Next Generation Sequencing
Illumina GA Workflow
Sample Prep Details

Robert Steen
Biopolymers Facility
Department of Genetics
Harvard Medical School

ABRF 2010
“I am sorry if this smacks of an investigator, hair on fire, running into your lab needing data yesterday for a grant submission….”

Jay Fox
Biopolymers Core Facility

Our Customers:
(over past 12 months)

5 Harvard Institutions
34 Non-Harvard Academic Institutions
6 Private Companies

* 438 Laboratories
* 1,611 Users

Our Platforms and Services:

* Affymetrix
* Agilent Bioanalyzer 2100
* Applied Biosystems 3730xl
* Applied Biosystems 7900HT
* Applied Biosystems SOLiD V3
* Axon GenePix 4000B
* Hamilton StarPlus Robot
* Two - Illumina Genome Analyzer II
* Luminex LS200
* NanoDrop ND-8000
* QIAcube Robot
* Tecan Genios

* DNA / RNA Analysis
* DNA / RNA Preps & Purification
* DNA Sequencing
* Gene Expression
* Micro RNA Analysis
* Next-Generation DNA Sequencing
  * ChIP-Seq / RIP-Seq
  * mRNA-Seq
  * small RNA-Seq
  * Exome Sequencing
  * Whole Genome Sequencing
* Protein Interaction Assays
* Quantitative PCR
* SNP / Microsatellite Genotyping
Applications

Next Gen-Omics

- Genome Resequencing
- mRNA Tag Profiling
- Methylation Analysis
- Small RNA Identification
- Functional Elements (ChIP-Seq, DNase-Seq)
- Transcriptome Sequencing

Slide Source: Nature Web
### Applications In Our Facility

<table>
<thead>
<tr>
<th>Applications</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Cells (~ 12 months)</td>
<td>92</td>
</tr>
<tr>
<td>Lanes</td>
<td>736</td>
</tr>
<tr>
<td>Samples</td>
<td>373</td>
</tr>
</tbody>
</table>

#### Lanes Run by Application

<table>
<thead>
<tr>
<th>Application</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChIP-Seq</td>
<td>97</td>
</tr>
<tr>
<td>Digital Gene Expression</td>
<td>68</td>
</tr>
<tr>
<td>Genomic (Single Read And Paired End of various types)</td>
<td>278</td>
</tr>
<tr>
<td>PhiX Genomic Control</td>
<td>123</td>
</tr>
<tr>
<td>Small RNA</td>
<td>24</td>
</tr>
<tr>
<td>Whole Transcriptome</td>
<td>114</td>
</tr>
</tbody>
</table>
NGS Workflow Summary

• LIMS & Website
  • Information Presentation
  • Sample Ordering
  • Sample Tracking
  • Data Management
  • Invoicing

• Sample Processing / Library Prep
  • Primary Sample QC
  • Library Sample Prep Protocols
  • Library QC

• Sequencing Run
  • Pre-Cluster QC
  • Cluster Generation
  • Sequencing
  • On Instrument QC

• Data Analysis & Distribution
  • Primary SCS 2.6 & RTA
  • Pipeline OLB V1.6
  • Internal Transfer, FTP, FEDEX HD
Next Generation Sequencing Services - Illumina GA II

We now offer ultra high throughput DNA sequencing services on the Illumina Genome Analyzer II (formerly known as the Solexa 1G) platform.

Funding for the instrument was generously provided by Harvard Medical School institutional support through the Taplin Funds for Discovery Program as well as funds from a five-department consortium.

This new service is available to all our users (both internal and external), but preference in queuing will be given to the laboratories that participated in the Taplin proposal.

If you have questions regarding the service please contact the Biopolymers Facility at nextgen@genome.med.harvard.edu.

- **Overview**
- **Applications**
- **Sample Submission Details**
- **Bioninformatics**
- **Pricing**
Do Your Customers Understand the Technology?

• New Customer or Someone New to NGS
  • Inquiry Form
  • Consultation -- This is often VERY important!

• Existing Customer or Knowledgeable User
  • Straight to Sample Submission
More Information is Better

Sample Assignment

| Sample Assignment | Sample name (ex. blank) | Organism | Library | Vol. (μl) | Date/Assay (μl) | Adaptor
|-------------------|-------------------------|----------|---------|-----------|-----------------|---------
| (example)         | eukaryotic              | Man musculus | generic | N         | 20-50           | 250     |
| 1                 |                         |           |         |           |                 |         |
| 2                 |                         |           |         |           |                 |         |
| 3                 |                         |           |         |           |                 |         |
| 4                 |                         |           |         |           |                 |         |
| 5                 |                         |           |         |           |                 |         |
| 6                 |                         |           |         |           |                 |         |
| 7                 |                         |           |         |           |                 |         |

Run Details

| Run/Flowcell Type | # of Barrs | Primer Name | Library
|-------------------|------------|-------------|---------
|                   | Single     |            |         |
|                   | Pair End   |            |         |
| 1                 | X          |            |         |
| 2                 | I          |            |         |
| 3                 | X          |            |         |
| 4                 | I          |            |         |
| 5                 | X          |            |         |
| 6                 | I          |            |         |
| 7                 | X          |            |         |

Library Types:
- sRNA: Grican
- sRNA-m: CHIR-m: Digester Genes
- sRNA-leaf: Digestion (XbaI) or (SpeI)
- F: Forward adapter: compatible with Single Read and Pair End

Additional Comments:

Delivered by: ____________________________ Date: ____________________________

Received by (BPF Staff member): ____________________________ Date (BPF): __________ Freezer (BPF): ________
Basic Sample Prep Outline

Isolate DNA / RNA (In our case customer always does this)

Fragment (Covaris, Bioruptor, Nebulize, Enzymes, Epicentre Transposase, etc.)

End Repair and Add “A” (Most cases)

Ligate Adaptors  (Most cases)

Size Select / Purify (Traditional Gels, Columns, E-Gel System, SPRI Beads)

Amplify  (No PCR Libraries | NATURE METHODS | VOL.6 NO.4 | APRIL 2009 )

Purify  (Columns, E-Gel)
Slide Source: Broad / Illumina
Important Garbage In = Garbage Out

Start with Good Quality DNA / RNA
Sufficient Quantity is Important
Require QC on Bioanalyzer Before Starting

Bad RNA

Good RNA

<table>
<thead>
<tr>
<th>Overall Results for sample 6: Shubha Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Area: 319.1 ng/ul</td>
</tr>
<tr>
<td>rRNA Ratio [28S/18S]: 0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fragment table for sample 6: Shubha Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>18S</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall Results for sample 5: Pr-014-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Area: 153.7 ng/ul</td>
</tr>
<tr>
<td>rRNA Ratio [28S/18S]: 1.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fragment table for sample 5: Pr-014-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>18S</td>
</tr>
<tr>
<td>28S</td>
</tr>
</tbody>
</table>
Stick to Your Lab’s Requirements

We go down to 2ug
We need 50ng

PCR Amplicons: 50ng
Other apps. evaluate on a case by case basis.
Fragmenting

We use Covaris S2 with mini-tube holder.
Precise and reproducible shearing.

Know what you are working with; Bioanalyze all samples.

Isolated ChIP DNA

After Shearing
Note About Kits

Our facility currently uses:

- Illumina kit for RNA-Seq library preps.
- NEB kit for Genomic, ChIP and Similar Preps.

Our customers often create their own kits from list of enzymes.

Important: Adaptor sequences matter (obviously).

Our lab only makes paired-end libraries now and we ask our users to do the same.

Very Important: Discuss barcoding / indexing.
- Potential for biasing reads (may or may not matter)
- Difficulty calling image data.
- Base balance your indexes or use a spike in such as PhiX.
Size Selection

Invitrogen’s E-Gel system with ibase.
- Very Fast ~15 mins.
- High recovery 50% to 80%.
- Very easy to collect multiple fractions (we collect 3-5).
- Also use in place of final column to clean up and verify.

ChIP DNA

Genomic DNA
Amplification

Important not to over-amplify.
Easy to introduce bias.

Currently using 10 to 18 cycles
Project / Sample dependent.

In place of QiaQuick PCR Purification Kit we use E-Gel.
Allows for efficient clean up, high recovery and library verification.

Would like to implement no-PCR libraries in coming year.
Library QC Part One

Agilent Bioanalyzer High Sensitivity DNA Chip.

- Library must be proper “size and shape”.
- Must have a minimum concentration.
Library QC Part Two

SYBR QPCR Assay

- Assay based on original design from Stuart Levine at MIT.
- Use BioA concentration to make initial 10nM Stock.
- Make Std Curve with Illumina PhiX Library (new lots need characterizing).
- Dilute curves and stds. 1:1000.
- Use KAPA BIOSYSTEMS SYBR Fast qPCR Kit.
- Run on AB 7900HT in 384 well plate.
- Based on CT derived concentrations, adjust libraries for clustering.

-Suggest using robot to reduce variability in std. curve and libraries.
-Investigating KAPA BIOSYSTEMS Library Quant Kit.

```
SYBR_F-AATGATACGGCGACCACCGA
5'  AATGATACGGCGACCACCGAATCTACACTCTTTCCCTACAGCTGTAGTCCGAGTCACACAGCTCTTCCGATCT................AGATCGGAAGAGCGTCGTAATGGGACACGCTTCTGTCTGCTCTGCTTG-3' 
3'  TTACTATAGCCGCTTGTTCCTTCTGAGATGTAGAAAGGGATGTACGAGAGACACAGCTAGA............TCTAGCCCTTCTGGAGATACGCAGCGAACGAGCAGAAAGACGCAACG-5' 
AGCATAACGCCGAGGAGACGCAAC-SYBR_R
```
Cluster and Sequence

Follow routine maintenance for Cluster Station and GA!

Always, do fluidics line checks on Cluster Station and GA!

Use fresh reagents and track lot #'s.
For longer runs, pre-mix cycle seq mix to reduce run variability

First base report is a good indicator of run quality but will not match final run analysis metrics exactly and that is okay.

Most often successful trouble shooting step when there are no or poor intensity clusters is to re-hyb the seq primer.
What Will Your Core Provide?

In Our Lab:

No detailed Bioinformatics provided.

We provide the middle of the pyramid.

We use Institutional Infrastructure for IT components.

Slide Image Source: Geospiza
Run SCS2.6 and RTA to determine real time QC Metrics for Run.

HOWEVER

All runs are re-analyzed on HMS Orchestra cluster. (HMS maintained cluster with 1,000’s of cpu’s.) Specifically: Seven IBM blades, 8 cores and 32G RAM each.)

WHY?

Seen an increased use of indexes and odd adaptors by users. Not base balancing the indexes / adaptors.

Significantly increase pass filter cluster yield by using ten cycles for cluster identification and specifying which base to start on. This is too compute intensive for RTA.
Thank You!