The Agilent Technologies
SureSelect Platform for Target Enrichment

SureSelect XT

Focus Your Sequencing on DNA that Matters

Sheila G Purim, PhD
Agenda

1. SureSelect Technology
2. Whole Exome Sequencing
3. Custom and Catalog Panels
4. Targeted RNA Sequencing
5. Automation
6. Data Analysis
Targeted Resequencing vs. Whole Genome

Required Throughput = Genome Size x Average Coverage

Human Genome
3Gb x 30 = 90Gb
Full Sequencing Run
1 Sample/Run

1 Sample/Run $$$$$$
Targeted Resequencing vs. Whole Genome

Required Throughput = Genome Size x Average Coverage

Target = 50Mb x 100 = 5Gb
Target = 5Mb x 100 = 500Mb
Target = 0.5Mb x 100 = 50Mb
Target = 50Kb x 100 = 5Mb
## Sequencing Needs

### Do you need to sequence the Whole Genome?

<table>
<thead>
<tr>
<th>Research</th>
<th>Sequencing Need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendelian Diseases</td>
<td>Exome Sequences</td>
</tr>
<tr>
<td>Follow-up GWAS</td>
<td>Selected Genomic Regions</td>
</tr>
<tr>
<td>Disease Research</td>
<td>Candidate Genes/Exome Sequences</td>
</tr>
<tr>
<td>Pharma Research</td>
<td>Genes from Biochemical Pathways or with related function</td>
</tr>
<tr>
<td>Clinical Research</td>
<td>Selected Genes</td>
</tr>
</tbody>
</table>
Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing

Andreas Gnrke, Alexandre Melnikov, Jared Maguire, Peter Rogov, Emily M LeProust, William Brockman, Timothy Fennell, Georgia Giannoukos, Sheila Fisher, Carsten Russ, Stacey Gabriel, David B Jaffe, Eric S Lander, & Chad Nusbaum

Targeting genomic loci by massively parallel sequencing requires new methods to enrich templates to be sequenced. We developed a capture method that uses biotinylated RNA ‘bait’ to fish targets out of a ‘pond’ of DNA fragments. The RNA is transcribed from PCR-amplified oligodeoxynucleotides originally synthesized on a microarray, generating sufficient bait for multiple captures at concentrations high enough to drive the hybridization. We tested this method with 170-mer baits that target >15,000 coding exons (2.5 Mb) and four regions (1.7 Mb total) using Illumina sequencing as read-out. About 90% of uniquely aligning bases fell on or near bait sequence, up to 50% lay on exons proper. The uniformity was such that ~60% of target bases in the exonic ‘catch’, and ~80% in the regional catch, had at least half the mean coverage. One lane of Illumina sequence was sufficient to call high-confidence genotypes for 89% of the targeted exon space.

The development and commercialization of a new generation of increasingly powerful sequencing methodologies and instruments have lowered the cost per nucleotide of sequencing data by several orders of magnitude. Within a short time, several individual human genomes have been sequenced on target sets complex enough to match the scale of current next-generation sequencing instruments. The first method, microarray capture, uses hybridization to arrays containing synthetic oligonucleotides that match the target sequence.
<3μg

NEW!
Library Preparation
• SureSelect XT for Illumina
• SureSelect XT for SOLiD

Hybridization / Capture

Bead Separation

Wash / Elution / Amp

SureSelect Target Enrichment System: Workflow

Baits:
- cRNA probes
- Long (120bp)
- Biotin labeled
- User-defined (eArray)
- SurePrint synthesis

Hybridization

24 hours

Wash Beads and Digest RNA

Sequencing
SureSelect Target Enrichment System

<3µg

Genomic Sample (Set of chromosomes)

NGS Kit

0.5µg

Genomic Sample (Prepped)

SureSelect KPB Buffer

SureSelect Biotinylated RNA Library “Baits”

Hybridization

24 hours

Streptavidin Coated Magnetic Beads

Unbound Fraction Discarded

Wash Beads and Digest RNA

Bead Capture

Amplify

Sequencing

Illumina GAIIx

HiSeq 2000

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
SureSelect Target Enrichment System

<3μg

0.5μg

SureSelect Capture Process

Hybridization

24 hours

STREPTAVIDIN COATED MAGNETIC BEADS

Wash Beads and Digest RNA

Bead capture

Sequencing

SOLiD 4

5500
SureSelect Target Enrichment System

<3μg

0.5μg

GENOMIC SAMPLE (Set of chromosomes)

SureSelect™ Target Enrichment System Capture Process

Hybridization

24 hours

STREPTAVIDIN-MADE MAGNETIC BEADS

UNBOUND FRACTION DISCARDED

Wash Beads and Digest RNA

Amplify

Sequencing

454 FLX

GS Junior

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
SureSelect Target Enrichment System

- 3 μg
- 0.5 μg

GENOMIC SAMPLE (Set of chromosomes)

SureSelect™ Target Enrichment System

Capture Process

24 hours

Hybridization

STREPTAVIDIN COATED MAGNETIC BEADS

UNBOUND FRACTION DISCARDED

Wash Beads and Digest RNA

Bead capture

Amplify

Sequencing

Pacific Biosciences

Ion Torrent - PGM

MiSeq

Coming Soon!
Best in Class just got better……

SureSelect<sup>XT</sup> – Complete Workflow Solution
SureSelect Baits
Better sensitivity

<table>
<thead>
<tr>
<th>Feature</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ultra-Long RNA Baits</strong></td>
<td>Only 24hr hybridization compared to 72hrs for competitor</td>
</tr>
<tr>
<td><strong>Binding strength</strong></td>
<td>Capture small and large genetic variations: SNP, InDels, CNV, fusions</td>
</tr>
<tr>
<td>RNA:DNA &gt; DNA:DNA</td>
<td>Low # PCR cycles, less artifacts</td>
</tr>
<tr>
<td></td>
<td>Better Allelic Balance - Better Uniformity</td>
</tr>
<tr>
<td></td>
<td>Low input DNA needed</td>
</tr>
</tbody>
</table>
Longer baits = Better Sensitivity

Really Long RNA Baits Tolerate Mismatches (even long ones)

Allele 1

Allele 2 - SNP

120-mer bait

Allelic Balance post-capture

0.5

0.5
Longer baits = Better Sensitivity

Really Long RNA Baits Tolerate Mismatches (even long ones)

Allele 1

Allele 2 – 25bp deletion

Allelic Balance post-capture

0.5

Smaller Probes will poorly hybridize!

0.5
SureSelect – Proven high sensitivity
Capture not only SNPs and Indels, but also CNVs!

Frequent Mutations of Chromatin Remodeling Gene ARID1A in Ovarian Clear Cell Carcinoma

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gene</th>
<th>Transcript accession</th>
<th>Nucleotide (genomic‡)</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCCO1PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO2PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO3PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO7PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO3PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO7PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO3PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO7PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO3PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO7PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO3PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
<tr>
<td>OCCO7PT</td>
<td>ARID1A</td>
<td>g.chr1:26972561_26972562insA</td>
<td>c.3079C&gt;T</td>
<td>11bp deletion</td>
</tr>
</tbody>
</table>

Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing

Table 3. Genomic deletions and duplication identified by the assay

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
<th>Start*</th>
<th>End*</th>
<th>Size (bp)</th>
<th>Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>15</td>
<td>41,226,145</td>
<td>41,327,157</td>
<td>101,013</td>
<td>0.509</td>
</tr>
<tr>
<td>BRCA1</td>
<td>14</td>
<td>41,230,562</td>
<td>41,325,836</td>
<td>95,275</td>
<td>1.578</td>
</tr>
<tr>
<td>BRCA1</td>
<td>13</td>
<td>41,203,975</td>
<td>41,229,279</td>
<td>15,323</td>
<td>0.519</td>
</tr>
<tr>
<td>BRCA1</td>
<td>12</td>
<td>41,219,394</td>
<td>41,219,795</td>
<td>160</td>
<td>0.489</td>
</tr>
<tr>
<td>BRCA2</td>
<td>17</td>
<td>32,889,026</td>
<td>32,890,990</td>
<td>1,881</td>
<td>0.489</td>
</tr>
<tr>
<td>BRCA2</td>
<td>12</td>
<td>32,950,734</td>
<td>32,952,070</td>
<td>1,337</td>
<td>0.544</td>
</tr>
</tbody>
</table>

Note: Results are ranked by Allele and ORDER pairs, which are not captured order base pairs for wildtype genotype.

25bp deletion
7bp deletion
10bp deletion
11bp deletion
Large deletions and duplications
Operational Excellence

**Automation:**
- Agilent’s Complete automation solution, for the whole workflow
- Validated protocols for SureSelect, customizable
- Agilent service and support

**Sample Multiplexing:**
- 12 Indexes or 16 Barcodes (Illumina/SOLiD)
- Coming soon: 96 indexes/barcodes!
- Validated and optimized protocols

**Compatibility:**
- The only complete agnostic technology on the market
- Compatible with all NGS platforms available

**Protocol:**
- SureSelectXT – complete workflow solution
- Fastest hybridization time on the market
- Validated protocols, continuously being improved
Exon capture is powerful to study diseases

- Most diseases are caused by coding mutations (with some exceptions)
- Exons are only ~1-1.4% of human genome (30-50 Mb)

**Benefits:**
- Much less sequencing: ~5% of WGS, so up to 20x more samples
- More interpretable
- Easier to follow up
SureSelect for Exome Sequencing

- **Human All Exon v1**
  - Chromosome X
  - 7674 Exons
  - 3Mb Capture

- **Human All Exon v2**
  - Chromosome X
  - 50Mb
  - 44MB
  - CCDS+RefSeq
  - Broad

- **50Mb**
  - CCDS+RefSeq+
  - GENCODE
  - Sanger
Expanding Exome Design to Animal Models

NEW!

Mouse All Exon

Exon definition derived from Ensembl + RefSeq
Designed against mm9 reference from UCSC
Number of genes covered: 24,306
Number of exons: 221,784
Total size of design: 49.6 MB

Performance Mouse All Exon Design with 5Gb sequencing

- % Reads On-Target +/- 200bp
- Uniformity (3/4 mean with upper tail)
- % Bases 1X Coverage
- % Bases 10X Coverage
- % Bases 20X Coverage

Promega  C3H  DBA  PWK  15NIH36a  15NIH49a

Agilent Technologies

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
### SureSelect Human All Exon Kits

#### Comprehensive coverage

<table>
<thead>
<tr>
<th></th>
<th>All Exon v2</th>
<th>All Exon 50Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCDS Sept. 2008 + additional RefSeq content including CCDS Sept. 2009 exons</td>
<td>GENCODE and Sanger (includes CCDS and Broad defined v2 content as well)</td>
</tr>
<tr>
<td><strong>Target Size</strong></td>
<td>44Mb</td>
<td>50Mb</td>
</tr>
<tr>
<td><strong>CCDS (Nov. 2010)</strong></td>
<td>98.2%</td>
<td>99.5%</td>
</tr>
<tr>
<td><strong>RefSeq Genes (Nov. 2010)</strong></td>
<td>96.9%</td>
<td>99.0%</td>
</tr>
<tr>
<td><strong>miRNA (miRBase v14)</strong></td>
<td>90.0%</td>
<td>92.8%</td>
</tr>
<tr>
<td><strong>Ensembl (Aug. 2010)</strong></td>
<td>90.9%</td>
<td>96.2%</td>
</tr>
<tr>
<td><strong>GENCODE v4</strong></td>
<td>92.9%</td>
<td>97.0</td>
</tr>
<tr>
<td><strong>Addition of Custom Exome</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Developed with</strong></td>
<td>Broad</td>
<td>Sanger</td>
</tr>
</tbody>
</table>

- Human All Exon kits can be **customized** (PLUS) with up to 6.8Mb additional custom content
- Human All Exon kits can be **multiplexed** on SOLiD4 and HiSeq2000
Human All Exon 50Mb – 5Gb coverage

Most comprehensive Human All Exon content available

Sequencing capacity:
- 0.5-1 sample / lane GAIIx
- 1-3 samples / lane HiSeq
- 5-10 samples / flowcell SOLiD4

Chemistry recommended:
- PE 2x76bp Illumina
- PE 50+25 SOLiD

Multiplexing:
- Illumina
- SOLiD
Comparison of SNP calls with HapMap

**Genotype Sensitivity vs. HapMap**

<table>
<thead>
<tr>
<th></th>
<th>Human All Exon v2</th>
<th>Human All Exon 50Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT is REF</td>
<td>99.13%</td>
<td>99.21%</td>
</tr>
<tr>
<td>GT is variant HOM</td>
<td>98.38%</td>
<td>98.02%</td>
</tr>
<tr>
<td>GT is variant HET</td>
<td>95.68%</td>
<td>94.90%</td>
</tr>
</tbody>
</table>

**Genotype Concordance vs. HapMap**

<table>
<thead>
<tr>
<th></th>
<th>Human All Exon v2</th>
<th>Human All Exon 50Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT is REF</td>
<td>99.76%</td>
<td>99.73%</td>
</tr>
<tr>
<td>GT is variant HOM</td>
<td>98.18%</td>
<td>98.11%</td>
</tr>
<tr>
<td>GT is variant HET</td>
<td>98.55%</td>
<td>98.26%</td>
</tr>
<tr>
<td>OVERALL</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Efficient Capture of 5bp deletion on the X-Chromosome: Menke’s Syndrome

**SureSelect™ Target Enrichment Kit Efficiently Captures 5 bp Mutant Readout on Illumina GA**

hg18_ChrX_77131408_77131467_ + : Wildtype Bait Design

```
CTATTGTTTATCAACCTCATCTTTATCTCAGTAGGGAAATGAAGAACAGATTGAAGCT
CTATTGTTTATCAACCTCATCTTT--------AGTAGAGGAAATGAAAA
ATTTTTATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAG
TTTTATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGGC
GGTTATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAC
TATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGAG
ATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGATTG
ATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGATTG
ATCAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGATTG
CAACCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGATTG
CCTCATCTTT--------AGTAGAGGAAATGAAAAAGCAGATTG
```
Applications to Mendelian disorders

**ARTICLES**

Exome sequencing disorder

Sarah B Ng

Exome capture and massively parallel sequencing identifies a novel HPSE2 mutation in a Saudi Arabian child with Ochoa (urofacial) syndrome

Wisam Al Badr, Suha Al Bader, Edgar Otto, Friedhelm Hildebrandt, Todd Ackley, Weiping Peng, Jishu Xu, Jun Li, Kailey M. Owens, David Bloom, Jeffrey W. Innes

Dietrich Matern, Soma Das, Darrel Waggoner, Dan L. Nicolae, and Carole Ober

Chromosomal Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; and School of Life Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom

The ITALSUG Consortium, Mauriana Galassi, Sona W. Scholz, J. Paul Taylor, Gabriella Natalega, Adriano Ciliberto, and Bryan J. Traylor

The ITALSUG Consortium, Mauriana Galassi, Sona W. Scholz, J. Paul Taylor, Gabriella Natalega, Adriano Ciliberto, and Bryan J. Traylor

Srl Jacek Majewski, and Loydye A. Jerome-Majewska

Froguel

Agata Fiumara, John M. Opitz, Efrat Levy-Lahad, Rachel E. Klevit, and Mary-Claire King

Pierre Dechelotte, Jacek Majewski, and Nada Jabado

Amal Abu Rayyan, Suheir Loulus, Karen B. Avraham, Mary-Claire King, and Moien Kanaan

Bert B A de Vries and Joris A Veltman

Jacy F. Hansen, Stacie K. Lottus, Karen Chong, James C. Mullikin, and Leslie G. Biesecker
• **The Problem:** Linkage analysis confirmed family inheritance pattern to a 28 Mb region (build 36)
  
  • >200 genes contained in 28 Mb region; daunting task to sequence and validate all the variants.
  
  • **Solution:** Identify a 2\textsuperscript{nd} family that manifests the syndrome, determine linkage and perform targeted sequencing using the X-demo kit on heterozygote females of both families
Findings:

- Family 1 heterozygote: frameshift due to an A insertion in *RBM 10* (RNA binding motif 10)
- Family 2 heterozygote: G to A substitution in *RBM 10* (RNA binding motif 10)
- Both mutations were confirmed by Sanger Sequencing
- Inhibition of the orthologous *Rbm 10* expression in developing mouse embryo using anti-sense probes showed patterns that were consistent with observed human malformations

**Target:** 2.7 Mb; 36 bp SE seq on Illumina GA
Schinzel-Giedion Syndrome:
- Severe mental retardation
- Distinctive facial features
- Multiple congenital malformations

De novo mutations of SETBP1 cause Schinzel-Giedion syndrome

Disease phenotype occurs sporadically, suggesting heterozygous de novo mutations in a single gene

Goal: Identify the gene associated with the disease

How: Exome Sequencing of 4 unrelated individuals
Filtering & prioritization of variants

Dominant disease-causing... so exclude variation present in unaffected controls

• dbSNP > 12,000,000 variants
• In-house database: > 3,500,000 variants
  – individually published genomes
  – >100 exomes performed at dept. of human genetics
  – 1,000 genomes project

Focus on variants most likely to affect protein structure

• Splice-site mutations
• Nonsense mutations
• Frameshift mutations
• Highly conserved missense mutations
...and filter more!

<table>
<thead>
<tr>
<th>Variants</th>
<th>Patient 1</th>
<th>Patient 2</th>
<th>Patient 3</th>
<th>Patient 4</th>
<th>Mean</th>
<th>Genes with variants in all samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total called</td>
<td>22,916</td>
<td>22,602</td>
<td>22,152</td>
<td>19,528</td>
<td>21,800</td>
<td>4,735</td>
</tr>
<tr>
<td>Exonic + SpliceSites(SS)</td>
<td>12,196</td>
<td>12,255</td>
<td>11,796</td>
<td>10,498</td>
<td>11,686</td>
<td>3,331</td>
</tr>
<tr>
<td>Non-synonymous (NS) + SS</td>
<td>5,556</td>
<td>5,618</td>
<td>5,427</td>
<td>4,802</td>
<td>5,351</td>
<td>1,634</td>
</tr>
<tr>
<td>Novel (not in dbSNP130)</td>
<td>405</td>
<td>401</td>
<td>390</td>
<td>387</td>
<td>396</td>
<td>35</td>
</tr>
<tr>
<td>Novel (not in in-house db)</td>
<td>299</td>
<td>289</td>
<td>275</td>
<td>288</td>
<td>288</td>
<td>12</td>
</tr>
<tr>
<td>Novel (not in ~100 exomes)</td>
<td>180</td>
<td>186</td>
<td>154</td>
<td>172</td>
<td>173</td>
<td>1</td>
</tr>
</tbody>
</table>
Unlocking the Cancer Genome

- Breast and Ovarian
- Leukemia
- Renal
- Pancreatic
- Colorectal
- Melanoma
- Hepatocellular
SureSelect Target Enrichment in Cancer Research

**Frequent Mutations of Chromatin Remodeling Gene ARID1A in Ovarian Clear Cell Carcinoma**

Siân Jones,1 Tian-Li Wang,2 Ie-Ming Shih,3 Tsui-Lien Mao,4 Kentaro Nakayama,5 Richard Roden,3 Ruth Glas,6 Dennis Slamon,6 Luis A. Diaz Jr.,3 Bert Vogelstein,3 Kenneth W. Kinzler,7,* Victor E. Velculescu,1,* Nickolas Papadopoulos1*  

Ovarian clear cell carcinoma (OCCC) is an aggressive human cancer that is generally resistant to therapy. To explore the genetic origin of OCCC, we determined the exonic sequences of eight tumors after immunoaffinity purification of cancer cells. Through comparative analyses of normal cells from the same patients, we identified four genes that were mutated in at least two tumors. PIK3CA, which encodes a subunit of phosphatidylinositol-3 kinase, and KRAS, which encodes a well-known oncoprotein, had previously been implicated in OCCC. The other two mutated genes were previously unknown to be involved in OCCC: PPP2R1A encodes a regulatory subunit of serine/threonine phosphatase 2, and ARID1A encodes adenine-thymine (AT)-rich interactive domain-containing protein 1A, which participates in chromatin remodeling. The nature and pattern of the mutations suggest that PPP2R1A functions as an oncogene and ARID1A as a tumor-suppressor gene. In a total of 42 OCCCs, 7% had mutations in PPP2R1A and 57% had mutations in ARID1A. These results suggest that aberrant chromatin remodeling contributes to the pathogenesis of OCCC.

Ovarian cancers are a heterogeneous group of diseases with distinct clinicopathological and molecular features (1). Among them, OCCCs, which account for 10% of epithelial ovarian cancers, is one of the most aggressive types because, unlike the high grade-serous type,

- OCCC — one of the most aggressive types of ovarian cancers;
- SureSelect Human All Exon: 8 tumor x normal pairs
- 4 genes found — 2 known, 2 unknown

*Jones et al; Science 330, 228(2010)*
### SureSelect Target Enrichment in Cancer Research

#### Table 1. Mutations in ARID1A, KRAS, PIK3CA, and PPP2R1A in human ovarian clear cell carcinomas.

<table>
<thead>
<tr>
<th>Sample†</th>
<th>Gene</th>
<th>Transcript accession</th>
<th>Nucleotide (genomic)‡</th>
<th>Nucleotide (cDNA)</th>
<th>Amino acid (protein)</th>
<th>Mutation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCC01PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26972561_26972562insA</td>
<td>c.3854_3855insA</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC02PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26896034C&gt;T</td>
<td>c.553C&gt;T</td>
<td>p.Q185X</td>
<td>Nonsense</td>
</tr>
<tr>
<td>OCC03PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26975790_26975792dupGCA (hom)</td>
<td>c.4001_4002dupGCA (hom)</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC07PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979804_26979805delTG (hom)</td>
<td>c.6828_6829delTG(hom)</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC14PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979286_26979288delTG</td>
<td>c.4012delTT (hom)</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC15PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979394G&gt;A</td>
<td>c.5202T&gt;A</td>
<td>p.W1545X</td>
<td>Nonsense</td>
</tr>
<tr>
<td>OCC16PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26895967_26895973delCGCCGCC (hom)</td>
<td>c.486_492delCGCCGCC (hom)</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC18PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979493delA</td>
<td>c.357delA</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC20PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979493delA</td>
<td>c.357delA</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC22PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26896379_268963790_i</td>
<td>c.9insGTC</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC23PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26979686_26979687insT</td>
<td>c.711insT</td>
<td>fs</td>
<td>Indel</td>
</tr>
<tr>
<td>OCC24PT</td>
<td>ARID1A</td>
<td>CDSS285.1</td>
<td>g.chr:26930542C&gt;T</td>
<td>c.1663C&gt;T</td>
<td>p.Q555X</td>
<td>Nonsense</td>
</tr>
</tbody>
</table>

*Indel: insertion or deletion; fs: frameshift; p.: premature stop codon.*

**Notes:**
- 25bp deletion
- 7bp deletion
- 10bp deletion
- 11bp deletion
Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma

Ignacio Varela¹, Patrick Tarpey¹, Keiran Raine¹, Dachuan Huang², Choon Kiat Ong², Philip Stephens¹, Helen Davies¹, David Jones¹, Meng-Lay Lin¹, Jon Teague¹, Graham Bignell¹, Adam Butler¹, Juok Cho¹, Gillian L. Daigle², Danushka Galappaththige¹, Chris Greenman³, Claire Hardy⁴, Mingming Jia¹, Calli Lattimer¹, King Wai Lau¹, John Marshall¹, Stuart McLaren¹, Andrew Menzies¹, Laura Mudie³, Lucy Stebbings¹, David A. Largaespada³, L. F. A. Wessels⁴, Stephane Richard⁵,⁶, Richard J. Kahnoski⁷, John Anema⁷, David A. Tuveson⁸, Pedro A. Perez-Mancera⁹, Ville Mustonen⁹, Andrej Fischer⁹,¹⁰, David J. Adams¹¹, Alistair Rust¹¹, Waraporn Chan-on², Chutima Subimerb², Karl Dykema¹², Kyle Purge¹², Peter J. Campbell¹, Bin Tean Teh²,¹³,¹⁴, Michael R. Stratton¹,¹⁵ & P. Andrew Futreal¹

- **SureSelect All Exon v1 and 50Mb – 7 tumors x 7 normal adjacent tissue**
- **Illumina GAIIx – Paired-end 2x76bp**
- **Found truncating mutations in 4 cases, in PBRM1 gene (Chromatin remodeling, implicated in replication and transcription)**

Varela at al, *Nature*, 2011
SureSelect Target Enrichment in Cancer Research

- Follow-up with 257 other cases – sequencing PBRM1
- Truncating mutations identified in 88/257

- Additional functional studies performed support the involvement of PBRM1 in tumorigenesis
Agenda

1. SureSelect Technology
2. Whole Exome Sequencing
3. Custom and Catalog Panels
4. Targeted RNA Sequencing
5. Automation
6. Data Analysis
Customize your SureSelect Kit

Create your own design or add extra custom sequence to a catalog design – up to 6.8Mb

https://earray.chem.agilent.com
Customize your SureSelect Kit
Create your own design or add extra custom sequence to a catalog design – up to 6.8Mb
SureSelect Custom Designs

Coming Soon – 96 Indexes!
HiSeq – 200Gb
SOLiD4 – 100Gb

1.5Mb-2.99Mb
500Kb-1.49Mb
200Kb-499Kb
<200Kb

6.8Mb
147-288 Samples/Run
333-672 Samples/Run
667-2328 Samples/Run
1536-256 Samples/Run
1536-256 Samples/Run

Agilent Technologies
Custom Designs
Applications for Ion Torrent and other desktop sequencers

- Custom designs up to 1Mb
- Custom designs up to 10Mb (with 1Gb throughput)
- DNA and RNA enrichment applications
- Clinical research panels
The Power of Smaller Panels

Association of TALS Developmental Disorder with Defect in Minor Splicing Component U4atac snRNA

Carrier Testing by Next-Gen

LETTER

Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies

Carrier Testing by Next-Gen

Carrier Testing by Next-Gen

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
Inherited mutations in the tumor suppressor genes BRCA1, BRCA2, and multiple other genes predispose to high risks of breast and/or ovarian cancer.

**Proof-of-Concept study:** developed an assay to capture [with Agilent’s custom SureSelect], and detect all mutations in 21 genes that predispose to breast of ovarian cancer, including BRCA1 and BRCA2,

- There were zero false-positive calls.
- This approach enables genetic testing and personalized risk assessment for breast and ovarian cancer.

Walsh et al; PNAS (2010) 107:28
Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing

Tom Walsh, Ming K. Lee, Silvia Casadei, Anne M. Thornton, Sunday M. Stray, Christopher Pennil, Alex S. Nord, Jessica B. Mandell, Elizabeth M. Swisher, and Mary-Claire King

*Departments of Medicine and Genome Sciences and Obstetrics and Gynecology, University of Washington, Seattle, WA 98195

Table 2. Point mutations and small insertions and deletions identified by the assay

<table>
<thead>
<tr>
<th>Gene</th>
<th>Nucleotide</th>
<th>Effect</th>
<th>Type</th>
<th>Size (bp)</th>
<th>Chromosome</th>
<th>Start</th>
<th>End</th>
<th>Wild type</th>
<th>Variant</th>
<th>% Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>4510 del3ins2</td>
<td>1465 stop</td>
<td>Deletion-insertion</td>
<td>1</td>
<td>17</td>
<td>41,228,596</td>
<td>41,228,597</td>
<td>525</td>
<td>596</td>
<td>0.53</td>
</tr>
<tr>
<td>BRCA1</td>
<td>5083 del19</td>
<td>1657 stop</td>
<td>Deletion</td>
<td>19</td>
<td>17</td>
<td>41,222,949</td>
<td>41,222,968</td>
<td>700</td>
<td>644</td>
<td>0.48</td>
</tr>
<tr>
<td>BRCA1</td>
<td>5382 insC</td>
<td>1829 stop</td>
<td>Insertion</td>
<td>1</td>
<td>17</td>
<td>41,209,080</td>
<td>41,209,081</td>
<td>606</td>
<td>596</td>
<td>0.50</td>
</tr>
<tr>
<td>BRCA2</td>
<td>999 del5</td>
<td>273 stop</td>
<td>Deletion</td>
<td>5</td>
<td>13</td>
<td>32,905,141</td>
<td>32,905,146</td>
<td>363</td>
<td>229</td>
<td>0.39</td>
</tr>
<tr>
<td>BRCA2</td>
<td>1983 del5</td>
<td>585 stop</td>
<td>Deletion</td>
<td>5</td>
<td>13</td>
<td>32,907,366</td>
<td>32,907,371</td>
<td>304</td>
<td>258</td>
<td>0.46</td>
</tr>
<tr>
<td>BRCA2</td>
<td>6174 delT</td>
<td>2003 stop</td>
<td>Deletion</td>
<td>1</td>
<td>13</td>
<td>32,914,438</td>
<td>32,914,439</td>
<td>565</td>
<td>661</td>
<td>0.54</td>
</tr>
<tr>
<td>BRCA2</td>
<td>9179 C &gt; G</td>
<td>2984 stop</td>
<td>Nonsense</td>
<td>1</td>
<td>13</td>
<td>32,953,650</td>
<td>32,953,650</td>
<td>391</td>
<td>361</td>
<td>0.48</td>
</tr>
<tr>
<td>BRIP1</td>
<td>3401 delC</td>
<td>1149 stop</td>
<td>Deletion</td>
<td>1</td>
<td>17</td>
<td>59,761,006</td>
<td>59,761,007</td>
<td>651</td>
<td>486</td>
<td>0.43</td>
</tr>
<tr>
<td>CDH1</td>
<td>591 G &gt; A</td>
<td>157 stop</td>
<td>Nonsense</td>
<td>1</td>
<td>16</td>
<td>68,842,406</td>
<td>68,842,406</td>
<td>421</td>
<td>359</td>
<td>0.46</td>
</tr>
<tr>
<td>CHEK2</td>
<td>1100 delC</td>
<td>381 stop</td>
<td>Deletion</td>
<td>1</td>
<td>22</td>
<td>29,091,857</td>
<td>29,091,858</td>
<td>3,293</td>
<td>586</td>
<td>0.15</td>
</tr>
<tr>
<td>MLH1</td>
<td>ivs14(-1) G &gt; A</td>
<td>568 stop</td>
<td>Splice</td>
<td>1</td>
<td>3</td>
<td>37,083,758</td>
<td>37,083,758</td>
<td>1,024</td>
<td>683</td>
<td>0.40</td>
</tr>
<tr>
<td>MSH2</td>
<td>1677 T &gt; A</td>
<td>537 stop</td>
<td>Nonsense</td>
<td>1</td>
<td>2</td>
<td>47,693,895</td>
<td>47,693,895</td>
<td>575</td>
<td>552</td>
<td>0.49</td>
</tr>
<tr>
<td>p53</td>
<td>721 G &gt; A</td>
<td>R175H</td>
<td>Missense</td>
<td>1</td>
<td>17</td>
<td>7,578,406</td>
<td>7,578,406</td>
<td>449</td>
<td>306</td>
<td>0.41</td>
</tr>
<tr>
<td>PALB2</td>
<td>509 delGA</td>
<td>183 stop</td>
<td>Deletion</td>
<td>2</td>
<td>16</td>
<td>23,647,357</td>
<td>23,647,359</td>
<td>1,283</td>
<td>1,233</td>
<td>0.49</td>
</tr>
<tr>
<td>STK11</td>
<td>ivs6(-1) G &gt; A</td>
<td>316 stop</td>
<td>Splice</td>
<td>1</td>
<td>19</td>
<td>1,221,947</td>
<td>1,221,947</td>
<td>722</td>
<td>572</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Excellent allelic balance

Deletion up to 19bp
Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing

Tom Walsh\(^a\), Ming K. Lee\(^a\), Silvia Casadei\(^a\), Anne M. Thornton\(^a\), Sunday M. Stray\(^a\), Christopher Pennil\(^b\), Alex S. Nord\(^a\), Jessica B. Mandell\(^a\), Elizabeth M. Swisher\(^b\), and Mary-Claire King\(^a,1\)

\(^a\)Departments of Medicine and Genome Sciences and \(^b\)Obstetrics and Gynecology, University of Washington, Seattle, WA 98195

Table 3. Genomic deletions and duplication identified by the assay

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic event</th>
<th>Chromosome</th>
<th>Start*</th>
<th>End*</th>
<th>Size (bp)</th>
<th>Ratio(^\d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>Deletion excns 1–15</td>
<td>17</td>
<td>41,226,145</td>
<td>41,327,157</td>
<td>101,013</td>
<td>0.509</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Duplication exon 13</td>
<td>17</td>
<td>41,230,562</td>
<td>41,235,836</td>
<td>5,275</td>
<td>1.578</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Deletion excns 14–20</td>
<td>17</td>
<td>41,203,975</td>
<td>41,229,297</td>
<td>25,323</td>
<td>0.519</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Deletion excn 17</td>
<td>17</td>
<td>41,219,596</td>
<td>41,219,755</td>
<td>160</td>
<td>0.495</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Deletion excns 1–2</td>
<td>13</td>
<td>32,889,020</td>
<td>32,890,900</td>
<td>1,881</td>
<td>0.489</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Deletion excn 21</td>
<td>13</td>
<td>32,950,734</td>
<td>32,952,070</td>
<td>1,337</td>
<td>0.544</td>
</tr>
</tbody>
</table>

*Breakpoints are flanked by Alu and other repeats, which are not captured.
†Reads per base pair for deletion or duplication/reads per base pair for wild-type genotype.
Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing

Tom Walsh, Ming K. Lee, Silvia Casadei, Anne M. Thornton, Sunday M. Stray, Christopher Pennil, Alex S. Nord, Jessica B. Mandell, Elizabeth M. Swisher, and Mary-Claire King

*Departments of Medicine and Genome Sciences and †Obstetrics and Gynecology, University of Washington, Seattle, WA 98195

- CNVs can be measured by comparing read depth to a reference
- Ability to measure both deletions and duplications
- Variants range from 100s bp – +100kb
More than 1139 genes have been described as associated with Mendelian Disorders

Until now, carrier test has been recommended to only 5 of them (in the US)

**Goal:** develop a preconception genetic test for 448 severe recessive childhood disorders

**2 Methods tested:** SureSelect in-solution enrichment x microdroplet PCR

**Criteria for selection of the best method:** analytical validity, concordance, high-throughput (scalability), cost-effectiveness

Bell et al (2011); ScienceTranslMed, 3:65
Mendelian Diseases Diagnosis and Carrier Testing

Performance Data: SureSelect

Agilent SureSelect Custom Panel (1.9Mb), Illumina GAIIx 2.6Gb sequencing: single-end 50-mer reads. (7717 target segments)

- 99% of bases with >0% coverage;
- 96% of bases with >20x coverage

Trimodal distribution for 92,106 SNP calls

- Accuracy – 98.8%
- Sensitivity – 95.6%
- Specificity – 99.99%
- PPV – Positive Predictive Value - 99.96%
- NPV – Negative Predictive Value - 98.5%
Mendelian Diseases Diagnosis and Carrier Testing

Ability to detect large deletions by local reduction in normalized aligned reads.

Detection of a gross deletion that is a cause of Duchenne muscular dystrophy (OMIM #310200, DMD exons 51 to 55 del; >146Kb deletion!)
Biomarker discovery: disease state

- Mutations linked to various disease states, including cancer
- Cell signaling pathways

SureSelect Human Kinome Kit
SureSelect Human Kinome Kit

612 Genes - 3.2Mb Capture

- 518 putative kinases
- 12 PI3K domain-containing genes
- 13 diglyceride kinases
- 6 PI3K regulatory components
- 9 inositol polyphosphate Kinases
- 9 PIP4/PIP5 Kinases
- 28 genes frequently mutated in human cancer
- 19 genes known to be mutated in breast cancer


Original content defined by Prof. René Bernards – NKI)
SureSelect Human Kinome DNA & RNA Kits

See the “omics” picture!
Same 612 Genes – DNA & RNA


Original content defined by Prof. René Bernards – NKI

Slide courtesy of Prof Rene Bernards
## Agenda

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SureSelect Technology</td>
</tr>
<tr>
<td>2</td>
<td>Whole Exome Sequencing</td>
</tr>
<tr>
<td>3</td>
<td>Custom and Catalog Panels</td>
</tr>
<tr>
<td>4</td>
<td>Targeted RNA Sequencing</td>
</tr>
<tr>
<td>5</td>
<td>Automation</td>
</tr>
<tr>
<td>6</td>
<td>Data Analysis</td>
</tr>
</tbody>
</table>

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
RNA Target Enrichment

Enrich your sample for a subset of transcripts of interest to...

- Get deeper coverage per transcript
- Discover splice junctions, fusions, SNPs, allelic expression
- Quantify gene expression with sensitivity similar to qPCR
- Have more cost-effective and faster experiments
- Study whole biochemical pathways or gene families in one experiment
- Easier data analysis

For Research Use Only. Not for use in diagnostic procedures.
RNA Target Enrichment in Cancer

Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts

Joshua Z Levin*, Michael F Berger†, Xian Adiconis*, Peter Rogov*, Alexandre Melnikov*, Timothy Fennell†, Chad Nusbaum*, Levi A Garraway†§ and Andreas Gnirke*

*Genome Biology 2009, 10:R115

- **RNA Target Enrichment of 467 Cancer Genes** (~all Tyrosine Kinases + Genes from Cancer Gene Census);
  - Overall >800 target transcripts

- **Sequencing Libraries**: K-562 chronic myeloid leukemia (CML) cell line cDNA libraries: with and without target enrichment
RNA Target Enrichment in Cancer

(a) Mean sequence coverage by region

Expression level of target transcripts

For Research Use Only: Not for use in diagnostic procedures.
## Comparison

<table>
<thead>
<tr>
<th>Feature</th>
<th>Without RNA Target Enrichment</th>
<th>With RNA Target Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Coverage</td>
<td>14.4x</td>
<td>606x</td>
</tr>
<tr>
<td>Fold-Enrichment</td>
<td>N/A</td>
<td>42x</td>
</tr>
<tr>
<td>Regions with 20x cov.</td>
<td>13%</td>
<td>63%</td>
</tr>
<tr>
<td>Common SNPs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel Variants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splice-Junctions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternatively spliced-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene Fusions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>5' Gene</th>
<th>5' Chr.</th>
<th>3' Gene</th>
<th>3' Chr.</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCR</td>
<td>22</td>
<td>ABLI</td>
<td>9</td>
<td>13</td>
<td>874</td>
</tr>
<tr>
<td>NUP214</td>
<td>9</td>
<td>XKR3</td>
<td>22</td>
<td>9</td>
<td>152</td>
</tr>
<tr>
<td>SNHG3-RCCI</td>
<td>1</td>
<td>PICALM</td>
<td>11</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>PRIMI</td>
<td>12</td>
<td>NACA</td>
<td>12</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>NCKIPSD</td>
<td>3</td>
<td>CELSR3</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SLC29A1</td>
<td>6</td>
<td>HSP90ABI</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

For Research Use Only: Not for use in diagnostic procedures.
SureSelect RNA Target Enrichment
First RNA Capture on the market!

SureSelect Custom RNA Capture

- Designs up to 6.8Mb

SureSelect Catalog Kits

- SureSelect Human Kinome RNA (612 Genes - 3.2Mb Capture)
SureSelect RNA Target Enrichment Protocol

Start with <1ug of total RNA or mRNA*

Similar process to DNA Target enrichment, except working with a cDNA NGS library

Complete protocol takes ~ 3-4 days

Protocols available for Illumina and SOLiD, for individual or multiplexed samples

*see protocol for details and recommendations
Performance of SureSelect RNA Target Enrichment

Gene Expression Levels

High correlation to unenriched RNA library = enrichment process doesn’t create any bias – reliable gene expression results!
Performance of SureSelect RNA Target Enrichment

- Target Enriched transcripts
- Non-Target Unenriched transcripts
Performance of SureSelect RNA Target Enrichment

Gene Expression Levels

High Reproducible process = reliable gene expression results

$R^2=0.998$
Performance of SureSelect RNA Target Enrichment

>90% on-target reads, for different sizes of custom designs
Performance of SureSelect RNA Target Enrichment

An enriched library clearly provides more reads/transcript and better sequence coverage
Performance of SureSelect RNA Target Enrichment

High coverage for splice junctions
SureSelect Human Kinome DNA & RNA Kits

See the “omics” picture!
Same 612 Genes – DNA & RNA

Human Kinome - DNA

Human Kinome - RNA


Original content defined by Prof. René Bernards – NKI

Slide courtesy of Prof Rene Bernards
Real world example in from transcriptome sequencing of the kinome in esophageal cancer samples: KRAS Splice Variants

Detect splice variants!
Detect fusion transcripts!
Agenda

1. SureSelect Technology
2. Whole Exome Sequencing
3. Custom and Catalog Panels
4. Targeted RNA Sequencing
5. Automation
6. Data Analysis
Automation of Agilent SureSelect\textsuperscript{XT} for Next Generation Sequencing

**Instruments Highlighted by NGS Workstation:**
- **G5409A Bravo Automated Liquid Handler**
  - On Deck Gripper Upgrade (Option 102)
  - 96LT Disposable Pipette Head (Option 178)
  - Peltier Heating/Cooling (Option 161)
  - Orbital Shaking Station (Option 159)
- **G5400A BenchCel Microplate Handling Workstation**
  - Standard Four Rack Configuration (Option 052)
  - Front Loading Racks for BenchCel (Option 066)
- **G5472A Labware MiniHub**
- **Desktop PC with VWorks Software**
  - Validated Protocols for running Agilent SureSelect workflows
- **Standalone PlateLoc**
Automation of Next-Gen Sequencing Using SureSelect

SureSelect XT Library Prep
- Shear Genomic DNA
- Repair Ends
- 3'-dA Addition
- Adapter Ligation
- PCR Enrichment
- Prepped Library

SureSelect Target Enrichment
- Automated Purification
- Automated Oligo Capture Library
- Library Hybridization
- Automated Bead Capture
- PCR
- Automated Purification
- QA

Illumina Sequencer
On and Near Target Performance Consistent between Well Position with Kinome Captures

Percentage of Reads On and Near Target

- Percentage reads in regions +/- 100bp:
- Percentage reads in regions +/- 200bp:
- Percentage reads in regions +/- 300bp:

Well Position on 96-well Plate
## Comparison of Timing and Throughput of Manual vs. Bravo SureSelect Sample Processing

<table>
<thead>
<tr>
<th></th>
<th>MANUAL METHOD</th>
<th>AUTOMATED METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hands-On Time</td>
<td>Maximum Number of Samples Processed/Day&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Library Prep</td>
<td>375</td>
<td>20</td>
</tr>
<tr>
<td>Pre-Capture PCR</td>
<td>90</td>
<td>20</td>
</tr>
<tr>
<td>Hybridization Preparation</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>Capture</td>
<td>210</td>
<td>20</td>
</tr>
<tr>
<td>Post-Capture PCR</td>
<td>90</td>
<td>20</td>
</tr>
<tr>
<td>Total Minutes</td>
<td>825</td>
<td>20 - 40 per week</td>
</tr>
<tr>
<td>Total Hours</td>
<td>13.75</td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>Assumes individual tube preps with an experienced user staggering sample processing

<sup>2</sup>Library prep master mixes done once

Up to 10x throughput per week vs. manual processing!
- Reduction of PCR cycles required – from 14 to 6-8 cycles
- 1200 samples processed/week, in batches of 96
Agenda

1. SureSelect Technology
2. Whole Exome Sequencing
3. Custom and Catalog Panels
4. Targeted RNA Sequencing
5. Automation
6. Data Analysis

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
Avadis NGS - Downstream Analysis of NGS Data

Primary Analysis

Control Software

Data File (Reads + Quality)

Secondary Analysis

Data File (Reads + Quality)

ELAND/BIOS COPE/BWA...

Reads aligned to genome

Tertiary Analysis

Reads aligned to genome

Avadis NGS

Chip-Seq

DNA-Seq

RNA-Seq

Agilent SureSelect™ Platform
Enabling Products for the Next-Generation Sequencing Workflow
DNA Targeted Resequecing with SureSelect

Coverage & Reads BEFORE Filtering

Coverage & Reads AFTER Filtering

Reads Not Falling on Targeted Region Can Be Removed To Focus Analysis

Targeted Regions (SureSelect Baits)
SNPs & InDel Detection and Annotation

The consequence column contains the effects of variants. For SNPs, the dbSNP Annotation column is provided. Indels are also included. Predicted Amino Acid Change for Non-Synonymous Coding SNPs is indicated.
Acknowledegements

- Collaborators:
  - Broad Institute
    - Chad Nussbaum et. al
    - Stacey Gabriel et. al
    - Sheila Fisher et. al
  - Sanger Institute
    - Daniel Turner et. al.
  - NKI
    - Rene Bernards
    - Ian Majewski
  - All our early access collaborators (over 20 institutions worldwide)
Thank You!