library preps – previously recommended workflow
- 2 replicate library preps – KAPA Library Preparation workflow
- Sequence Capture separately with SeqCap EZ Exome v3 probes
- 2x100bp PE sequencing (HiSeq2000)
- Analysis pipeline incorporating SOAPaligner (v2.18) (http://soap.genomics.org.cn/)
- Measure coverage over capture targets according to % G+C content of the capture target
High and Low–G/C targets

Uniform performance across a wider range of GC contents

Coverage Comparisons Across GC Fractions of EZ Exome v3

Previous Workflow- Replicate 1
Previous Workflow- Replicate 2
High and Low–G/C targets

Uniform performance across a wider range of GC contents

Coverage Comparisons Across GC Fractions of EZ Exome v3
High and Low–G/C targets

Uniform performance across a wider range of GC contents

Coverage Comparisons Across GC Fractions of EZ Exome v3
Sequence Capture with Low Sample Input

Uniform performance across a wide range of DNA input

Experiment:

- Kapa library prep workflow
- 1μg, 100ng, 10ng DNA input (HapMap sample NA12891)
  - Increase PCR cycles and reduce adapter concentration when using lower input amount
- Sequence Capture with SeqCap EZ Exome v3 probes - 2 replicates for each DNA amount
- 2x100bp PE sequencing (HiSeq2000)
- Analysis pipeline incorporating SOAPaligner (v2.18) and SOAPsnp (Crossbow implementation of v1.01) (http://soap.genomics.org.cn/)
Sequence Capture with Low Sample Input

Uniform performance across a wide range of DNA input

Percent Bases Covered of Exome v3 for Various DNA Input Levels

Randomly sampled to 75 Million 2x100 Reads

(sample NA12891, 29179 known SNPs, 2x100 bp reads)
## Sequence Capture with Low Sample Input

*Uniform performance across a wide range of DNA input*

### Coverage Metrics – Exome v3

<table>
<thead>
<tr>
<th>DNA input amount</th>
<th>All reads (M)</th>
<th>Raw reads mapped (%)</th>
<th>Mapped duplicates (%)</th>
<th>Mapped, unique reads on target (%)</th>
<th>Average coverage</th>
<th>Median coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ug</td>
<td>122</td>
<td>80.9</td>
<td>3.1</td>
<td>77.7</td>
<td>92.7</td>
<td>84.6</td>
</tr>
<tr>
<td>100 ng</td>
<td>119</td>
<td>80.7</td>
<td>6.8</td>
<td>80.5</td>
<td>89.8</td>
<td>83.0</td>
</tr>
<tr>
<td>10 ng</td>
<td>126</td>
<td>80.4</td>
<td>11.8</td>
<td>81.5</td>
<td>90.7</td>
<td>84.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampled reads (M)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ug</td>
<td>75</td>
<td>80.9</td>
<td>2.1</td>
<td>77.7</td>
<td>57.5</td>
<td>52.2</td>
</tr>
<tr>
<td>100 ng</td>
<td>75</td>
<td>80.7</td>
<td>4.6</td>
<td>80.6</td>
<td>57.9</td>
<td>53.0</td>
</tr>
<tr>
<td>10 ng</td>
<td>75</td>
<td>80.4</td>
<td>7.7</td>
<td>81.7</td>
<td>56.6</td>
<td>52.2</td>
</tr>
</tbody>
</table>
Sequence Capture with Low Sample Input
Uniform performance across a wide range of DNA input

**SNP Calling Metrics**

<table>
<thead>
<tr>
<th>DNA input amount</th>
<th>Raw reads (M)</th>
<th>Filtered SNPs (≥4x, Q20)</th>
<th>Het True Positive Rate (%)</th>
<th>Hom True Positive Rate (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ug</td>
<td>122</td>
<td>56,214</td>
<td>99.4</td>
<td>99.2</td>
<td>97.9</td>
<td>99.3</td>
</tr>
<tr>
<td>100 ng</td>
<td>119</td>
<td>55,982</td>
<td>99.3</td>
<td>99.2</td>
<td>97.9</td>
<td>99.3</td>
</tr>
<tr>
<td>10 ng</td>
<td>126</td>
<td>56,077</td>
<td>99.4</td>
<td>99.2</td>
<td>97.9</td>
<td>99.3</td>
</tr>
<tr>
<td>Sampled reads (M)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 ug</td>
<td>75</td>
<td>54,863</td>
<td>99.4</td>
<td>99.2</td>
<td>97.5</td>
<td>99.3</td>
</tr>
<tr>
<td>100 ng</td>
<td>75</td>
<td>54,908</td>
<td>99.3</td>
<td>99.3</td>
<td>97.5</td>
<td>99.3</td>
</tr>
<tr>
<td>10 ng</td>
<td>75</td>
<td>54,696</td>
<td>99.5</td>
<td>99.2</td>
<td>97.4</td>
<td>99.4</td>
</tr>
</tbody>
</table>
Quantitating circulating tumor DNA

- CAPP-Seq: Cancer personalized profiling by deep sequencing
- Created a NimbleGen design for non-small-cell lung cancer
- Detect mutant alleles down to ~0.02%
Roche NimbleGen’s Target Enrichment Solutions

Overview

Why Targeted Sequencing

Technology Overview

• Workflow
• NimbleGen Probe Library
• SeqCap Designs

SeqCap EZ Data

SeqCap EPI Overview

SeqCap EPI Data
DNA Methylation
Essential part of genomics research

Gene Expression
Dosage Compensation

Imprinting
Genome Stability
Development

Nature 441, 143-145 (11 May 2006)
DNA Methylation & Human Health

Altered methylation found in virtually every type of tumor

- Cytosine methylation is the most common post-replicative DNA modification in animals and plants
  - Essential for mammalian development
  - Essential for genomic integrity in plants
- Altered DNA methylation linked to Cancer
  - > 4,000 publications
  - >1,200 in past three years

(Jones and Baylin, 2002)
5mC monitoring technology

Bisulfite mediated deamination of cytosine

- Bisulfite is the “Gold” standard technology for enabling single base resolution
- Uracil is replaced with thymine by DNA Pol during amplification ("T" read as unmeth C)
- BOTH 5mC and 5hmC are substantially resistant to the mutagenesis conditioned by HSO$_3^-$
- “C” interpreted as either 5mC or 5hmC in the pre-treated DNA
- Destroys as much as ~95% of the sample via acid-hydrolysis

Whole Genome Bisulfite Sequencing (WGBS)

Genome-wide methods

<table>
<thead>
<tr>
<th>Genome Analyser II Reads* (Illumina, 2x36):</th>
<th>H1 human ES cells</th>
<th>IMR90 fetal lung fibroblasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence yield:</td>
<td>1.16 billion</td>
<td>1.18 billion</td>
</tr>
<tr>
<td>Average read depth (per strand):</td>
<td>87.5 Gb**</td>
<td>91.0 Gb</td>
</tr>
<tr>
<td></td>
<td>14.2×</td>
<td>14.8×</td>
</tr>
</tbody>
</table>

- >86% of both strands covered by at least one sequence read; covered 94% of the cytosines in the hg18 genome
- ~25% of DNA methylation in H1 cells found in non-CG contexts, mCHG and mCHH (H = A, C or T),

* 376 total lanes; ** ~3 HiSeq 2500 lanes (2x100)
Lister R et al., Nature. 2009 Nov; doi:10.1038/nature08514
Gaps and Limitations of Current Technologies
Issues with breadth, depth and throughput

Current technologies have a series of limitations

- **WGBS (whole genome bisulfite sequencing):**
  - Cost-prohibitive
  - Poor depth of coverage
  - Complex data analysis

- **Microarrays:**
  - Limited SNP calling on CpG
  - Missing allelic info

Big gaps in current enrichment tools for bisulfite-treated DNA

- **Current enrichment products have drawbacks:**
  - Company A: Limited molecular complexity
    - High PCR duplication
    - High sample input
    - Targets only one strand
  - RRBS: Fixed content limited by enzyme sites
    - Missing data

- **No effective custom solution**
Bisulfite Sequencing: Challenges of Targeting Approaches

More complex than standard genomics analysis

ATCGTCTAGCGCGAAT
TAGCAGATCGCGCTTA

Bisulfite Conversion

m

ATUGTUTAGUGUGAAT
TAGUAGATUGUGUTTA

ATCGTCTAGCGCGAAT
TAGCAGATCGCGCTTA

PCR amplification

m

m

m

ATTTTTAGTTGAAT
TAACAAATCACACTTA

ATCGTTTAGCGCGAAT
TAGCAAAATCGCGCTTA

ATCATCTAACACAAAT
TAGTAGATTGTTTA

ATCGTCTAACGCGAAT
TAGCAGATTGGGCTTA
SeqCap Epi Enrichment System Workflow

Bisulfite treatment and then capture

Target Regions | Prepare a SeqCap Epi Library | Bisulfite Conversion & Amplification | SeqCap Epi (solution capture probes target CG, CHG, CHH) | Post-Capture Amplification of Library | Sequence (MiSeq or HiSeq)

Genomic DNA | Library Preparation with methylated adapters | Chemical Mutagenesis (C>T conversion) and amplification with uracil tolerant polymerase | Hybridization | Capture and Washing | Amplification and QC | Sequencing

SeqCap Epi Accessory kit: KAPA HiFi Uracil+ ReadyMix, Bisulfite Capture Enhancer, KAPA HiFi HotStart ReadyMix, & Bisulfite Conversion Control

Focus bisulfite sequencing to any C of interest
Capture *before* bisulfite conversion leads to more severe bottlenecking of genomic information.

**SeqCap Epi**
- MethylSeq Library Prep
- Bisulfite Conversion
- Pre-Capture Amplify Library
- Capture
- Post-Capture Amplify Library
- Sequence

**Alternative**
- MethylSeq Library Prep
- Capture
- Bisulfite Conversion
- Capture
- Post-Capture Amplify Library
- Sequence
Capture *before* bisulfite conversion leads to more severe bottlenecking of genomic information.

**SeqCap Epi**
- MethylSeq Library Prep
- Bisulfite Conversion
- Pre-Capture Amplify Library
- Capture
- Post-Capture Amplify Library
- Sequence

**Alternative**
- MethylSeq Library Prep
- Capture
- Bisulfite Conversion
- Post-Capture Amplify Library
- Sequence

100% input complexity

tighter bottleneck
Capture before bisulfite conversion leads to more severe bottlenecking of genomic information.

**SeqCap Epi**
- MethylSeq Library Prep
- Bisulfite Conversion
- Pre-Capture Amplify Library
- Capture
- Post-Capture Amplify Library
- Sequence

+ Higher output complexity
+ Lower sample input required
+ Higher reproducibility
- Higher probe density required

**Alternative**
- MethylSeq Library Prep
- Capture
- Bisulfite Conversion
- Capture
- Post-Capture Amplify Library
- Sequence

- Lower output complexity
- Higher sample input required
- Lower reproducibility
+ Lower probe density required

Tighter bottleneck before bisulfite conversion leads to more severe bottlenecking of genomic information.
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

Probes designed to fully-methylated targets

Probes designed to fully un-methylated targets

Probes designed to the flanks of methylate-able targets
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

- Probes designed to fully methylated targets
- Probes designed to fully un-methylated targets
- Probes designed to the flanks of methylate-able targets

Probes:
- NNNNNNNCCCCCN
- NNNNNNNCTCCCC
- NNNNNNNCCCTCC
- NNNNNNTCTCTC
- NNNNNNTCTTTT
- NNNNNNTCTTTT
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

Probes designed to fully-methylated targets

Probes designed to fully un-methylated targets

Probes designed to the flanks of methylate-able targets

NNNNNNNNNNNNNNN
NNNNNNCTCCCCCCNNNNNN
NNNNNNCCCTCCNNNNNN

NNNNNNCTCTCTCNNNNNN
NNNNNNCCCTTTTNNNNNN
NNNNNNTTTCCCNNNNNN

NNNNNNTTTTTTTNNNNNN
NNNNNNTTTTTCTNNNNNN
NNNNNNTTCTTTTNNNNN
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

Probes designed to fully methylated targets

Probes designed to fully un-methylated targets

Probes designed to the flanks of methylate-able targets
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

Probes designed to fully-methylated targets

Probes designed to fully un-methylated targets

Probes designed to the flanks of methylate-able targets
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

Probes designed to fully-methylated targets

Probes designed to fully unmethylated targets

Probes designed to the flanks of methylate-able targets
Innovative Probe Design & Manufacture

SeqCap Epi Probe Pool

- Probes designed to fully-methylated targets
- Probes designed to fully un-methylated targets
- Probes designed to the flanks of methylate-able targets
Roche NimbleGen’s Target Enrichment Solutions

Overview

Why Targeted Sequencing

Technology Overview

• Workflow
• NimbleGen Probe Library
• SeqCap Designs

SeqCap EZ Data

SeqCap EPI Overview

SeqCap EPI Data
Benchmarking Technical Performance

**General strategies**

- **Assess reproducibility of measurement from samples**
  - Low technical variation from same sample compared means you can trust any differences observed between samples

- **Assess biological variation**
  - Compare a “normal” cell line to a cancer cell line
    - Compare results from NA12762 (CEPH) and NA04671 (Burkitt’s Lymphoma)
    - Expect increase in 5mC at CpG Island in cancer

- **Assess capacity of probe response across methylation range**
  - Use a hypomethylated sample (HCT116 DKO)
  - Full and intermediate methylation *in vitro*

- **Compare with Whole Genome Bisulfite Sequencing (“Gold Standard”)**

- **Compare with another target enrichment platform**
**Workflow Reproducibility**

*Low Technical Variation in Methylation Data*

- High fold enrichment (3.2 Mb target)
- Low duplicate read rate
- High correlation ($R^2$) between methylation data from same sample

<table>
<thead>
<tr>
<th>Sample</th>
<th>PF Reads aligned</th>
<th>% Reads on target</th>
<th>Fold enrichment</th>
<th>% Dup reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA04671_#1</td>
<td>4,427,673</td>
<td>53.1%</td>
<td>654.78</td>
<td>2</td>
</tr>
<tr>
<td>NA04671_#2</td>
<td>4,732,482</td>
<td>47.3%</td>
<td>583.32</td>
<td>5</td>
</tr>
<tr>
<td>NA04671_#3</td>
<td>4,990,449</td>
<td>51.3%</td>
<td>632.84</td>
<td>3</td>
</tr>
<tr>
<td>NA04671_#4</td>
<td>5,401,533</td>
<td>52.4%</td>
<td>646.29</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R^2$</th>
<th>NA04671 #1</th>
<th>NA04671 #2</th>
<th>NA04671 #3</th>
<th>NA04671 #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA04671 #1</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>NA04671 #2</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>NA04671 #3</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>NA04671 #4</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Burkitt’s Lymphoma (NA04671=4)

Data presented by J. Greally at ASHG Oct 2013
Benchmarking Technical Performance

General strategies

- Assess reproducibility of measurement from samples
  - Low technical variation from same sample compared means you can trust any differences observed between samples

- **Assess biological variation**
  - Compare a “normal” cell line to a cancer cell line
    - Compare results from NA12762 (CEPH) and NA04671 (Burkitt’s Lymphoma)
    - Expect increase in 5mC at CpG in cancer

- **Assess capacity of probe response across methylation range**
  - Use a hypomethylated sample (HCT116 DKO)
  - Full and intermediate methylation *in vitro*

- **Compare with Whole Genome Bisulfite Sequencing (“Gold Standard”)**

- **Compare with another target enrichment platform**
Assess Biological Variation
Reproducible DMR identification

Hyper Meth in Cancer

Primary Target
Probe Coverage
% G+C

#1
#2
#3
#4

NA12762
(HapMap)

NA04671
(Lymphoma)

CpG

Male CEPH (NA12762 =4) vs. Burkitt’s Lymphoma (NA04671=2)

3.2Mbp capture design targeting 500 genes’ promoters
(Developed with John Greally at Albert Einstein School of Medicine)

Data presented by J. Greally at ASHG Oct 2013
Benchmarking Technical Performance

General strategies

• Assess reproducibility of measurement from samples
  - Low technical variation from same sample compared means you can trust any differences observed between samples

• Assess biological variation
  - Compare a “normal” cell line to a cancer cell line
    • Compare results from NA12762 (CEPH) and NA04671 (Burkitt’s Lymphoma)
    • Expect increase in 5mC at CpGI in cancer

• Assess capacity of probe response across methylation range
  - Use a hypomethylated sample (HCT116 DKO)
  - Full and intermediate methylation in vitro

• Compare with Whole Genome Bisulfite Sequencing (“Gold Standard”)

• Compare with another target enrichment platform
Capacity of Probe Response Across Methylation Range
Capture all methylation states with confidence

- **Capture DNA from a hypomethylated human genome: HCT116 DKO**
  - Commercially available colon cancer cell line
  - Low methylation due to double KO
    - DNMT1 – CG methyltransferase
    - DNMT3A – CNG methyltransferase

- **Methylate HCT116 DKO *in vitro* with CpG Methyltransferase (M.SssI)**
  - Fully methylated state after 60 minute treatment
  - Intermediate methylated state after 15 minute treatment
Capacity of Probe Response Across Methylation Range

Capture all methylation states with confidence

Data presented by J. Greally at ASHG Oct 2013
Capacity of Probe Response Across Methylation Range

Capture all methylation states with confidence

Data presented by J. Greally at ASHG Oct 2013
Capacity of Probe Response Across Methylation Range

Capture all methylation states with confidence

(T = 60 min MSssl)  Data presented by J. Greally at ASHG Oct 2013
Capacity of Probe Response Across Methylation Range

Capture all methylation states with confidence

Data presented by J. Greally at ASHG Oct 2013

hg19 chr 7 q22.1 TAF6 CpG

(T= 0 + 60 min MSssl) Data presented by J. Greally at ASHG Oct 2013
Capacity of Probe Response Across Methylation Range

Capture all methylation states with confidence

hg19 chr 7 q22.1 TAF6 CpGI

(T= 15 min MSssl) Data presented by J. Greally at ASHG Oct 2013
Benchmarking Technical Performance

General strategies

- Assess reproducibility of measurement from samples
  - Low technical variation from same sample compared means you can trust any differences observed between samples

- Assess biological variation
  - Compare a “normal” cell line to a cancer cell line
    - Compare results from NA12762 (CEPH) and NA04671 (Burkitt’s Lymphoma)
    - Expect increase in 5mC at CpGI in cancer

- Assess capacity of probe response across methylation range
  - Use a hypomethylated sample (HCT116 DKO)
  - Full and intermediate methylation in vitro

- **Compare with Whole Genome Bisulfite Sequencing (“Gold Standard”)**

- **Compare with another target enrichment platform**
Compare with Whole Genome Bisulfite Sequencing

High Correlation with WGBS Data

- IMR90 cell line DNA
  - 2 x 100 HiSeq sequencing
  - 10x mean coverage WGBS
  - 30x mean coverage SeqCap Epi

- Left-right symmetry indicates little systematic bias in % methylation measurements
  - Correlation R = 0.929
Benchmarking Technical Performance

**General strategies**

- **Assess reproducibility of measurement from samples**
  - Low technical variation from same sample compared means you can trust any differences observed between samples

- **Assess biological variation**
  - Compare a “normal” cell line to a cancer cell line
    - Compare results from NA12762 (CEPH) and NA04671 (Burkitt’s Lymphoma)
    - Expect increase in 5mC at CpGI in cancer

- **Assess capacity of probe response across methylation range**
  - Use a hypomethylated sample (HCT116 DKO)
  - Full and intermediate methylation in vitro

- Compare with Whole Genome Bisulfite Sequencing (“Gold Standard”)

- **Compare with another target enrichment platform**
Direct Comparison

The Importance of Targeting Watson and Crick

- **Agilent SureSelect Human Methyl-Seq** (targets one strand)

- **Resulted in a 470x strand coverage imbalance (1 HiSeq lane per captured sample)**
Direct Comparison
The Importance of Targeting Watson and Crick

- **NimbleGen SeqCap Epi (targets both strands)**
- **Provides the capacity to identify the effect of sequence polymorphisms on 5mC calls**
The Importance of Targeting Watson and Crick

SNPs are still important in the epigenetic world

- The most common type of SNP between any two individuals is a C->T transition. Many of these occur at CpG (methylation sites)
- C->T transition @ CpG is 6.7 fold over-represented from expectation
- If a sample has a SNP at a CpG (i.e. TpG), sequence analysis will report “unmethylated” following bisulfite conversion
  - The only way to resolve this with confidence is to read the sequence on the other strand (A vs. G)
  - If a capture method targets only 1 strand, this cannot be resolved
  - SNP sites = ERRORS
Direct Comparison

SeqCap Epi Choice v. Agilent SureSelect Human Methyl-Seq

- Agilent Methyl-Seq
  - 84 Mb design covering CpG islands, CpG island shores, undermethylated regions, promoters, and DMRs
  - Followed manufacturers' user guide recommendations
    - 3 ug DNA input, XT Library Prep kits, Capture then Convert, etc..

- SeqCap Epi Choice L
  - 83.9 Mb design (Primary Capture Target) to mimic the Agilent Methyl-Seq design
  - Followed manufacturers' user guide recommendations
    - 1 ug DNA input, KAPA Library Prep kits, Convert then Capture, etc.

- Sequenced on Illumina HiSeq, 2x100 and direct comparison data subsampled to 55 million reads
## SeqCap EPI: RNG Clear Technology Leader

*Direct comparison with Agilent SureSelect Methyl-Seq*

<table>
<thead>
<tr>
<th></th>
<th>SeqCap Epi</th>
<th>SureSelect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capture Specificity:</strong></td>
<td>70%</td>
<td>90%</td>
</tr>
<tr>
<td><strong>Duplicate Read Rate:</strong></td>
<td>6%</td>
<td>20%</td>
</tr>
<tr>
<td><strong>Cov. Completeness (@10X):</strong></td>
<td>90%</td>
<td>86%</td>
</tr>
<tr>
<td><strong>Cov. Uniformity (1/nc80):</strong></td>
<td>~2</td>
<td>~3</td>
</tr>
<tr>
<td><strong>Strand Bias:</strong></td>
<td>0.5</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>CpG Surveyed (millions):</strong></td>
<td>3.43</td>
<td>3.02</td>
</tr>
</tbody>
</table>
SeqCap Epi CpGiant Enrichment Kits
Upgrade your research from microarrays

“CpGiant” is an epigenome-wide design targeting Illumina’s Infinium Human Methylation 450 Bead Chip sites with “smart” padding to maximize CpG

<table>
<thead>
<tr>
<th>Product comparison</th>
<th>SeqCap Epi CpGiant</th>
<th>Illumina Infinium 450K</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CpG’s Targeted</td>
<td>&gt; 5.5 Million</td>
<td>485,512</td>
</tr>
<tr>
<td>Discern SNPs from Methylation?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Discover Novel CpG’s?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Scalable</td>
<td>Process up to 96 samples at once</td>
<td>12 samples per array</td>
</tr>
</tbody>
</table>
SeqCap Epi CpGiant Enrichment Kits
Upgrade your research from microarrays

~4 samples per HiSeq2000 lane
Mean 30x coverage, 2x100 reads
SeqCap Epi Choice / Epi Developer Enrichment Kits

Custom human designs up to 210Mb

- First fully customizable target enrichment solution for bisulfite sequencing

- 4 sizes available, for human or non-human species
  - Small (<30Mb)
  - Medium (30-60Mb)
  - Large (60-90Mb)
  - Extra Large (90-210Mb)

- Designs created by Roche NimbleGen Bioinformatics experts
  - Just supply your genome build and coordinates
What’s next for SeqCap?

Targeted cDNA sequencing:


- We are working to optimize and simplify the workflow
- John will present data at the RNA-Seq Summit 2014 in Boston (July 2014)
NimbleGen SeqCap Products

*Discover More, Sequence Less*

- Discover more variants
- Minimize sequencing costs
- Work with more samples each time
- Capture YOUR regions of interest efficiently
- Manage your workflow with maximum convenience
Disclaimer and Trademark Statement

Unless explicitly stated otherwise, all Roche products and services referenced in this presentation/document are intended for the following use:

For life science research only.
Not for use in diagnostic procedures.

NIMBLEDESIGN, NIMBLEGEN and SEQCAP are trademarks of Roche.
KAPA is a trademark of Kapa Biosystems.
All other product names and trademarks are the property of their respective owners

© 2013 Roche NimbleGen, Inc.
Questions?
Doing now what patients need next
Appendix

- Product timeline
- Understanding target sequencing further
- Technology details
- 3rd party comparison details
- Multiplexing data
- Cost comparison with Illumina
- Automation solution
# SeqCap EZ Product Portfolio

**Continuous expansion for the evolving capture market**

<table>
<thead>
<tr>
<th>Year</th>
<th>Product Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td><strong>Exome v2.0</strong>&lt;br&gt;Focused content from well known databases.</td>
</tr>
<tr>
<td>2011</td>
<td><strong>Exome v3.0, Choice, Developer</strong>&lt;br&gt;Continuous improvement of known genes for the exome, largest exome capture; plus products for customers to study their regions of interest in the human genome or other organisms.</td>
</tr>
<tr>
<td>2012</td>
<td><strong>Developer Reagent, Exome +UTR, Exome Plus, Oncology, Neurology</strong>&lt;br&gt;Expanded portfolio of Exome products for largest capture, and custom designs covering oncology and neurology genes. Developer Reagent manufactured to make sequencing of non-human genomes more efficient; technical note applied to soy exome.</td>
</tr>
<tr>
<td>2013</td>
<td><strong>Reagents + Library Prep Kits</strong>&lt;br&gt;A more streamlined workflow solution, providing customer convenience; offering reagents for indexed library generation, library processing, oligos for blocking secondary capture and beads for capture and purification.</td>
</tr>
<tr>
<td>2014</td>
<td><strong>Epi Products: CpGiant, Choice, Developer &amp; Reagents</strong>&lt;br&gt;Expanding the SeqCap portfolio to epigenomic research through introduction of our novel bisulfite conversion before capture target enrichment line.</td>
</tr>
</tbody>
</table>
Targeted Exome vs. Whole-Genome Sequencing

More samples increase the power of research studies

- Sequence the exome at a fraction of the cost for a whole genome.
- Allow more samples per study by targeted sequencing at a lower per sample cost.
- Simplify data analysis by focusing on known functional units.

Whole Genome Sequencing
► Sequence 5 Genomes

OR

Targeted Exome Sequencing
► Sequence 50 Exomes

For life science research only. Not for use in diagnostic procedures.
Why Perform Targeted Resequencing?
Focus your research on relevant regions

- **Genetic Diseases:** identify mutations that correlate to disease or contribute to genetic risk.
  - Follow up of GWAS, re-sequence of disease associated regions
  - Exome sequencing for discovery of mutations in rare diseases
  - Candidate genes panels involved in disease pathways

- **Cancer:** identify somatic mutations involved in tumorigenesis or metastasis, discover germ line mutations that contribute to cancer risk.
  - Focus on exome or cancer gene panel for comprehensive coverage of known and rare mutations
  - Deep sequencing of cancer related regions to identify rare mutations present in heterogeneous tumor tissues
NimbleGen SeqCap EZ Library

Reduce sequencing costs – capture uniformity

Good depth  Average depth  Poor depth

Exon 1  Exon 2  Exon 3

........ Minimum data required

$
NimbleGen SeqCap EZ Library
Reduce sequencing costs – capture uniformity

Good depth | Average depth | Poor depth

Exon 1 | Exon 2 | Exon 3

••••• Minimum data required
NimbleGen SeqCap EZ Library

Reduce sequencing costs – capture uniformity

Good depth  Average depth  Poor depth

Exon 1  Exon 2  Exon 3

· · · · · Minimum data required
NimbleGen SeqCap EZ Library - summarize

Capture and sequencing uniformity reduces sequencing costs

Exon 1  Exon 2  Exon 3

· · · · · Minimum data required
Exome Capture Comparison Study - Stanford

Sept 25, 2011

Performance comparison of exome DNA sequencing technologies

Clark et al., Performance comparison of exome DNA sequencing technologies (2011)
Nature Biotechnology Published online 25 September 2011 doi:1038/nbt.1975

Exome Capture Comparison Study – BGI
Sept 28, 2011

Research

Comprehensive comparison of three commercial human whole-exome capture platforms
[no first name] Asan, Yu Xu, Hui Jiang, Chris Tyler-Smith, Yali Xue, Tao Jiang, Jiawei Wang, Mingzhi Wu, Xiao Liu, Geng Tian, Jun Wang, Jian Wang, Huangming Yang and Xiuqing Zhang

• Study compared the performance of three capture technologies:
  • NimbleGen SeqCap EZ Exome Library v1.0
  • NimbleGen Sequence Capture 2.1M Array
  • Agilent SureSelect Exome

“…the NimbleGen platforms showed better uniformity of coverage and greater genotype sensitivity at 30-100 folds sequencing depth.”

Asan et al., Comprehensive comparison of three commercial human whole-exome capture platforms (2011)
Genome Biology Published online 28 September 2011, 12:R95
http://genomebiology.com/2011/12/9/R95/abstract
Exome Capture Comparison Study – FIMM
Sept 28, 2011

- Study compared the performance of two capture technologies:
  - NimbleGen SeqCap EZ Exome Library v2.0
  - Agilent SureSelect Exome

“In our data, libraries captured with NimbleGen kits aligned more accurately to the target regions. NimbleGen SeqCap v2.0 most efficiently covered the exome with a minimum coverage of 20x....”

Sulonen et al., *Comparison of solution-based exome capture methods for next generation sequencing* (2011)
Genome Biology Published online 28 September 2011, 12:R94

http://genomebiology.com/2011/12/9/R94/abstract
One Lane of HiSeq v3 sequenced at 100 bp reads and subsampled to 75 Million Reads
## SeqCap EZ Exome v3 Pre-Capture Multiplexing

*Equivalent post-capture and pre-capture performance*

<table>
<thead>
<tr>
<th></th>
<th>Post Capture – 3 samples</th>
<th>Pre Capture – 3 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>% On Target</td>
<td>~72%</td>
<td>~71%</td>
</tr>
<tr>
<td>Average Coverage</td>
<td>~50x</td>
<td>~46x</td>
</tr>
<tr>
<td>Median Coverage</td>
<td>~40x</td>
<td>~38x</td>
</tr>
<tr>
<td>% bases at ≥ 10 coverage</td>
<td>~93%</td>
<td>~93%</td>
</tr>
<tr>
<td>% PCR Duplicates</td>
<td>~1%</td>
<td>~2.5%</td>
</tr>
<tr>
<td>% Sensitivity</td>
<td>~97%</td>
<td>~97%</td>
</tr>
<tr>
<td>% Specificity</td>
<td>~99%</td>
<td>~99%</td>
</tr>
</tbody>
</table>

*One Lane of HiSeq v3 sequenced at 100 bp reads and subsampled to 75 Million Reads*
SeqCap EZ Exome v3 Pre-Capture Multiplexing

Balanced distribution of indexed reads

One Lane of HiSeq v3 sequenced at 100 bp reads and raw reads/not subsampled
SeqCap EZ Exome v3 Pre-Capture Multiplexing

Consistent performance shown by 7-plex pre-capture

Three separate HiSeq lanes - One Lane of HiSeq v3 sequenced at 100 bp reads and raw reads/not subsampled
## SeqCap EZ Exome v3 Pre-Capture Multiplexing

*Consistent performance: 3-plex vs 7-plex comparison*

<table>
<thead>
<tr>
<th></th>
<th>Pre Capture – 3 samples</th>
<th>Pre Capture – 7 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>% On Target</td>
<td>~71%</td>
<td>~72%</td>
</tr>
<tr>
<td>Average Coverage</td>
<td>~75x</td>
<td>~33x</td>
</tr>
<tr>
<td>Median Coverage</td>
<td>~60x</td>
<td>~26x</td>
</tr>
<tr>
<td>% bases at &gt; 10 coverage</td>
<td>~95%</td>
<td>~87%</td>
</tr>
<tr>
<td>% PCR Duplicates</td>
<td>~4%</td>
<td>~2.5%</td>
</tr>
<tr>
<td>% Sensitivity</td>
<td>~98%</td>
<td>~93%</td>
</tr>
<tr>
<td>% Specificity</td>
<td>~99%</td>
<td>~98%</td>
</tr>
</tbody>
</table>

*One Lane of HiSeq v3 sequenced at 100 bp reads and raw reads/not subsampled*
Comprehensive Cancer Design Performance

Sample multiplexing capabilities

Coverage of target bases at 75x, 100x, and 150x sequencing depths using Illumina 2x76 bp reads

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Illumina HiSeq (34 Gb*)</th>
<th>Illumina MiSeq (4 Gb*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75x</td>
<td>14 samples</td>
<td>2 samples</td>
</tr>
<tr>
<td>100x</td>
<td>12 samples</td>
<td>1 sample</td>
</tr>
<tr>
<td>150x</td>
<td>7 samples</td>
<td>-</td>
</tr>
</tbody>
</table>

*Estimates for sequencing throughput per lane is based on manufacturer’s specifications.
Neurology Panel Design Performance

Sample multiplexing capabilities

Coverage of target bases at 30x, 50x, and 100x sequencing depths using Illumina 2x76bp reads
Pre-capture Multiplexing with Illumina TruSeq Libraries

*Balanced distribution of indexed reads*

- 8 Samples (4 tumor/normal pairs)
- Each capture reaction was sequenced with 1 lane on HiSeq (2x100 bp paired-end)
Data suggests ~2.5x more sequencing with Illumina to obtain similar results to Roche NimbleGen

**SeqCap EZ Exome vs. Illumina TruSeq Exome Performance comparison**

<table>
<thead>
<tr>
<th></th>
<th>Illumina TruSeq Exome (62 Mb)</th>
<th>NimbleGen SeqCap EZ Exome (64 Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes of Sequencing per Sample</td>
<td>0.83 lanes</td>
<td>0.33 lanes</td>
</tr>
<tr>
<td>Average Coverage</td>
<td>~64x</td>
<td>~75x</td>
</tr>
<tr>
<td>% bases at &gt; 10-12.5 coverage*</td>
<td>~90%</td>
<td>~95%</td>
</tr>
</tbody>
</table>

Illumina TruSeq Exome data from [Illumina’s TruSeq Exome Data Sheet](#)
NimbleGen SeqCap EZ Exome data from Roche NimbleGen RnD

- Data suggests ~2.5x more sequencing with Illumina to obtain similar results to Roche NimbleGen

*10x coverage at 90% or greater are globally accepted metrics for new genetic disease discovery*
## SeqCap EZ Exome vs. Illumina TruSeq Exome

### Cost comparison

<table>
<thead>
<tr>
<th></th>
<th>Price per Sample (6 plex)</th>
<th>Sequencing price per sample</th>
<th>Price per sample including sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Illumina TruSeq Exome</strong></td>
<td>~$50</td>
<td>$1370</td>
<td>~$1420</td>
</tr>
<tr>
<td><strong>NimbleGen SeqCap Exome (48 pack)</strong></td>
<td>$83</td>
<td>$545</td>
<td>$628</td>
</tr>
</tbody>
</table>

- Due to NimbleGen performance advantages as shown in previous slide the list price per sample is more than twice as much for Illumina compared to NimbleGen ($1420 vs. $628)
  - Cost saving per sample is approximately $800. A Roche NimbleGen customer can do two samples for less than the price of one with Illumina

* Assuming the current cost of Illumina reagents at list price, $1650/ lane
SeqCap EZ Workflow

Automation using the Caliper Sciclone NGS Workstation

- Optimized scripts written for the SeqCap EZ workflow
- Capability to load tips as needed based on number of samples
- 24 deck locations for more walk-away time
- Small size to fit on bench top
- Fully enclosed system for contamination protection
- Process 288 samples per week