Target Enrichment Strategies for Next Generation Sequencing

Anuj Gupta, PhD
Agilent Technologies, New Delhi

Genotypic Conference, Sept 2014
FIGURE 19.1  Timeline of introduction of DNA sequencing technologies and platforms.
Information burst

- Nearly 30,000 human genomes sequenced
- 391 terabases deposited in Genbank
- 112 vertebrate, 455 non-vertebrate eukaryotic, 8760 prokaryotic genomes sequenced till 2013
- 1542 GWAS published
- 62.6 million SNP reported in 2013
- 1.4 million short stretches of InDels
- 14000 large deletions
More to come..

- Cost reduced to $1,000
- Tremendous progress in terms of read length and throughput
- Faster library preparation technologies
- Faster sequencing technologies
What NEXT...??
3 Ways to pursue Next Generation Sequencing

More sequencing to identify Rarest of Rare

Identify interested regions and find variation

Identify Rare.. but phenotypically relevant

Whole Genome Sequencing

Panel Sequencing

Exome Sequencing
Approaches of Next Generation Sequencing

- **Discovery**
  - R of R Variations
  - Variation anywhere
  - Any gene
  - Research
  - Whole Genome
  - GWAS

- **Follow-up**
  - Frequency of Variations
  - Variation in all proteins coding genes
  - Research & Application
  - Whole Exome
  - Exome+UTRs etc

- **Application**
  - Variations within
  - Variations in selected genes
  - Focused application
  - Pathway panels
  - Disease panels
## Sequencing Needs

### Why do we need Targeted sequencing?

<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>
World wide trend of Sequencing

Whole genome sequencing
NCBI publication No. 190
From 2011 - Current

Targeted Resequencing
NCBI publication No. 490

Why Targeted Sequencing Publication Increased in NCBI?

Exome capture enables more samples per run

Whole Genome
6 samples per run
90 GB per sample

Exome
150 samples per run
4 GB per sample

Exome sequencing requires ~23x less sequencing—allowing 25x greater sample throughput and faster time to discovery.
Target Enrichment: Focus on what Matters

Sequencing of small regions, few genes, a chromosome, a pathway etc..

Key Goals of Target Enrichment

• Focus on a subset of the genome / transcriptome

• Achieve greater depth in targeted regions
  – Discover more mutations, SNPs and Indels

• Lower sequencing costs and obtain faster time to result
Targeted sequencing approaches

PCR Approach

Hybridization Approach

Circularized Amplicons

Transposon Approach
Targeted sequencing approaches

What to see before choosing TE

- Sensitivity & Specificity
- Coverage
- Uniformity
- Reproducibility
- Cost
- Ease of use
- Amount of DNA required
- Multiplexing capability
- Platform compatibility
- Designing
- Content / Database covered for Designing
- TOT turn around time – hybridization
- Variant reliability
TARGET ENRICHMENT

PCR based approaches
PCR-based Target Enrichment

Normal PCR sequencing

5' - CTGACTATTAACTGAA-3'
3' - ACTT-5'

DNA (template strand)
Primer

5' - CTGACTATTAACTGAA-3'
ddGACTGATAATTGACTT

3' - ACTT-5'
12 new DNA fragments

Termination Method for Sequencing DNA

separated by electrophoresis + Fluorescence detection

DNA polymerase

dATP ddATP
dCTP ddCTP
dTTP ddTTP
dGTP ddGTP

Agilent Technologies
PCR-based Target Enrichment

Target Enrichment by Multiplex PCR

Use of multiple primer sets to amplify multiple targets on a genomic template

Designed to include adapter sequences

Ideal for preparing amplicon libraries
Target Enrichment by microDroplet PCR

Aqueous droplets possessing volumes from nanoliters to femtoliters are used as individual biochemical reactors.

Picoliter volume PCR reactions

Each droplet represents one PCR reaction.

10 million discrete reactions per hour.
Target Enrichment by Microfluidics

Multiple samples and Primer sets are processed in microfluidics chip

Picoliter volume PCR reactions

Combination of samples and amplicons

>450 amplicons can be processed from 40-50 samples
PCR-based Target Enrichment

ADVANTAGES

• **Fast workflow**
  Single day from DNA sample to result

• **Low cost**
  Due to use of primers, simple PCR reagents are required

• **Low DNA input**
  PCR amplifies.. so few ng DNA is sufficient

• **Well established**
  Everybody know PCR process.. so simple and established
LIMITATIONS

• **Primer Cross-reactivity**
  Primers usually cross-react to produce non-specific products

• **Limited in target regions**
  Limited by amplicon length and number of primers that could be accommodated, e.g. 5-6 K

• **Long optimization**
  Specially for new designs

• **Dropouts**
  Variation at primer binding sites could lead to amplicon dropout

• **Artifacts**
  Variation is true or artifict..?
Artifacts and Dropouts in Multiplex PCR

If an unknown mutation appears in a primer site, it causes a complete dropout in the target region. If a variant occurs, it is hard to know if it is a real mutations and not a PCR artifact.
TARGET ENRICHMENT

Hybridization based approaches
Hybridization based Target Enrichment

Array capture technology

Probes that are fixed to a chip are hybridized to fragmented genomic DNA, immobilizing complementary target sequences.

Targeted DNA sequences can be eluted and used as a sample for library preparation.
Hybridization based Target Enrichment

Array capture technology

**Good coverage**

**Low cost**

Need higher amount of starting DNA material - 10 ug

Less easy workflow

Less targets to be captured
Hybridization based Target Enrichment

Solution capture technology

Complimentary probes are hybridized to fragmented genomic DNA in solution.

Probes are biotinylated.

Targeted DNA sequences are captured by magnetic beads coated with streptavidin. Beads are separated with the help of magnet.

Targets are eluted and used as a sample for library preparation.
Hybridization based Target Enrichment

Solution capture technology

Allele 1

120-mer bait

Allele 2 - SNP

Allelic Balance post-capture

0.5

0.5

Agilent Technologies

Agilent SureSelect Platform
For Research Use Only
Hybridization based Target Enrichment

Solution capture technology

Allele 1

0.5

Allelic Balance post-capture

Allele 2 – 25bp deletion

0.5
Hybridization based Target Enrichment

Advantages

• Good recovery of sequences targeted by the probe set
• Probes being mobile (solution) the probability of probe-target hybridization is high
• Better representation of large variations due to hybridization probes
• No dropouts and artifacts
• Less PCR duplicates
• Easy to design and tile
• Portable to automated workstations and is therefore easily scalable
• DNA baits / RNA baits

Limitations

Not as fast as PCR based methods
Require little more DNA – now possible with low input
Costlier than PCR based methods
TARGET ENRICHMENT

Combination of Hybridization and PCR
(Haloplex technology)
Combination of Hybridization and PCR for Target Enrichment

Capture by Circularization

1. Digest and denature sample DNA
   - Target Region

2. Hybridize probe library to DNA targets
   - Biotin
   - Sequencing primer motif
   - Index
   - Bridge or emulsion
   - PCR primer
   - Target & complementary probe sequence

3. Purify and ligate targets
   - Probe/Fragment hybrids are retrieved

4. Amplify enriched fragments by PCR
   - PCR primers

   - Only circular DNA targets are amplified.

200ng input sample is fragmented
- 16 RE
- 8 double digests

Research Use Only. Not for use in diagnostic procedures.
Hybrid/PCR

With HaloPlex each target base is covered by multiple amplicons (different start and stop sites)!

If a variant occurs, it can be checked by multiple amplicons; with HaloPlex if a variant occurs it is hard to know if it is a real mutation and not a PCR artifact.

PCR

If an unknown mutation appears in a restriction site, it may affect one or two fragments, but all others will be present.

If an unknown mutation appears in a primer site, it causes a complete dropout in the target region.
Advantages

- Amplicon tiling improves coverage by design
- Amplicon redundancy reduces risk of losing completeness if a probe fails; protects against primer site mutations
- Library preparation free
- No cross reactivity

Limitation

- Depends upon restriction enzymes cut sites
- Fixed target capture sites
TARGET ENRICHMENT

Capture by Transposon based technology
Transposon based Target Enrichment

Capture by transposases

A 19-bp region borders the native transposon

Transposase enzyme binds to this region and covalently attaches it to target DNA

A single stranded tag is added to provide a PCR primer binding site

Enrichment and library preparation occurs simultaneously
Transposon based Target Enrichment

Capture by transposases

**Advantages**

- Very fast process – total time 3-4 hours only including 1 hr hybridization
- Less input DNA is required
- Not limited to human
- Works for Targeted Sequencing as well as WGS

**Limitation**

- Very sensitive to DNA quantification
- Little costlier than PCR technology
What we do with Targeted Sequencing

1. Exome Sequencing
2. Panel Sequencing
3. Custom sequencing
What we do with Targeted Sequencing

1. Exome Sequencing
2. Panel Sequencing
3. Custom sequencing
Exome

Exons are 1-2 % of whole genome

There are 180,000 exons, which constitute about 1% of the human genome, or approximately 37 million base pairs

Harbors most of the highly penetrant phenotypically / clinically relevant variations

It is believed that the exome contains about 85% of heritable disease-causing mutations.
LOOKING FOR RELEVANT

Good choice for those who are looking for rare mutations, especially as a complement to studies of common variation like GWAS

RARE VARIANT MAPPING IN COMPLEX DISORDERS

Rare variants in complex diseases have been found in low freq. Large sample size needed, which is not cost-effective for WGS

DISCOVERY OF MENDELIAN DISORDERS

Disease-causing variants of large effect have been found to lie within exomes.

CLINICAL DIAGNOSTICS

Diagnostics, Therapeutic management, Risk assessment
EXOME SEQUENCING

**Low Cost**

Overall cost is low due to less target size

**Rapid Method**

Exome sequencing workflows enable cancer researchers to identify coding variants in tumor samples up to 70% faster than WGS

**High throughput**

Allows sequencing of at least 20 times as many samples compared to whole genome sequencing

**Less sequencing**

Only 10-20% sequencing is required to generate 20 fold higher confidence

**Less analysis**

Less data output, easy analysis
Identification of mutation for Inflammatory bowel disease


The efforts of a group of researchers at the Medical College of Wisconsin to diagnose and treat a young boy with an inexplicable and exceptionally severe case of inflammatory bowel disease.

In what turned out to be the first clinical use of exome sequencing, researchers identified a single point mutation in the X-linked inhibitor of apoptosis (XIAP) gene.

That information suggested a possible therapeutic strategy, umbilical cord blood transplant.

Since then, the college has applied the approach to 25 additional cases, obtaining a “definitive diagnosis” in 27% of them.
Clinical diagnosis of Bartter syndrome

Choi et al 2009 reported successful clinical diagnosis of a suspected Bartter syndrome patient of Turkish origin. Bartter syndrome is a renal salt-wasting disease.

Exome sequencing revealed an unexpected well-conserved recessive mutation in a gene called SLC26A3 which is associated with congenital chloride diarrhea (CLD).

This molecular diagnosis of CLD was confirmed by the referring clinician.

This example provided proof of concept of the use of whole-exome sequencing as a clinical tool in evaluation of patients with undiagnosed genetic illnesses.

This report is regarded as the first application of next generation sequencing technology for molecular diagnosis of a patient.
Type of Exome Capture

**Full Human Exome**
Target all 1,80,000 exons and flanking regions

**Extended Exome (Exome + UTRs)**
Target all exons and other important regions such as UTRs, miRNA, other ncRNA

**Exome + Other custom regions**
Exome capture can be combined with custom targets of interest

**Clinical Exome**
Coverage of Exome with additional deep coverage of disease associated regions
What we do with Targeted Sequencing

1. Exome Sequencing
2. Panel Sequencing
3. Custom sequencing
Panel Sequencing

- Throat Cancer
- Cancer of the Oesophagus
- Lung Cancer
- Stomach Cancer
- Bowel Cancer
- Cancer of the Bladder
- Prostate Cancer
- Testicular Cancer
- Skin Cancer
- Ovarian Cancer

- Throat Cancer
- Cancer of the Oesophagus
- Lung Cancer
- Breast Cancer
- Stomach Cancer
- Bowel Cancer
- Cancer of the Bladder
- Cancer of the Uterus or Cervix
- Skin Cancer

- Wheat field

- Heart diagram
- Fish images

Agilent Technologies
Panel Sequencing

Cancer

Cardiomyopathy

Noonan Syndrome

ICCG (ISCA)

X Chromosome

Connective Tissue Disorder

Arrhythmia
What we do with Targeted Sequencing

1. Exome Sequencing
2. Panel Sequencing
3. Custom sequencing
Custom Sequencing

Creation of a Panel

I don’t like these panels

I have targets and want to design my own panel

CUSTOMIZATION
Custom Sequencing

Creation of a Panel

Any Panel (Catalog or Pre-Designed) + Custom Targets = The Best research panel for your lab!

SureDesign
Custom Sequencing

Creation of a Panel
The Power of Custom Panels

- All 54 known deafness genes
- 21 breast and ovarian cancer genes
- Detects fusions for 90 human tyrosine kinases
- Exons – Chrom
- 437 Genes related to Mendelian Disorders
- 3.19Mb Custom Design

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

Carrier Testing by Next-Generation Sequencing

Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies

Cancerogene

 detecting...
THANK YOU

Happy Sequencing