Sample Quality Control:
Qualifying Renewable Biological Resources

Dr. Andrew Brooks

University of Medicine and Dentistry of New Jersey
Rutgers University
Director, Bionomics Research and Technology Center
Director, Technology Development and Implementation
Rutgers University Cell and DNA Repository
Environmental and Occupational Health Science Institute
Human Genetics Institute of New Jersey
RUCDR enables sharing programs (DNA, RNA, cell lines, tissue and clinical data) for NIH Institutes, research advocacy groups & biotechnology corporations

- Speeding discovery of genes for complex diseases by sharing well annotated, high quality human samples

- >$30M annual grant & contract support
SELECTED RUCDR PROJECTS

- **NIDDK**
  - Diabetes Type I and Type II (also HBDI)
  - Inflammatory Bowel Disease
  - Kidney and Liver Diseases

- **NIMH**
  - Alzheimer Disease
  - Autism (also CAN/AGRE)
  - Bipolar Disorder
  - Schizophrenia
  - Pharmacogenetic (clinical) trials

- **NIDA**
  - Tobacco
  - Opiates
  - Cocaine
  - Clinical trials

- **NIAAA / COGA**
  - Alcoholism
    - Simons Simplex Collection
    - Autism
    - Immune Tolerance Network
5 Major Program Functions

- Sample acquisition
- Processing
- Storage
- Distribution
- Analysis
Functional Essentials: Maximizing Biological Resources

- **Maximal use of primary samples**
  - Undefined application for downstream analyses

- **Efficient processing**
  - Maximizing extraction technologies to improve yield and quality

- **Appropriate storage**
  - Defining storage formats and temperatures to maximize storage infrastructure

- **Nucleic acid amplification / Cell line establishment**
  - Creating renewable resources to preserve primary sample and/or precious collections

- **Appropriate distribution guidelines**
  - Define needs for specific downstream applications to preserve sample resources
Repository Management Operations

- Standards & Policies
- Controls
- Equipment & Supplies
- Infrastructure
- Scalable & Dedicated Resources
- Liability / Business Continuity
- Processes
- Technology
- Sample Management Operations
- Project Oversight
- Data Integrity & Management
- Compliance & Quality
- Logistics
- Performance

Courtesy of BST 2009
What is the difference between a “Stored” sample and a “Biobank “Sample? 

- A banked sample is proactively acquired for future testing or analysis 
- A banked sample is often sent to multiple different recipients

**A BANKED SAMPLE SHOULD NEVER BE DEPLETED**
Some Challenges for Genetics Repositories

○ Most DNA are genotyped (e.g., SNPs) soon after collection and provided to several labs who may compare data.

  ➢ Errors are revealed quickly!

○ Samples must be of high quality and uniform concentration

  ➢ Requirement of high throughput assays

○ Must accommodate up to a 5-fold daily variation in number of samples received (labor, space and supplies issues)

MUST BE ABLE TO MANAGE BANDWIDTH!
Sources of errors...

- Sample identity errors are often revealed by lack of Mendelian relationship between samples.
  - Non-paternity, non-maternity (adopted)
  - Mislabeling in field (most common error)
    - Mixing samples from two individuals (especially common when collecting family samples at the same time)
- Repository errors
  - QA procedures and sample tracking systems allow historic dissection of mislabeling errors (which can then be corrected)
    - Photographing blood tubes/ saving blood sample
    - No manual transcription
    - Capture data on all processing and QA/QC steps
Application “Independent” Workflows

- Sample Pre-Registration
  - Hundreds of sites globally
- Sample Accessioning
  - Scalable and qualitative
- Sample Validation
  - Process initiation
- Processing / QC
  - Analytical and Functional Measurements
- Sample Storage
  - Variable temperatures and formats
- Sample Distribution
  - Custom requests and sample management
Workflow Analysis

Sample → DNA/RNA Extraction → Nucleic Acid Amplification

Sample Archival → Sample Distribution

Qualitative & Quantitative Analyses → Comprehensive Tracking
Integrated QC Processes

- **Sample Quality Control**
  - DNA – Spectroscopy, RUCDR ID™ SNP Profiling
  - RNA – Spectroscopy, Bioanalyzer, cDNA fidelity testing (QPCR-ICED)

- **Nucleic Acid Amplification for Expression**
  - Currently 50% of all RNA samples are “amplified”
  - 2010 projection for 100% pre-amplification of all expression studies

- **Archival**
  - Rigorous storage requirements for all nucleic acid samples
  - Renewable resource for investigators expanding the discovery/screening process
Process Redefined...

- Qualitative
- Quantitative
- Functional
Quality Control / Quality Assurance

- **Process Quality Control**
  - Sample Collection
  - Sample Processing

- **Storage Quality Control**
  - Storage Format
  - Temperature

- **Analytical Quality Control**
  - Volume, Concentration, Fidelity

- **Functional Quality Control**
  - Application specific analysis

- **Distribution Quality Control**
Analytical Quality Control

- **Volume**
  - Non-contact vs. contact
  - How accurate do measurements need to be?
  - How do you define a “fudge factor” for lost volume during sampling

- **Concentration**
  - How “homogeneous” is the sample you are measuring
  - When is the right time to sample for measurement?
  - Where do you sample from?
  - What technologies are available?
Purity / Fidelity
- What are the right measurements to record?
- How are purity metrics determined, empirically?
- Sample “clean up” quality control
- Establishing acceptable criteria for downstream applications

Weight
- An alternative to volume measurements
- “To tare or not to tare”
Annotation

- Consistency for sample annotation is key
- Samples can be defined by their quality control metrics
- Make sure sample QC encompasses “industry standards” that are often sample type specific

Sample Retesting

- When does analytical analysis need to be repeated (if ever)?
- If retests are run, what do you do with historical data?
Functional Quality Control

- **DNA (gDNA, WGA, Free floating DNA)**
  - More downstream applications then ever before in this field
  - Importance of high molecular weight DNA vs. low molecular weight DNA
  - Choose application(s) that have the most correlative value for analysis

- **RNA (Total RNA, mRNA, miRNA)**
  - Sample quality is of paramount importance!
  - Fidelity doesn’t necessarily ensure reproducibility
Functional Quality Control II

- **Protein** (lystaes, serum, plasma)
  - Qualitative vs. quantitative analysis
  - Defining stability measurements
  - How many end point measurements is enough?

- **Tissue** (fresh, fixed, post-mortem)
  - Pathology verification
  - Verification of storage formats
  - Molecular vs. Histological Analyses
Functional Quality Control III

○ Functional Analysis Over Time
  ● Is it your responsibility to monitor potentially labile samples over time?
  ● What are the appropriate intervals for testing?
  ● How is change in sample quality reported?
  ● When new downstream applications arise is additional functional testing required?
## RUCDR ID™ 96 SNP Panel

<table>
<thead>
<tr>
<th>RUID</th>
<th>dbSNP</th>
<th>Category</th>
<th>RUID</th>
<th>dbSNP</th>
<th>Category</th>
<th>RUID</th>
<th>dbSNP</th>
<th>Category</th>
<th>RUID</th>
<th>dbSNP</th>
<th>Category</th>
<th>RUID</th>
<th>dbSNP</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>hu1</td>
<td>rs1471939</td>
<td>I</td>
<td>hu29</td>
<td>rs4746136</td>
<td>A,I</td>
<td>hu54</td>
<td>rs9319336</td>
<td>I</td>
<td>hu79</td>
<td>Rs1336071</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu2</td>
<td>rs4666200</td>
<td>I</td>
<td>hu30</td>
<td>rs4821004</td>
<td>A,I</td>
<td>hu55</td>
<td>rs1019029</td>
<td>I</td>
<td>hu80</td>
<td>Rs740598</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu3</td>
<td>rs7554936</td>
<td>I</td>
<td>hu31</td>
<td>rs13218440</td>
<td>I</td>
<td>hu56</td>
<td>rs1358856</td>
<td>P</td>
<td>hu81</td>
<td>Rs12997453</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu4</td>
<td>rs9530435</td>
<td>A,I</td>
<td>hu32</td>
<td>rs1523537</td>
<td>I</td>
<td>hu57</td>
<td>rs279844</td>
<td>P</td>
<td>hu82</td>
<td>Rs2352476</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu5</td>
<td>rs6104567</td>
<td>I</td>
<td>hu33</td>
<td>rs1058083</td>
<td>I</td>
<td>hu58</td>
<td>rs1623718</td>
<td>I</td>
<td>hu83</td>
<td>Rs1554472</td>
<td>A,I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu7</td>
<td>rs2272998</td>
<td>P</td>
<td>hu34</td>
<td>rs1344870</td>
<td>P</td>
<td>hu59</td>
<td>rs2503107</td>
<td>I</td>
<td>hu84</td>
<td>Rs10007810</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu9</td>
<td>rs560681</td>
<td>A,I</td>
<td>hu35</td>
<td>rs7704770</td>
<td>I</td>
<td>hu61</td>
<td>rs10236187</td>
<td>A,I</td>
<td>hu85</td>
<td>rs1760921</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu11</td>
<td>rs5691147</td>
<td>A,I</td>
<td>hu36</td>
<td>rs1410059</td>
<td>A,I</td>
<td>hu62</td>
<td>rs1513181</td>
<td>I</td>
<td>hu86</td>
<td>Rs1040045</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu12</td>
<td>rs321198</td>
<td>P</td>
<td>hu37</td>
<td>rs5768007</td>
<td>I</td>
<td>hu63</td>
<td>rs7657799</td>
<td>A,I</td>
<td>hu87</td>
<td>Rs10496971</td>
<td>A,I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu13</td>
<td>rs3784230</td>
<td>I</td>
<td>hu38</td>
<td>rs260690</td>
<td>A,I</td>
<td>hu64</td>
<td>rs2504853</td>
<td>A,I</td>
<td>hu88</td>
<td>Rs7803075</td>
<td>A,I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu14</td>
<td>rs870347</td>
<td>I</td>
<td>hu39</td>
<td>rs13400937</td>
<td>I</td>
<td>hu65</td>
<td>rs772262</td>
<td>I</td>
<td>hu89</td>
<td>Rs987640</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu15</td>
<td>rs2946788</td>
<td>I</td>
<td>hu41</td>
<td>rs4918842</td>
<td>A,I</td>
<td>hu66</td>
<td>rs3737576</td>
<td>I</td>
<td>hu90</td>
<td>Rs6444724</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu16</td>
<td>rs4891825</td>
<td>I</td>
<td>hu42</td>
<td>rs9809104</td>
<td>I</td>
<td>hu67</td>
<td>rs7520386</td>
<td>P</td>
<td>hu91</td>
<td>Rs10092491</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu17</td>
<td>rs10108270</td>
<td>A,I</td>
<td>hu43</td>
<td>rs2073383</td>
<td>P</td>
<td>hu68</td>
<td>rs445251</td>
<td>P</td>
<td>hu92</td>
<td>Rs735612</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu18</td>
<td>rs2397060</td>
<td>A,I</td>
<td>hu44</td>
<td>rs1821380</td>
<td>P</td>
<td>hu69</td>
<td>rs10468710</td>
<td>A</td>
<td>hu93</td>
<td>Rs985492</td>
<td>A,I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu20</td>
<td>rs7229946</td>
<td>A</td>
<td>hu45</td>
<td>rs279844</td>
<td>P</td>
<td>hu70</td>
<td>rs722969</td>
<td>P</td>
<td>hu94</td>
<td>Rs338882</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu21</td>
<td>rs13182883</td>
<td>A,I</td>
<td>hu46</td>
<td>rs952718</td>
<td>P</td>
<td>hu71</td>
<td>rs1109037</td>
<td>A</td>
<td>hu95</td>
<td>Rs9511711</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu22</td>
<td>rs1876482</td>
<td>P</td>
<td>hu47</td>
<td>rs447818</td>
<td>P</td>
<td>hu72</td>
<td>rs3780962</td>
<td>I</td>
<td>hu96</td>
<td>Rs3907047</td>
<td>A,I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu23</td>
<td>rs315791</td>
<td>A,I</td>
<td>hu48</td>
<td>rs13134862</td>
<td>P</td>
<td>hu73</td>
<td>rs7997709</td>
<td>I</td>
<td>hu98</td>
<td>Rs1865580</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu24</td>
<td>rs7205345</td>
<td>P</td>
<td>hu49</td>
<td>rs4463276</td>
<td>I</td>
<td>hu74</td>
<td>rs4670767</td>
<td>I</td>
<td>hu103X</td>
<td>Rs525869</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu25</td>
<td>rs786443</td>
<td>A,I</td>
<td>hu50</td>
<td>rs6845457</td>
<td>I</td>
<td>hu75</td>
<td>rs6522149</td>
<td>I</td>
<td>hu106Y</td>
<td>Rs2058276</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu26</td>
<td>rs4717865</td>
<td>A,I</td>
<td>hu51</td>
<td>rs3943253</td>
<td>I</td>
<td>hu76</td>
<td>rs4998343</td>
<td>A,I</td>
<td>hu107X</td>
<td>Rs2040962</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu27</td>
<td>rs2416791</td>
<td>A,I</td>
<td>hu52</td>
<td>rs6548616</td>
<td>A,I</td>
<td>hu77</td>
<td>rs6451722</td>
<td>I</td>
<td>hu109X</td>
<td>Rs530501</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu28</td>
<td>rs2125345</td>
<td>A,I</td>
<td>hu53</td>
<td>rs731257</td>
<td>A,I</td>
<td>hu78</td>
<td>rs12629908</td>
<td>I</td>
<td>hu111Y</td>
<td>Rs2032624</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
High-throughput Allelic Discrimination

96.96 Array

Fluidigm

Polymorphic

Gender

SNP: rs1, Call Rate: 100.00%, Confidence: 99.51

SNP: rs1, Call Rate: 100.00%, Confidence: 99.51

SNP: rs1, Call Rate: 100.00%, Confidence: 99.51

SNP: rs1, Call Rate: 100.00%, Confidence: 99.51
Data → Analysis → Interpretation

or

STOP
RUCDR ID™ Data Resource

- Millions of data points collected
- 10K+ samples/month
- Rapidly determine sample contamination/processing errors
- Proactively address sample registration errors
- Catalogue all RUCDR DNA samples continuously

RUCDR DNA QC Database
Important Metrics for RNA Quality Control

- Ribosomal RNA as a surrogate for mRNA
- When is QC most appropriate
- What is the best measure of RNA quality as a function of gene expression measurements?
- Is RNA the best biorepository source for expression studies?
  - Amplified cDNA for distribution (NuGEN Inc.)
Metrics for RNA Quality:
Sample Stability

% Sample Integrity

<table>
<thead>
<tr>
<th>Time</th>
<th>DNA</th>
<th>RNA</th>
<th>cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Month</td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>6 Month</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>1 Year</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>3 Year</td>
<td>92</td>
<td>92</td>
<td>92</td>
</tr>
</tbody>
</table>
Metrics for RNA Quality:
Degradation as a Function of Use

% Sample Integrity

Thaw Number

DNA
RNA
cDNA
Metrics for RNA Quality: Temporal and Technical Variation

QPCR

% Concordance

Lab 1  Lab 2  Lab 3  Lab 4  Lab 5

DNA  RNA  cDNA
Quality Control – RNA Integrity

- QPCR is an ESSENTIAL component of RNA quality assessment for all gene expression studies.
- The Bioanalyzer is a good “gross” measure of RNA integrity only.
- Biological “specific” QPCR approach provides more useful and functional information and can be used to correlate sample performance.
Raising the Bar for Sample Validation: A Standard Approach

cDNA yield = 8.5 ug

Genes Present = 49%
B-actin 3’/5’ ratio = 9.6
GAPDH 3’/5’ ratio = 1.6
Background = 53

Genes Present = 51%
B-actin 3’/5’ ratio = 5.4
GAPDH 3’/5’ ratio = 1.8
Background = 52
Reasons for Concern: Biology or Technology?

Ratio Distribution

Signal Distribution

116 197

Ratio Distribution

Signal Distribution
cDNA distribution (focused)=
Sample Quality Assessment

cDNA yield = 8.5 ug
Gene 1a (3’) Pass
Gene 1b (M) Pass
Gene 1c (5’) Pass

Gene 2a (3’) Pass
Gene 2b (M) Pass
Gene 2c (5’) Pass

Gene 3a (3’) Pass
Gene 3b (M) Pass
Gene 3c (5’) Pass

High Expresser
Pass
Pass

Medium Expresser
Pass
Pass/Fail

Low Expresser
Fail
Fail

Independent Consistent Expression Discriminator (ICED)

Step 1: Assigning Weights

\[ W_1(g) = \frac{1}{m} \sum_{i=1,m} |g_{2i} - \mu_{1,n}(g)| \]
\[ W_2(g) = \frac{1}{n} \sum_{j=1,n} |g_{1j} - \mu_{2,m}(g)| \]

Step 2: Assigning Votes

\[ V_1(g) = W_2(g) \cdot |g_x - \mu_{2,TR,m}(g)| \]
\[ V_2(g) = W_1(g) \cdot |g_x - \mu_{1,TR,n}(g)| \]

Step 3: Counting Votes

\[ P(x) = \frac{q \cdot \sum_{i=1,q} V_1(g_i) - p \cdot \sum_{i=1,p} V_2(g_i)}{q \cdot \sum_{i=1,q} V_1(g_i) + p \cdot \sum_{i=1,p} V_2(g_i)} \]
QPCR: cDNA is better indicator of sample quality for gene expression

- Utilizes standard chemistries for QPCR
- Multiple probes per gene (multiplexed)
- "Biological Representation" – CRITICAL
  - Neurobiology, Cardiovascular, Oncology
- Range of sensitivities - CRITICAL
- Correlation to reference database
- Sample "pass" or "fail"
  - Assessing sample quality on the fly...
    accounting for amplified product sizes
DNA/RNA Analysis: Downstream Applications

- **Non-QPCR**
  - Microarrays
  - NextGen Sequencing

- **QPCR/QFPCR**
  - Research applications
  - Diagnostic applications
  - High Throughput Technologies

Flexibility is ESSENTIAL
Sensitivity is CRITICAL
Quantitation is CRITICAL
Biobanking: Is there a real need?

- Are all samples limiting?
- Is having comparable data essential?
- Integrated Sample Quality Control

“Keep eating... we will make more”

“Keep analyzing... we will make more”
Applying Technological & Business Infrastructure to Complex Disease Research

http://www.RUCDR.org
brooks@biology.rutgers.edu