Bridging the gap to utilising whole genome data in drug development:

*Leveraging next generation sequencing in somatic cancer genetics*

TAT 2011

Andy Futreal
Cancer Genome Project
Wellcome Trust Sanger Institute

Heteroduplex/PCR resequencing of a handfull of genes

<table>
<thead>
<tr>
<th>BRAF mutations</th>
<th>Cancer cell lines</th>
<th>Primary tumours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Mel.</td>
<td>(2) Colo. ca.</td>
</tr>
<tr>
<td>Nucleotide</td>
<td>Amino acid</td>
<td></td>
</tr>
<tr>
<td>G1388A</td>
<td>G463E</td>
<td>1</td>
</tr>
<tr>
<td>G1388T</td>
<td>G463V</td>
<td>1</td>
</tr>
<tr>
<td>G1394C</td>
<td>G465A</td>
<td>1</td>
</tr>
<tr>
<td>G1394A</td>
<td>G465E</td>
<td>1</td>
</tr>
<tr>
<td>G1394T</td>
<td>G465V</td>
<td>1</td>
</tr>
<tr>
<td>G1403C</td>
<td>G468A</td>
<td>2</td>
</tr>
<tr>
<td>G1403A</td>
<td>G468E</td>
<td>2</td>
</tr>
<tr>
<td>G1753A</td>
<td>E585K</td>
<td></td>
</tr>
<tr>
<td>T1782G</td>
<td>F504L</td>
<td></td>
</tr>
<tr>
<td>G1783C</td>
<td>G595R</td>
<td>1</td>
</tr>
<tr>
<td>C1786G</td>
<td>L596V</td>
<td>1</td>
</tr>
<tr>
<td>T1787G</td>
<td>L596R</td>
<td>1</td>
</tr>
<tr>
<td>T1796A</td>
<td>V599E</td>
<td>19</td>
</tr>
<tr>
<td>TG1796-97AT</td>
<td>V599D</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: BRAF mutations in human cancer

Amino acid residues are grouped in blocks. Three further BRAF coding sequence variants were identified (G2041A R681Q in the HEK1A endometrial cancer cell line. T974C I325T in the ZR-75-30 breast cancer cell line, and C2180T A727V in the H33AJ-1A1 T-ALL cell line). These were not present in 341 control DNAs. Lane numbers (in parentheses) are provided for convenience. Mel., melanoma; Colo. ca., colorectal cancer; Mel. STC, melanoma short-term culture.

* Four out of ten LMP (low malignant potential); 1 out of 25 malignant epithelial.
†Glioma (n = 15), breast cancer (n = 33), prostate cancer (n = 23), HNSCC (head and neck squamous cell carcinoma) (n = 19), lung cancer (n = 14).
What we learned from earlier PCR-based systematic screens

• The majority of somatic mutations identified in large-scale resequencing screens are likely to be passenger events

• There is evidence for multiple infrequently mutated genes under positive selection

• There are likely to be many genes that can contribute to oncogenesis when mutated

• A large number of genes sequenced in a large number of cancers of “same” type will be needed to begin to fully elucidate the complement of cancer genes
PCR exon-resequencing ccRCC study

101 ccRCC cases
(96 clinical samples + 5 matched pair cell lines)

Screen the coding exons of 3,544 genes
(750 Mb total)

Copy number and expression

Follow-up series of 311 clear cell carcinoma clinical samples with matching normals
Results

VHL point mutations in ~60%, VHL/Hypoxia signature in ~85%

Very quiet on SNP6.0, no high level focal amplicons

75/91 (82%) of cases assessed for expression had upregulation of genes associated with cellular hypoxia
Four/five significantly mutated genes in ccRCC are histone methylase/demethylases

UTX – H3K27 demethylase and MLL2/3 complex in H3K4 methylation
12/407 cases mutated (12/12 truncating)

SETD2 – H3K36 methyltransferase
15/407 cases mutated (12/15 truncating)

JARID1C – H3K4 demethylase
14/407 cases mutated (12/14 truncating)

MLL2 – H3K4 methyltransferase
17/407 (including a silent, 6 truncating, 5 missense)

The allelic architecture of driver mutations in clear cell renal cancer
Whole ‘exome’ sequencing

Rearrangement screens

Whole genome shotgun
Solution Hybrid Capture of all coding exons and miRNA genes

Exome is GenCode/ICGC includes 21,416 protein coding genes + 1664 miRNA genes
<table>
<thead>
<tr>
<th>Sample</th>
<th>Sex</th>
<th>Age</th>
<th>Grade</th>
<th>Histology</th>
<th>VHL mutation^</th>
<th>SETD2 mutation</th>
<th>UTX mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD2125a</td>
<td>M</td>
<td>82</td>
<td>4</td>
<td>Clear Cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2126a</td>
<td>F</td>
<td>74</td>
<td>1</td>
<td>Clear Cell</td>
<td>c.236_241delGCAGTC; p.R79_P81&gt;P</td>
<td>c.1801T&gt;A; p.R601*</td>
<td></td>
</tr>
<tr>
<td>PD2127a</td>
<td>F</td>
<td>59</td>
<td>4</td>
<td>Clear Cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2144a</td>
<td>F</td>
<td>63</td>
<td>4</td>
<td>Clear Cell</td>
<td>c.525delC; p.Y175*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2147a</td>
<td>F</td>
<td>50</td>
<td>2</td>
<td>Clear Cell</td>
<td></td>
<td></td>
<td>c.4161_4162delTG; p.Y1387fs*1</td>
</tr>
<tr>
<td>PD3295a</td>
<td>M</td>
<td>62</td>
<td>4</td>
<td>Clear Cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3441a</td>
<td>M</td>
<td>69</td>
<td>1</td>
<td>Clear Cell</td>
<td>c.223_225delATC; p.F76_C77&gt;C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^ VHL mutations in PD2126a and PD3441a were not "re-discovered" in exome sequencing due to poor coverage of the highly GC-rich first exon.
PBRM1 is somatically mutated in 40% (92/227) of ccRCC

ARID1A ** also mutated in ccRCC

Nature, Jan 20, 2011
36/38 PBRM1 mutant ccRCC have hypoxia signature

55/107 cases with a demonstrable* VHL mutation have a PBRM1 mutation

9/9 cases with a SETD2 mutation have mutation in either VHL or PBRM1

6/9 SETD2 mutant cases have a mutation in all three gene

3 tumour suppressor genes unmasked with only 4 hits

3p LOH is most frequent marker in ccRCC

*point mutations only
Breast Cancer Exome Sequencing

- 72 cases
  - 46 ER+ (HER2-)
  - 11 Triple Negative
  - 12 HER2+
    - 1 BRCA1 mut, 2 BRCA2 mut

- Exome is GenCode/ICGC includes 21,416 protein coding genes + 1664 miRNA genes

Solution hybrid capture (Agilent) followed by sequencing on Illumina and variant calling as just described

Follow-up in 300 cases
<table>
<thead>
<tr>
<th>Sample</th>
<th>TOTAL mutations</th>
<th>AKT1</th>
<th>AKT2</th>
<th>CDH1</th>
<th>GATA3</th>
<th>KRAS</th>
<th>MAP2K4</th>
<th>NF1</th>
<th>PIK3CA</th>
<th>PTEN</th>
<th>RB1</th>
<th>SETD2</th>
<th>STK11</th>
<th>SMAD4</th>
<th>TP53</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD3852a</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3854a</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3856a</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3857a</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3858a</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3985a</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3986a</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3989a</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3993a</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3994a</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3995a</td>
<td>39</td>
<td>AKT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3996a</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3997a</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD3999a</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4000a</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4004a</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4087a</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4092a</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4093a</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4094a</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4101a</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4104a</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4105a</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4108a</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4110a</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4111a</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD4134a</td>
<td>51</td>
<td>AKT2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Triple Negative Cases

<table>
<thead>
<tr>
<th>Sample</th>
<th>TOTAL mutations</th>
<th>AKT1</th>
<th>AKT2</th>
<th>CDH1</th>
<th>GATA3</th>
<th>KRAS</th>
<th>MAP2K4</th>
<th>NF1</th>
<th>PIK3CA</th>
<th>PTEN</th>
<th>RB1</th>
<th>SETD2</th>
<th>STK11</th>
<th>SMAD4</th>
<th>TP53</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD3987a</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4002a</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4003a</td>
<td>126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4091a</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4098a</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4102a</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4107a</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4109a</td>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4113a</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4130a</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
</tr>
<tr>
<td>PD4133a</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RB1</td>
</tr>
</tbody>
</table>
Whole ‘exome’ sequencing

Rearrangement screens

Whole genome shotgun
Tumour-specific rearrangements
Plasma DNA

Cancer

Normal tissues

DNA with tumour-specific rearrangement
Serial measurements

Estimated tumour DNA / mL serum (pg)

- Detectable at limit of sensitivity
- Undetectable

Months after diagnosis

- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

Chemotherapy:
- Ifosfamide + doxorubicin
- Bortezomib
- Paclitaxel

CT scan:
- Localised deposits around T9-10
- Widespread soft-tissue metastases

Rearrangement 1
Rearrangement 2
Whole 'exome' sequencing
Rearrangement screens
Whole genome shotgun
Comprehensive catalogues of somatic mutations in cancer

• Detection of all classes of somatic variant base substitutions
  insertions and deletions
  rearrangements
  copy number

• Detection of somatic variants in all genomic regions
  coding exons
  noncoding exons
  introns
  intergenic regions
### Somatic mutations in Colo-829 and NCI-H209

<table>
<thead>
<tr>
<th></th>
<th>Colo-829</th>
<th>NCI-H209</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Somatic Substitutions</strong></td>
<td>33,345</td>
<td>22,910</td>
</tr>
<tr>
<td><strong>Insertion/deletion</strong></td>
<td>66</td>
<td>65</td>
</tr>
<tr>
<td><strong>Rearrangements</strong></td>
<td>37</td>
<td>58</td>
</tr>
<tr>
<td><strong>CODING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missense</td>
<td>168</td>
<td>92</td>
</tr>
<tr>
<td>silent</td>
<td>104</td>
<td>36</td>
</tr>
<tr>
<td>nonsense</td>
<td>14</td>
<td>4</td>
</tr>
</tbody>
</table>
ICGC Goal: To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.


ICGC Public Presentation April 15, 2010: PDF | PPT

International Cancer Genome Consortium (ICGC) Goals, Structure, Policies and Guidelines: HTML | PDF

Members of the ICGC
Committed Projects to date: 23
Some Key (immediate) Issues and Challenges

Tumour Complexity/Heterogeneity

Ability to utilise small biopsy and FFPE samples

Informatics

....INFORMATICS
Cancer Genome Project

Sally Bamford
David Beare
Graham Bignell
Nidhi Bindal
Adam Butler
Helen Davies
Charlotte Dunham
Simon Forbes
Christopher Greenman
Claire Hardy
Mingming Jia
David Jones
Chai Yin Kok
Calli Latimer
King Wai Lau
Kenric Leung
Meng-lay Lin
Mark Maddison
John Marshall
David McBride
Stuart McLaren
Andrew Menzies
Lina Chen
Juok Cho
Danushka Galappaththige
Catherine Leroy

Laura Mudie
Keiran Raine
Rebecca Shepherd
Lucy Stebbings
Philip Stephens
Patrick Tarpey
Jonathan Teague
Tony Webb
Stacey Price
Ignacio Varela

Mathew Garnett
Anne McLaren-Douglas
Andrew Bar thorpe
Patrick Brien
Laura Hirst
Frances Jewitt
Tatiana Mironenko
Jorge Soares
Ian Thompson

Serena Nik-Zainal
Elizabeth Murchison
Jennifer Yen

Sancha Martin
Angela Macharia
Wendy McLaughlin

Erin Pleasance
Gillian Dalgliesh

ccRCC/PBRM1 Studies
Bin Teh
Dachuan Huang
Choon Kiat Ong
Waraporn Chan-on
Chutima Subimerb
Kyle Furge
Karl Dykema

ICGC Breast Cancer Working Group

Mike Stratton, Peter Campbell, Ultan McDermott
Paired-end Reads

Random 400bp fragments of matching cancer and normal genomes
Hundreds of millions of individual molecules sequenced simultaneously

Map paired sequences back to reference genome
Compare and look for tumour-specific variants
Assay design

Copy number

Chr 11

780 kb deletion

70.7Mb

71.8Mb

1st round PCR

Nested real-time PCR