Sequencing the “Cancer Genome”

Richard K. Wilson, Ph.D.
Washington University
School of Medicine

rwilson@watson.wustl.edu
PCR-based re-sequencing

- List of candidate genes
- Large collection of patient samples
TCGA: Targeted Re-sequencing in GBM

• Phase I (n=601)
  – Published literature and unpublished data in the GBM community
  – Emphasize genes with genetic evidence of alterations (e.g. amplification/deletion; mutation) or with clinical outcome correlation

• Phase II (n=725)
  – Initial GBM genome characterization data by CGCC
  – Genetic elements of interest (GEOI) defined by integrative analysis of copy number and expression in GBM across multiple platforms (include both coding mRNA and non-coding microRNA)
  – Conserved regions across evolution
  – Tumor-specific spliced variants
GBM: Somatic mutations

(Shared 20 genes, 84 tumors; orthogonal validation pending)

Pink: 1 mutation  Red: 2 mutations

*Indels included
GBM: Somatic mutations

(601 Phase I genes, 84 tumors; orthogonal validation pending)
## Next Generation DNA Sequencing Technologies

Genome size: 3000 Mb  
Req'd coverage: 6 12 25

<table>
<thead>
<tr>
<th></th>
<th>3730</th>
<th>454 FLX</th>
<th>Solexa</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp/read</td>
<td>600</td>
<td>250</td>
<td>32</td>
</tr>
<tr>
<td>Reads/run</td>
<td>96</td>
<td>400,000</td>
<td>40,000,000</td>
</tr>
<tr>
<td>bp/run</td>
<td>57,600</td>
<td>100,000,000</td>
<td>1,280,000,000</td>
</tr>
<tr>
<td>/runs req'd</td>
<td>312,500</td>
<td>360</td>
<td>59</td>
</tr>
<tr>
<td>Cost per run</td>
<td>$48</td>
<td>$6,800</td>
<td>$9,300</td>
</tr>
<tr>
<td>Total cost</td>
<td>$15,000,000</td>
<td>$2,448,000</td>
<td>$544,922</td>
</tr>
</tbody>
</table>

© R.K.Wilson 2008
454-based Mutation Detection

- DNA from X tumor samples
- Pooled with equal concentration
- PCR amplification with Y primer pairs
- Pool PCR products
- 454 sequencing
- SNP/Indel Detection Using ssahaSNP and BreakPointRead

- Reads with G12 mutation in KRAS
- Reads with 15 bp deletion in EGFR
BCM/Nimblegen Approach to Exon Sequencing

1. Elute gDNA
2. Fragment and hybridize to Nimblegen capture array
3. Elute
4. 454 Sequencing
5. Analyze Exon sequences

Courtesy of R.A. Gibbs
Solexa-based Whole Genome Sequencing

Solexa flow cell

~50M clusters are sequenced per flow cell.
“933124”

- 57 y/o Caucasian female
- *De novo* M1 AML
- >95% blasts in initial BM sample
- Relapsed and died at 11 months
- Normal cytogenetics
- No LOH/CNV on Affy 6.0 SNP array
- Informed consent for whole genome sequencing
AML Tumor Genome Sequence

- 77% diploid coverage was obtained for an AML M1 tumor genome with 22x sequence coverage (97.4% haploid coverage) using Solexa (Illumina) sequencers.
- 10x sequence coverage of normal genome (epidermal).
- 2.1M sequence variants were detected (consistent with the known human polymorphism rate).
- ~495,000 are novel variants: SNPs vs. somatic mutations
- 3,731 non-synonymous coding variants were detected; most are likely rare SNPs (compare to germline & cDNA).
- Two somatic mutations have been detected and confirmed; both are coding sequence insertions.
- Currently validating a small number (<30) of additional candidate somatic mutations...
AML Tumor Genome Sequence

- FLT3: FMS-like tyrosine kinase
- NPM1: nucleophosmin
  - Both are small insertion mutations.
  - Both have been previously implicated in myeloproliferative disease.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>15</th>
<th>17</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td>8</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWI-EAS68__10151_5_84_616_593</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td>27</td>
</tr>
</tbody>
</table>

© R.K. Wilson 2008
Detection of alternative splicing

454/Solexa AML cDNA sequencing: 350 novel transcripts in ~300 genes.
Acknowledgments

• **WU Genome Sequencing Center**
  Elaine Mardis, Li Ding, Mike McLellan, Brian Dunford-Shore, Ken Chen, David Dooling, Sean McGrath, Matt Hickenbotham, Lisa Cook, Tracie Miner, Lucinda Fulton, *et al.*

• **WU Siteman Cancer Center**
  Tim Ley, Mark Watson, Matt Walter, Rhonda Ries, Jackie Payton, John DiPersio, Dan Link, Michael Thomasson, Tim Graubert, Sharon Heath

• **TSP/TCGA Colleagues**
  Baylor HGSC, Broad Institute, many others…

• **Funding sources**
  NHGRI (Wilson), NCI (Ley), Alvin J. Siteman (AML WGS)

[genome.wustl.edu](http://genome.wustl.edu)