Unravelling cancer genomes in the NGS era: from drug discovery to fusion transcript detection

Iwanka Kozarewa
Senior Scientific Officer-NGS
Breakthrough Breast Cancer Research Centre
The Institute of Cancer Research
Why do we need to dissect cancer genomes?

We need:

- New therapeutic targets.
- Biomarkers that direct use of therapy.
- Dissect heterogeneity.
- Understand cancer progression.
- Ultimately, be able to provide personalized therapy.
Ultra-high throughput functional screening: pairing RNAi & Solexa sequencing

The purposes of the screens are:

- Identify new therapeutic targets.
- Optimize use of existing drugs and identify combination of drugs.
- Dissect drug resistance.
Ultra-high throughput functional screening: theory (1)
Ultra-high throughput functional screening: theory (2)

pGIPZ constructs encode miRNA precursors of shRNAs targeting number of genes

10,000 constructs pooled together and packaged in 293 cells

Pools of packaged virus produced and used to transduce target cells

Level of infection monitored by GFP detection and enriched by FACSort or puromycin selection

Target cells are exposed to a form of selection (drug resistance, viability in culture etc.) or sorted according to protein expression. Genes involved in determining the phenotype of interest are identified by relative enrichment or depletion of shRNAs that target their mRNA sequence.
Ultra-high throughput functional screening: quantification

- shRNA depletion or enrichment is monitored in the final cell population by PCR amplification of shRNA sequences followed by massively parallel sequencing using the Illumina GAIIx.

- Genomic DNA is amplified using primers complimentary to two of the constant regions found in all shRNA sequences.

- These primers amplify a gene-specific region in each shRNA ("variable stem 2"). They also are extended by the P5 and P7 sequences that allow hybridization to the flowcell and subsequent sequencing.

- In order to reproducibly generate optimal cluster density and, thus, maximize yield, we quantify the half-hairpin libraries using qPCR.
Ultra-high throughput functional screening: validation

- Western blot analysis revealing effective p53 and PARP1 silencing by pGIPZ constructs 48 hours after infection of MCF7 cells.

- GIPZ virus can robustly infect a wide range of cell lines, including difficult to transfect cells.

- Comparison of shRNA frequency indicated that shRNA depletions of >25% can be detected.

- Reproducibility of detection was assessed using replica experiments, indicating a highly robust system ($r^2$ replica experiments = 0.93)
Three different shRNAs targeting geneR scored as conferring resistance to tamoxifen treatment.

Two other members of the same core network, here labelled as geneR’ and geneR”, also scored.

PTEN knockdown known to cause resistance to tamoxifen behaved as expected, confirming our novel targets discovery.

The same hits were obtained in an independent biological replicate of the screen.

Preliminary validation studies employing shifts in GFP biclonal populations and siRNA assays ratified these genes as determinants of tamoxifen response.
Agilent SureSelect Human Exome Kit

- Easily scalable in-solution enrichment.
- No special equipment required.
- <3 μg DNA required.
- 2X76bp runs used.
- Two lanes per sample: ~80X coverage.
- Done for SNP and INDEL discovery.
- For structural rearrangements and CNV identification supplemented with whole genome, low-depth, PCR-free sequencing.

Targeted exome re-sequencing: theory
Targeted exome re-sequencing: validation

BRCA2: c.5946ΔT

BRCA2: Exon 10 T Deletion

IFSTA_{1981} sgksvq...

> IFSTA_{1981} renlsr...
Targeted exome re-sequencing: novel INDEL discovery

EPHB2:
c.2673ΔC

EPHB2:Exon14 C Deletion
PNSLK$_{89}$ amapl... > PNSLK$_{89}$ pwrps...
Indel connected component

Deletion
Insertion
Connectivity
Frameshift
mRNA sample preparation (1)

**Illumina protocol**

- mRNA purification
- Fragmentation for 5 min
- 1st and 2nd strand synthesis
- End repair, A-tailing and adapter ligation

**In-house protocol**

- mRNA purification
- Fragmentation for 2 min
- 1st and 2nd strand synthesis
- End repair, A-tailing and adapter ligation

Adapter amount adjusted according to the cDNA concentration
mRNA sample preparation (2)

**Qiagen column clean up**

**SPRI beads clean up**

**Size selecting material in the 200 25bp range**

**Size selecting of material between 375-450 bp**

**Standard Phusion-based PCR of 16 cycles**

**Herculase-based PCR of 10-16 cycles depending on the starting amount of total RNA**

**Qiagen column clean up**

**SPRI beads clean up with reduced ratio beads: sample**
mRNA sample protocols’ comparison

PE-RNASeq (Illumina protocol)

PE-RNASeq (‘in-house’ protocol)

PE-Exome

Error rates by cycle for file_s_1_0001_rescore.txt

Error rates by cycle for file_s_2_0044_rescore.txt

Error rates by cycle for file Stats/s_1_0001_rescore.txt
Secretory breast carcinoma sample (SEC 419184) was sequenced: 2X54bp run: mappable yield of 4.8 Gb.

9 pairs of reciprocal gene fusions were detected (methods as described in Maher et al., 2009)
RNA sequencing: novel fusion transcript discovery

**ELMO2**: may function in phagocytosis of apoptotic cells and in cell migration.

**RAE1**: homolog of the yeast RAE1 (involved in RNA export).

2 X 54 bp run
3.8 Gb yield

**MCP-7105**

**ELMO2-RAE1 (chr1)**
**SLC2A1**: encodes a major glucose transporter in the mammalian blood-brain barrier. Mutations have been found in a family with paroxysmal exertion-induced dyskinesia.

**FAF1**: binds to FAS antigen and can initiate apoptosis or enhance apoptosis initiated through the FAS antigen.

**RNA sequencing**: novel fusion transcript discovery

2 X 54 bp run
3.8 Gb yield
**Acknowledgements**

<table>
<thead>
<tr>
<th>NGS Core Team</th>
<th>High throughput Functional Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerry Fenwick</td>
<td>Cristina Naceur-Lombardelli</td>
</tr>
<tr>
<td>Konstantinos Mitsopouloos</td>
<td>Ana Mendes-Pereira</td>
</tr>
<tr>
<td>David Sims</td>
<td>Maria Cerone</td>
</tr>
<tr>
<td>Chris Lord</td>
<td>Darren Burgess</td>
</tr>
<tr>
<td>Alan Ashworth</td>
<td>Jessica Taylor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Targeted Exome Re-sequencing</th>
<th>Cancer Informatics</th>
<th>RNA sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Louise Barber</td>
<td>Jarle Hakas</td>
<td>Rachel Natrajan</td>
</tr>
<tr>
<td>Juan Manuel Rosa-Rosa</td>
<td>Marketa Zvelebil</td>
<td>Chris Maher</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jorge Reis-Filho</td>
</tr>
</tbody>
</table>

**Funding**

Institute of Cancer Research  
Breakthrough Breast Cancer Research Centre