A Bioinformatics Core Matures: Growing and maintaining customer base while approaching cost-recovery

James Cavalcoli, PhD
University of Michigan
Ann Arbor
Outline

• Cost recovery

• Expanding services

• Identifying new customers

• Retention of Customers and Staff/Faculty
The Bioinformatics Core was created in CCMB (ca. 2009) and established as a Biomedical Research Core Facility by the Office of Research in 2012.

Our mission is to meet the growing bioinformatics analysis needs of researchers.

We help interpret complex, high-throughput biological data (DNA, RNA, Protein).

Bioinformatics: The science of analyzing and interpreting complex biological data.
Cost recovery

• In general, Cores at UM are supposed to be self-supporting through recharges back to researchers

• Bioinformatics Core’s major costs are salaries
  – software, cluster usage
  – Hardware - servers
Cost recovery - hurdles

- Developing methods and pipelines without a specific paying project.
- Accurately tracking time on custom projects
- Being able to charge enough to cover costs without breaking investigators banks.
Cost recovery – cont’d

- Service charges need to include the cost of developing the pipelines / automation for repeatable tasks.
- Need to recover the time for “learning” a new method, not just the time to run the method.
- Accurate estimation of time needed is difficult to obtain.
- Project management effort too!
Current Trend in Growth

Growth of Core Utilization

- Revenue was $221,558. Goal was 265K – pretty close!!
- # Projects and Investigators are on the right trajectory
- Changes in revenue from recharge rate restructuring and from increase in projects too

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Departments</td>
<td>12</td>
<td>27</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>Investigators</td>
<td>13</td>
<td>20</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td>Projects</td>
<td>16</td>
<td>25</td>
<td>55</td>
<td>109</td>
</tr>
</tbody>
</table>
Designing Services

- Identifying repeatable tasks
  - A process/method that many will use
- Quantitating time / effort into a per sample cost
  - Resistance and questioning of “hourly” charges
  - Incorporating effort to develop the repeatable pipeline
- Estimating time/effort for novel / custom projects
  - Custom projects have differential risk, and are assigned to bins of size/risk/effort
Expanding Services

- New methods come out (Tophat, STAR, etc)
- New Technology / platform (PacBio, minIon?)
- New utilization / request from researchers (gene panels, Hi-C)

** Need to validate new methods to insure proper results are given and reproducible. **
When to expand?

• When is a new service warranted??
  – The genomics core is adding a platform
  – Multiple requests for the same type of analysis

• When do you turn a service into a pipeline?
  – Need to improve the time to deliver, reduce internal effort
  – Large increase in requests and samples
How to Expand Customer Base?

- **Roadshows !!**
  - Present at departmental faculty meetings
  - Describe the core, services
  - Give examples of projects we’ve worked on

- **Recruit / Identify External Customers**
  - Other colleges and universities
  - Biotech / Commercial

- *Create a regional resource ?*
We help you throughout your research lifecycle

**Project Planning**
- Grant Support (Letters)
- Education and Training

**Experiment Design**
- Sequencing Methods
- Data Requirements
- Sample Size
- Data Analysis Options

**Execution**
- Data Analysis
- Methods/Software Development
- Annotation, Integration, Visualization
- Results Interpretation

**Publication**
- Validation Support
- Post-review Follow-up Analysis

---

Roadshow example slides
Our Standard Services

**Types**
- Whole Genome Seq (WGS)
- Exome Seq (WES)
- Amplicon
- Alignment, Variant identification
- Nucleotide polymorphisms, Copy number and structural variations

**Methods**
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**Results**
- DNA Methylation and Protein Binding
- Detect regulatory elements and chromatin structure
- Transcription factor binding, Histone modification, Methylation state
- Draft genome, Contigs, Functional annotation

**Genome Assembly**
- De-novo sequencing
- Metagenomics
- De-novo assembly, Homology mapping, gene prediction
- Draft genome, Contigs, Functional annotation

**RNA Expression**
- Analyze gene expression and alternative splicing
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**Genome Re-sequencing**
- Identify disease associated variations in DNA
- Whole Genome Seq (WGS)
- Exome Seq (WES)
- Amplicon
- Alignment, Variant identification
- Nucleotide polymorphisms, Copy number and structural variations

**DNA Methylation and Protein Binding**
- Detect regulatory elements and chromatin structure
- Transcription factor binding, Histone modification, Methylation state

**RNA Expression**
- Analyze gene expression and alternative splicing
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**Genome Assembly**
- Explore genomes of un-sequenced organisms
- De-novo sequencing
- Metagenomics
- De-novo assembly, Homology mapping, gene prediction
- Draft genome, Contigs, Functional annotation

**Genome Re-sequencing**
- Identify disease associated variations in DNA
- Whole Genome Seq (WGS)
- Exome Seq (WES)
- Amplicon
- Alignment, Variant identification
- Nucleotide polymorphisms, Copy number and structural variations

**RNA Expression**
- Analyze gene expression and alternative splicing
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**DNA Methylation and Protein Binding**
- Detect regulatory elements and chromatin structure
- Transcription factor binding, Histone modification, Methylation state

**Genome Assembly**
- Explore genomes of un-sequenced organisms
- De-novo sequencing
- Metagenomics
- De-novo assembly, Homology mapping, gene prediction
- Draft genome, Contigs, Functional annotation

**Results**
- DNA Methylation and Protein Binding
- Detect regulatory elements and chromatin structure
- Transcription factor binding, Histone modification, Methylation state

**RNA Expression**
- Analyze gene expression and alternative splicing
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**Genome Assembly**
- Explore genomes of un-sequenced organisms
- De-novo sequencing
- Metagenomics
- De-novo assembly, Homology mapping, gene prediction
- Draft genome, Contigs, Functional annotation

**Genome Re-sequencing**
- Identify disease associated variations in DNA
- Whole Genome Seq (WGS)
- Exome Seq (WES)
- Amplicon
- Alignment, Variant identification
- Nucleotide polymorphisms, Copy number and structural variations

**RNA Expression**
- Analyze gene expression and alternative splicing
- RNA-Seq
- Microarray
- Alignment, Transcript assembly
- Differential expression
- Differential gene expression, Alternative Splicing, Novel transcripts
- Alignment, Peak/signal detection
- Transcription factor binding, Histone modification, Methylation state

**DNA Methylation and Protein Binding**
- Detect regulatory elements and chromatin structure
- Transcription factor binding, Histone modification, Methylation state

**Genome Assembly**
- Explore genomes of un-sequenced organisms
- De-novo sequencing
- Metagenomics
- De-novo assembly, Homology mapping, gene prediction
- Draft genome, Contigs, Functional annotation
Custom Support for your Research

When you need something beyond an established support service …

*Come talk with us!*

---

**Research Support**

- Develop new methods (amplicon seq; pangenome assembly)
- Implement novel analysis workflows (RNA-DNA Editing)
- Evaluate and integrate third-party tools and data sources (Test a new aligners or variant callers)
- Specific analysis for publication figures

**New algorithms and software**

- Data integration and visualization (e.g. Jacquard, Winnow)
- Database design and implementation (DB for RNA-seq metadata and results; Kretzler)
Genomic Drivers in Leukemia and Lymphoma (Sami Malek)

2012 - 2014: 6 projects (4 Exome-Seq, 1 Gene Panel, 1 Custom), 152 samples

Somatic Variant Detection

Exome-Seq
FL, CLL

Prevalence of targeted variants in patient populations

Cross-platform Comparison (Exome vs. Microarray)

Targeted Gene Panel

10 genes
60 FL samples

Custom Analysis

Publication Support

Analysis Request

• How well do Exome SNP calls match up to SNP6.0 array calls?

Method

• Exome Sensitivity/Specificity analysis against SNP6.0

Result

• Exome SNP calls had 90% sensitivity and 99% specificity

• Novel coding variants identified

• Validated with Sanger sequencing

Gene-wise listing of variants across 60 samples with functional impact

Functional Impact of novel coding variants (SNPS, Indels)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Pos</th>
<th>Ref/Alt</th>
<th>Var Type</th>
<th>Effect</th>
<th>Class</th>
<th>AA</th>
<th>GENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>chr11</td>
<td>A/G</td>
<td>SNP</td>
<td>STOP</td>
<td>NONSENSE</td>
<td>E104*</td>
<td>Gene A</td>
</tr>
<tr>
<td>Sample 1</td>
<td>chr11</td>
<td>C/T</td>
<td>SNP</td>
<td>NON</td>
<td>NONSYNONMOUS</td>
<td>N3173D</td>
<td>Gene A</td>
</tr>
<tr>
<td>Sample 2</td>
<td>chr10</td>
<td>J/T</td>
<td>INDEL</td>
<td>NON</td>
<td>NONSYNONMOUS</td>
<td>D361H</td>
<td>Gene B</td>
</tr>
<tr>
<td>Sample 2</td>
<td>chr10</td>
<td>T/T</td>
<td>DEL</td>
<td>NON</td>
<td>NONSYNONMOUS</td>
<td>E306K</td>
<td>Gene B</td>
</tr>
</tbody>
</table>

Novel coding variants identified

Validated with Sanger sequencing
How to maintain growth and quality?

• Core service offerings need to be managed effectively to insure:
  – Uniformity of service (cost, time)
  – Quality and reproducibility
  – Resource allocation
  – Customer satisfaction

*Project & Portfolio Management*
Customer Retention

• Personal service
• Informed data handoff; Quality of data
• Clarity of deliverables and costs
• Responsiveness to (reasonable) requests
• Discussing the science before and after the experiment and analysis (Collaboration!)
Staff/faculty Retention

• Avoiding boredom and burnout
  – New Challenges; continuous learning

• Building teamwork and cross-training.
  – Maintain a ‘research’ experience where the core can
drive the science

• Competing with Industry salaries?
  – University vs. Commercial – tough to do
  – Make sure the environment is so good they won’t
    leave!
Summary

• Cost Recovery improved by good time tracking, appropriate rate structures; increasing throughput
• Improving and expanding services is a constant challenge
• Outreach is essential for new customer ID
• Collaboration and listening to existing customers improves retention
• Keeping your staff/faculty engaged is important to long-term consistency
Thank you!

The UM Bioinformatics Core is supported by the Office of Research, Medical School Administration and the Biomedical Research Core Facilities at the University of Michigan School of Medicine.