Sequencing Technologies: Overview & Capabilities

Chris Daum
Genomic Technologies Workshop
March 22, 2016
cgdaum@LBL.gov
Sequencing Technologies Group

- Define Project
- Obtain Metadata
- Product catalog

work order & flow

Sequencing Technology ($15M)

Sample Management (5 FTEs)
Library Construction (11 FTEs)
Library QC, qPCR, pooling (3 FTEs)
Sequencing (5 FTEs)

Area Leads (3); Resource & NTI manager (1); R&D Project manager (1)

- Streamlined Process from Sample In to Sequence Data Out
- Perform Process Optimization & Development:
  - New Preps, Applications, Sequencing Technologies
  - Continuous Improvement & Lean Manufacturing Six-Sigma
Sequencing Technologies: Sample to Data

Sample QC/ aliquot → Library preparation → Sequencing: Illumina / Pacbio

Supported by ITS-GLS LIMS system
Staying State of the Art

Sanger Sequencing to Next-Gen Sequencing by Synthesis

- **ABI 3730 Sanger reduced**
- **MegaBase Sanger offline**
- **454 in Production**
- **454 Titanium**
- **Solexa early access**
- **Solexa in Production**
- **SOLiD early access**
- **Illumina GAIx**
- **Illumina HiSeq 2000**
- **PacBio early access**
- **Oxford Nanopore early access**

Timeline:
- **07/2005**
- **04/2007**
- **07/2008**
- **05/2009**
- **12/2009**
- **11/2010**
- **3/2011**
- **12/2011**
- **04/2013**
- **06/2014**

- **454 offline**
- **Illumina MiSeq**
- **Illumina HiSeq 1T**
- **PacBio RSII**
- **Illumina HiSeq 2500**
JGI Yearly Sequence Output

FY Total Bases (Gb) Sequenced - All Platforms

As of March 21, 2016

44,000 human genome equivalents!

<table>
<thead>
<tr>
<th>FY</th>
<th>Total Bases (Gb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY2005</td>
<td>34</td>
</tr>
<tr>
<td>FY2006</td>
<td>33</td>
</tr>
<tr>
<td>FY2007</td>
<td>39</td>
</tr>
<tr>
<td>FY2008</td>
<td>126</td>
</tr>
<tr>
<td>FY2009</td>
<td>1,004</td>
</tr>
<tr>
<td>FY2010</td>
<td>6,041</td>
</tr>
<tr>
<td>FY2011</td>
<td>29,903</td>
</tr>
<tr>
<td>FY2012</td>
<td>56,115</td>
</tr>
<tr>
<td>FY2013</td>
<td>70,863</td>
</tr>
<tr>
<td>FY2014</td>
<td>100,608</td>
</tr>
<tr>
<td>FY2015</td>
<td>143,177</td>
</tr>
<tr>
<td>FY2016</td>
<td>58,172</td>
</tr>
</tbody>
</table>
# JGI Sequencing Platforms Portfolio

<table>
<thead>
<tr>
<th></th>
<th>Illumina HiSeq 1T</th>
<th>Illumina HiSeq 2500</th>
<th>Illumina HiSeq 2000</th>
<th>Illumina NextSeq 500</th>
<th>Illumina MiSeq</th>
<th>PacBio RSII</th>
<th>PacBio Sequel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Units</strong></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><strong>Reads (Single-Read)</strong></td>
<td>&gt;1,500 Million per Flowcell</td>
<td>200 Million per Flowcell</td>
<td>&gt;1,000 Million per Flowcell</td>
<td>400 Million per Flowcell</td>
<td>&gt;10 Million per Flowcell</td>
<td>0.06 Million per SMRT Cell</td>
<td>0.4 Million per SMRT Cell</td>
</tr>
<tr>
<td><strong>Readlength</strong></td>
<td>2 X 150bp Max*</td>
<td>2 X 250bp Max</td>
<td>2 X 150bp Max*</td>
<td>2 X 150 Max</td>
<td>2 X 300bp Max</td>
<td>&gt;12,500bp Avg; &gt;40,000bp Max</td>
<td>12,000bp Avg; &gt;40,000bp Max</td>
</tr>
<tr>
<td><strong>Total Bases</strong></td>
<td>500 Gb per Flowcell</td>
<td>130 Gb per Flowcell</td>
<td>350 Gb per Flowcell</td>
<td>&gt;100 Gb per Flowcell</td>
<td>5-20 Gb per Flowcell</td>
<td>&gt;0.4 Gb per SMRT Cell for 2hr runs; &gt;0.8Gb for 4hr runs</td>
<td>&gt;2.7 Gb per SMRT Cell for 2hr runs; &gt;5.0 Gb for 4hr runs</td>
</tr>
<tr>
<td><strong>Run Time</strong></td>
<td>7 Days for 2 X 150</td>
<td>4.5 Days for 2 X 250</td>
<td>16 Days for 2 X 150</td>
<td>1 Days for 2 X 150</td>
<td>2 Days for 2 X 300</td>
<td>0.08-0.12 Days (2-4 hours)</td>
<td>0.08-0.12 Days (2-4 hours)</td>
</tr>
<tr>
<td><strong>Applications</strong></td>
<td>Primary Sequence Generator at JGI</td>
<td>Rapid output HiSeq</td>
<td>Supplement / Backup Platform</td>
<td>Rapid mid-range output; Single Cell</td>
<td>16S/18S iTags, Library QC, R&amp;D</td>
<td>Assembly improvement, de novo, SynBio validation, methylation/epigenetics</td>
<td>Early Access Testing</td>
</tr>
</tbody>
</table>
Major Investment in Automation: Process

Sample Archive & Retrieval

Sample QC & Aliquot

Library Construction

Library QC, qPCR, pooling

SAM

Starlet x 2

Sciclon G3 x3

Star x 2

Automated Decapper

BioMicroLab Volume Check

BioTek Synergy H1 Plate Reader

5ul PCR

Labcyte Echo

Maximize Consistency, Throughput and Reliability
Automated Library Creation

• Plate based Automated library preps:
  – Implemented into Production in late 2011
  – 3 PerkinElmer Sciclone NGS robots
  – 12 production sample prep methods are supported
    • >24,000 sample sequencing libraries prepared in last year

• Supporting Equipment:
Sciclone Automated Preps

Supported Preps

- Illumina gDNA PCR-free WGS Fragment Library prep
- Illumina TruSeq RNA-seq stranded library preps:
  - PolyA selection of mRNA for eukaryotes
  - rRNA depletion for microbes & metatranscriptomes
- Illumina small RNA & miRNA for eukaryotes
- Illumina iTags (16S proks; 18S euks; Fungal ITS)
- Illumina Exome Capture (NimbleGen SeqCap) prep for targeted resequencing
- Illumina NexteraXT prep
- PacBio DNA 2kb libs, and >10kb libs with enzymatic shearing

Preps in Development

- Illumina Methyl-Seq (bisulfite conversion) prep
- Illumina 3' RNAseq prep
Continued Growth in Number of Samples Handled

**Samples Received**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>420</td>
<td>1,380</td>
<td>2,300</td>
<td>5,200</td>
<td>8,675</td>
<td>16,659</td>
<td>22,636</td>
</tr>
</tbody>
</table>

**Library Plates Processed**

<table>
<thead>
<tr>
<th>Year</th>
<th>FY2012</th>
<th>FY2013</th>
<th>FY2014</th>
<th>FY2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>iTag Plates</td>
<td>57</td>
<td>91</td>
<td>223</td>
<td>327</td>
</tr>
<tr>
<td>RNA Plates</td>
<td>16</td>
<td>18</td>
<td>37</td>
<td>71</td>
</tr>
<tr>
<td>DNA Plates</td>
<td>41</td>
<td>73</td>
<td>79</td>
<td>72</td>
</tr>
</tbody>
</table>

**Libraries Created**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>382</td>
<td>1,239</td>
<td>2,009</td>
<td>4,552</td>
<td>7,840</td>
<td>19,077</td>
<td>27,802</td>
</tr>
</tbody>
</table>
A Growing Portfolio of Library Capabilities

**Genome:** DNA
- WGS
- WGS: Tight insert
- LMP
- Pacbio long reads
  - Genome assembly
  - Structural variation
  - Comparative genomics
  - Genotyping

**Transcriptome:** RNA
- Stranded RNA-seq
  - Poly A
  - rRNA depletion
  - FFPE/LCM
  - smRNA
  - Iso-seq
  - Gene annotation
  - Gene expression

**Epigenome:** DNA-protein
- Bisulfite-seq
- FAIRE-seq
- ChIP-seq
- ChIA-PET
  - Methylation
  - Gene regulation
  - Chromatin conformation
  - Protein binding

Chew Yee Ngan
A Growing Portfolio of Library Capabilities

Genome Fragments
- 270bp
- 500bp
- Low input- Nextera

Genome Fragments
- Tight insert 400bp
- Tight insert 800bp

Genome Fragments
- Pacbio 2kb
- Pacbio 3kb
- Pacbio >10kb
  - Low input 10kb
  - Pacbio 20kb
  - Oxford Nanopore

Genome assembly
- 100ng
- 1ng

Example: DNA Products

Tight insert
- 270bp
- 500bp
- Low input- Nextera

Tight insert
- 400bp
- 800bp

Long Mate Pairs
- 2.5kb
- 4kb
- 8kb

Long Reads

Long Mate Pairs

Genetic Diversity

Structural variation

Comparative genomics

Chew Yee Ngan
RNA Sample Prep: Gene expression

- **RNA-seq**
  - RNA amount:
    - 1ug
    - 100ng
    - 10ng
  - PCR amplification:
    - 8
    - 10
    - 15
  - Poly A selection rRNA depletion (probes)

- **Small RNA-seq**
  - Eukaryotes: 20-40bp
  - Prokaryotes: 50-150bp

**Strandedness**
- Forward strand
- Reverse strand

**Reproducibility**
- Pearson correlation = 0.9992

**Coverage**
- Novel miRNA

Chew Yee Ngan
Epigenomic: DNA modifications

5 Methyl C
Bisulfite sequencing

Methyl A
Pacbio sequencing

Unmodified DNA template

Modified DNA template

Library preparation/sequencing

Delay in base incorporation opposite methyl adenine.

Stephen Mondo’s Presentation
Epigenomic: Chromatin structure

Regulatory elements/open chromatin: FAIRE-seq

Histone modifications

Protein binding sites: ChIP-seq

FAIRE
Cells crosslinked with formaldehyde

Shear by sonication
Perform phenol/chloroform extraction

Construct library
Sequence

Cross-link & shear

DNA-protein complexes

DNA sequencing

Purify

compared our data to DNase-hypersensitive sites identified in and with transposase sequence preference. Ties also correlated well with markers of active chromatin and not of DNase and ATAC-seq peaks (and the majority of reads within peaks came from intersections). We found that ATAC-seq paired-end reads produced detailed information of chromatin as defined by previous models. Normalizing the insert size distribution according to functional classes disclosed nucleosome positions from ~3–5 orders of magnitude more cells than ChIP-seq (Duke) and FAIRE-seq (University of North Carolina). This fragment size distribution also showed clear periodicity equal to the helical pitch of DNA. Single nucleosome–associated fragments. Transcribed and promoter–flanking regions were enriched for longer multinucleosomal fragments. Transcribed and promoter–flanking regions were enriched for longer multinucleosomal fragments. Transcribed and promoter–flanking regions were enriched for longer multinucleosomal fragments. ATAC-seq fragment sizes generated from GM12878 nuclei. (500 cells) showed highly reproducible between technical replicates (5,000 or 500 human nuclei as starting materials). This result demonstrated that ATAC-seq provides genome-wide information on chromatin state previously defined by ChIP-seq. Furthermore, data generated using three different sources of starting material correlated highly between ATAC-seq and two different sources of ChIP-seq data. The ratio similar to that of DNase-seq, whose data were generated from ~100× the number of cells (50,000 cells per replicate). This fragment size distribution also showed clear periodicity persists to DECEMBER 2013.

New Technologies

• **3’ RNA-seq**
  - **Total RNA**

• **ATAC-seq (Assay of Transposase Accessible Chromatin)**
  - Identification of genome wide regulatory elements

• **Gene expression study**
• **Reduced sequencing cost**

- **Coverage across transcripts**

- **Reduced sample requirement**
- **Organism dependent**
  - requires customization

Major User Request: Decrease Sample Amount Requirements

The need for cheaper nano-scale sample prep

Input; cost

Throughput; consistency

Microfluidic platform

2014-2016 ETOP
Blainey Lab MIT

Fluidigm C1™
Echo labcyte
GenCell CLiC
Illumina Neoprep

Closed Commercial Microfluidic Platforms
Emerging Sequencing Technologies

- Oxford Nanopore Technology – find out more from Juna Lee next!