Sequencing Cancer Gene Mutations

Ramakrishna Sompallae
Department of Pathology & Bioinformatics core, Iowa Institute of Human Genetics
University of Iowa
Overview

- General introduction
- Sequencing workflow
- Data Analysis
- Future directions
Genetic aberrations as basis of cancer

- **Initiation**
  - Mutation
- **Promotion**
  - Cell proliferation
  - Mutation or epigenetic change
  - Proliferation and selection
- **Tumor progression**
  - Mutation or epigenetic change
  - Proliferation and selection

- Mutations
  - SNVs – missense, nonsense, INDELs
  - Long INDELs
  - Copy number changes
  - Translocations

- Epigenetic changes
  - Methylation
  - Acetylation
Landscape of somatic aberrations

Mutational heterogeneity

Diagnosis & Personalized therapy

- **Diagnostic**
  - PML-RARA – APL
  - c-KIT e11 - GIST
  - Papillary thyroid carcinoma (Metastatic) - BRAFV600E

- **Prognostic (risk stratification)**
  - AML – FLT3(poor), NPM1(favourable) (gain of function)
  - CLL – del 11q22-23 (ATM, poor), 13q14.3(miR, favourable) & 17p13(P53, poor)

- **Predictive (for Targeted therapy)**
  - EGFR1, e18-21 – TK inhibitor, gefitinib, erlotinib
  - BRAFV600E – vemurafinib (TK inhibitor)

- **Monitoring (stable somatic mutations)**
  - BCR-ABL1 – CML(imatinib)
Personalized therapy

1. Database
2. Identify targets
3. Sequencing of tumor tissue
4. Bioinformatics analysis

Diagram showing the process of personalized therapy.
Current methods of testing

- Sanger sequencing
- QPCR
- Mass spectrometry
Ion torrent sequencing
Advantages of NGS

Four main advantages are:

- High throughput (multiple genes)
- Cost
- Sample size
- Accuracy

NGS is significantly cheaper, quicker, needs significantly less DNA and is more accurate and reliable than Sanger sequencing.
Ion Torrent PGM workflow

Nucleotide incorporates into DNA

Hydrogen ion is released

H⁺
Ampliseq cancer panel

- 22,027bp
- 207 amplicons
- 50 Oncogenes
- ~2,800 somatics mutations
- **Melanoma** [BRAF, KRAS, MAPK, MET, NRAS, PIK3CA, PTEN], **Endometrial & Ovarian** [KRAS, PTEN, TP53], **Pancreatic Cancer** [KRAS, NRAS, PIK3CA, TP53]

### The Ion AmpliSeq™ Cancer Panel targets 50 genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL1</td>
<td>EZH2</td>
<td>JAK3</td>
<td>PTEN</td>
</tr>
<tr>
<td>AKT1</td>
<td>FBXW7</td>
<td>IDH2</td>
<td>PTPN11</td>
</tr>
<tr>
<td>ALK</td>
<td>FGFR1</td>
<td>KDR</td>
<td>RB1</td>
</tr>
<tr>
<td>APC</td>
<td>FGFR2</td>
<td>KIT</td>
<td>RET</td>
</tr>
<tr>
<td>ATM</td>
<td>FGFR3</td>
<td>KRAS</td>
<td>SMAD4</td>
</tr>
<tr>
<td>BRAF</td>
<td>FLT3</td>
<td>MET</td>
<td>SMARCB1</td>
</tr>
<tr>
<td>CDH1</td>
<td>GNA11</td>
<td>MLH1</td>
<td>SMO</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>GNAS</td>
<td>MPL</td>
<td>SRC</td>
</tr>
<tr>
<td>CSF1R</td>
<td>GNAQ</td>
<td>NOTCH1</td>
<td>STK11</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>HNF1A</td>
<td>NPM1</td>
<td>TP53</td>
</tr>
<tr>
<td>EGFR</td>
<td>HRAS</td>
<td>NRAS</td>
<td>VHL</td>
</tr>
<tr>
<td>ERBB2</td>
<td>IDH1</td>
<td>PDGFRA</td>
<td></td>
</tr>
<tr>
<td>ERBB4</td>
<td>JAK2</td>
<td>PIK3CA</td>
<td></td>
</tr>
</tbody>
</table>
Versatile platform

PGM

314
Ion 314™ Chip
1 million wells
30 to 50 Mb output
for 200-base sequencing

316
Ion 316™ Chip
6 million wells
300 to 600 Mb output
for 200-base sequencing

318
Ion 318™ Chip
11 million wells
600 Mb to 1 Gb output
for 200-base sequencing

Proton

Proton I*
The Ion Proton™ I Chip
165 million wells
2 human exomes

Proton II*
The Ion Proton™ II Chip
660 million wells
1 human genome
Ampliseq Workflow

Starting Material DNA/RNA

Library Construction

Templated Bead Preparation

Sequence + Run Evaluation

Data Analysis

Prepare Library

Clonal Amplification

Isolate Positive Ion Sphere™ Particles

Load Chip and Sequence

Starting Material DNA/RNA

Library Construction

Templated Bead Preparation

Sequence + Run Evaluation

Data Analysis

Prepare Library

Clonal Amplification

Isolate Positive Ion Sphere™ Particles

Load Chip and Sequence
Ion Torrent PGM workflow

Prepare Library → Clonal Amplification → Isolate Positive Ion Sphere™ Particles → Load Chip and Sequence

**BIOINFORMATICS**
- Raw sequence data
- Sequence Mapped (TMAP)
- Variants called (TSVC)

**Torrent Suite**

**Custom pipeline**
- Annotation of Variants
  - variant quality, > 25
  - Coverage > 100
  - Nonsynonymous etc, Minor allele freq ≤ 0.1
- Filtered list of Variants
- Report

ANNOVAR – to annotate all variant calls
Validation samples

11 adenocarcinoma
3 melanoma
2 acute monoblastic leukemia
2 gastrointestinal stromal tumor
3 unknown
Platform validation

log10 Coverage of amplicons

samples
Annotation and Filtering of Variants

Custom pipeline

- Annotation of Variants
  -_variant quality, > 25
  - Coverage > 100
  - Nonsynonymous etc,
  - Minor allele freq ≤ 0.1

- Filtered list of Variants
- Report

- Filter out low frequency and synonymous variants
- Filter out low quality and common variants
- Prioritize remaining

Prioritize remaining variants using:
- SIFT
- POLYPHEN-2
- PhyloP
- Mutationtaster
- LRT
- GERP++
<table>
<thead>
<tr>
<th>ID</th>
<th>UI Run 1</th>
<th>UI Run 2</th>
<th>Mayo Ion</th>
<th>Ion Torrent Run</th>
<th>Previous mutation</th>
<th>Previous assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>R40-CHPv2_BC002</td>
<td>R42-CHPv2_BC001</td>
<td>IX001_GUT-120</td>
<td>GUT-120</td>
<td>EGFR c.2240-2257del18 p.L747_P753del</td>
<td>Sanger</td>
</tr>
<tr>
<td>M2</td>
<td>R44-CHPv2_BC010</td>
<td>ND</td>
<td>IX002_GUT-120</td>
<td>GUT-120</td>
<td>NPM1/FLTD835</td>
<td>Fragment analysis</td>
</tr>
<tr>
<td>M5</td>
<td>R42-CHPv2_BC004</td>
<td>R43-CHPv2_BC004</td>
<td>IX005_GUT-120</td>
<td>GUT-120</td>
<td>KRAS c.34G&gt;T p.G12C</td>
<td>SNaPshot</td>
</tr>
<tr>
<td>M6</td>
<td>R42-CHPv2_BC008</td>
<td>R43-CHPv2_BC008</td>
<td>IX006_GUT-120</td>
<td>GUT-120</td>
<td>NPM1/FLTD835</td>
<td>Fragment analysis</td>
</tr>
<tr>
<td>M7</td>
<td>R40-CHPv2_BC005</td>
<td>ND</td>
<td>IX007_GUT-120</td>
<td>GUT-120</td>
<td>KRAS c.35G&gt;T p.G12V</td>
<td>SNaPshot</td>
</tr>
<tr>
<td>M8</td>
<td>R45-CHPv2_BC009</td>
<td>ND</td>
<td>IX008_GUT-120</td>
<td>GUT-120</td>
<td>NPM1/FLTD835</td>
<td>Fragment analysis</td>
</tr>
<tr>
<td>M9</td>
<td>R45-CHPv2_BC015</td>
<td>ND</td>
<td>IX009_GUT-120</td>
<td>GUT-120</td>
<td>BRAF c.1799_1799GT&gt;AA p.V600K</td>
<td>Sanger</td>
</tr>
<tr>
<td>M10</td>
<td>R40-CHPv2_BC003</td>
<td>ND</td>
<td>IX010_GUT-120</td>
<td>GUT-120</td>
<td>KRAS c.34G&gt;A p.G12S</td>
<td>SNaPshot</td>
</tr>
<tr>
<td>M12</td>
<td>R45-CHPv2_BC013</td>
<td>R47-CHPv2_BC005</td>
<td>IX012_GUT-121</td>
<td>GUT-121</td>
<td>KIT c.1679-1681del p.Val560del</td>
<td>Sanger</td>
</tr>
<tr>
<td>M13</td>
<td>R45-CHPv2_BC014</td>
<td>R46-CHPv2_BC013</td>
<td>IX013_GUT-121</td>
<td>GUT-121</td>
<td>KIT c.1504-1509dupGCCTAT p.Al502_tyr503dup</td>
<td>Sanger</td>
</tr>
<tr>
<td>M14</td>
<td>R45-CHPv2_BC012</td>
<td>R47-CHPv2_BC006</td>
<td>IX014_GUT-121</td>
<td>GUT-121</td>
<td>EGFR c.2237_2254del18 p.E746_S752&gt;A</td>
<td>Sanger</td>
</tr>
<tr>
<td>M15</td>
<td>R42-CHPv2_BC007</td>
<td>R43-CHPv2_BC007</td>
<td>IX015_GUT-121</td>
<td>GUT-121</td>
<td>EGFR c.2235_2249del p.E746_A750del</td>
<td>Sanger</td>
</tr>
<tr>
<td>M16</td>
<td>R42-CHPv2_BC006</td>
<td>R43-CHPv2_BC006</td>
<td>IX016_GUT-121</td>
<td>GUT-121</td>
<td>KIT c.1679_1681delTTG p.V560del</td>
<td>Sanger</td>
</tr>
<tr>
<td>M17</td>
<td>R40-CHPv2_BC004</td>
<td>ND</td>
<td>IX001_GUT-121</td>
<td>GUT-121</td>
<td>KIT c.1676T&gt;A p.V559D</td>
<td>Sanger</td>
</tr>
<tr>
<td>M18</td>
<td>R40-CHPv2_BC001</td>
<td>ND</td>
<td>IX002_GUT-121</td>
<td>GUT-121</td>
<td>BRAF c.1789_1790CT&gt;TC p.L597S</td>
<td>Sanger</td>
</tr>
<tr>
<td>M19</td>
<td>R40-CHPv2_BC008</td>
<td>ND</td>
<td>IX003_GUT-121</td>
<td>GUT-121</td>
<td>EGFR c.2235_2249del p.Glu746_Ala750del</td>
<td>Sanger</td>
</tr>
<tr>
<td>M20</td>
<td>R40-CHPv2_BC007</td>
<td>ND</td>
<td>IX004_GUT-121</td>
<td>GUT-121</td>
<td>EGFR c. 2236_2250del15 c.2361G&gt;A p.Q787Q</td>
<td>Sanger</td>
</tr>
</tbody>
</table>
Concordance with Mayo Sequencing

% calls in agreement

M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15 M16 M17 M18 M19 M20 M21

Pct_obs_Run1
Pct_obs_Run2
# Validation data

<table>
<thead>
<tr>
<th># Previously identified mutation</th>
<th># PGM detected</th>
<th># avg variants</th>
<th>sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 KIT ex. 10, 11, 17</td>
<td>10</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>9 KRAS ex. 2</td>
<td>9</td>
<td>1.9</td>
<td>100%</td>
</tr>
<tr>
<td>8 EGFR ex. 18-21</td>
<td>8</td>
<td>1.0</td>
<td>100%</td>
</tr>
<tr>
<td>7 BRAF ex. 15 (V600E/K, K601E)</td>
<td>7</td>
<td>0.7</td>
<td>100%</td>
</tr>
<tr>
<td>4 PIK3CA ex. 21</td>
<td>4</td>
<td>3.0</td>
<td>100%</td>
</tr>
<tr>
<td>3 FLT3 D835 (ex. 20)</td>
<td>3</td>
<td>1.0</td>
<td>100%</td>
</tr>
<tr>
<td>3 NPM1 ex. 12</td>
<td>3</td>
<td>3.0</td>
<td>100%</td>
</tr>
<tr>
<td>2 CTNNB1 ex. 3</td>
<td>2</td>
<td>0.5</td>
<td>100%</td>
</tr>
</tbody>
</table>
## Validation with Coriell sample

<table>
<thead>
<tr>
<th>Barcode Name</th>
<th>Sample</th>
<th>Bases</th>
<th>≥ Q20</th>
<th>Reads</th>
<th>Mean Read Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>No barcode</td>
<td>NOSM</td>
<td>3.7M</td>
<td>2.7M</td>
<td>40365</td>
<td>93 bp</td>
</tr>
<tr>
<td>IonXpress_017</td>
<td>Coriell 10851_Lot6</td>
<td>74.3M</td>
<td>64.8M</td>
<td>698559</td>
<td>106 bp</td>
</tr>
<tr>
<td>IonXpress_018</td>
<td>Coriell 12156_LotA12</td>
<td>86.9M</td>
<td>76.5M</td>
<td>807667</td>
<td>107 bp</td>
</tr>
<tr>
<td>IonXpress_019</td>
<td>Coriell 12878_LotE2</td>
<td>90.4M</td>
<td>78.6M</td>
<td>842269</td>
<td>107 bp</td>
</tr>
<tr>
<td>IonXpress_020</td>
<td>12-901</td>
<td>69.7M</td>
<td>61.4M</td>
<td>691350</td>
<td>100 bp</td>
</tr>
</tbody>
</table>
Validation with Coriell sample

![Validation with Coriell sample](image-url)
Detecting variations at lower frequency

Tumor A
(KIT 6bp ins)

Tumor B
(EGFR 18bp del)

Graph showing the variation frequency in KITins and EGFR del between Tumor A and Tumor B.
Mixing Sample

Tumor A (PI3CA\textsubscript{SNP}) and Tumor B (HRAS\textsubscript{SNP})
Summary

- Validation with both FFPE and blood samples.
  - High concordance with variant calls from other services

- With our current workflow we could detect
  - SNVs and MNVs
  - Small and long Indels (18bp)

- On our way to implementation
  - Optimizing software, Validation reports, SOPs,
Other useful resources
Acknowledgments

Dept. of Pathology
- Dr. Jon heusel
- Natasha Guseva

IIHG - Bioinformatics Core
- Tom Bair
- Diana Kolbe