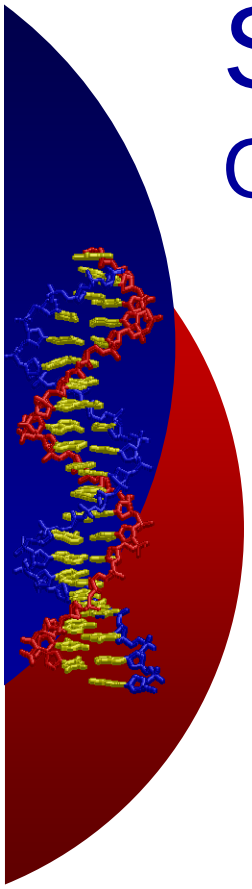


Sample Quality Control:
Qualifying Renewable Biological Resources

Dr. Andrew Brooks

University of Medicine and Dentistry of New Jersey

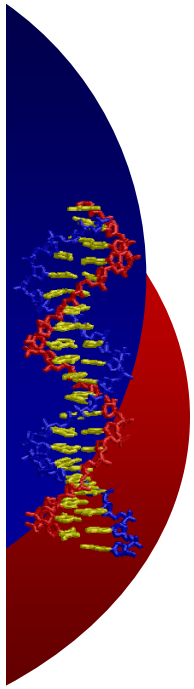
Rutgers University

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

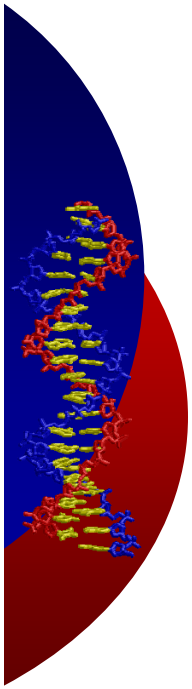


Mission

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

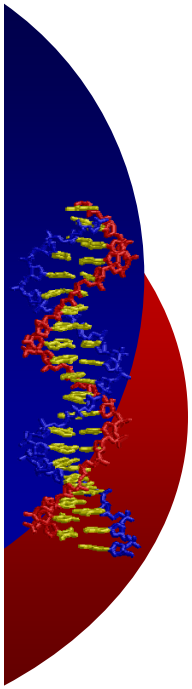
SELECTED RUCDR PROJECTS



- 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

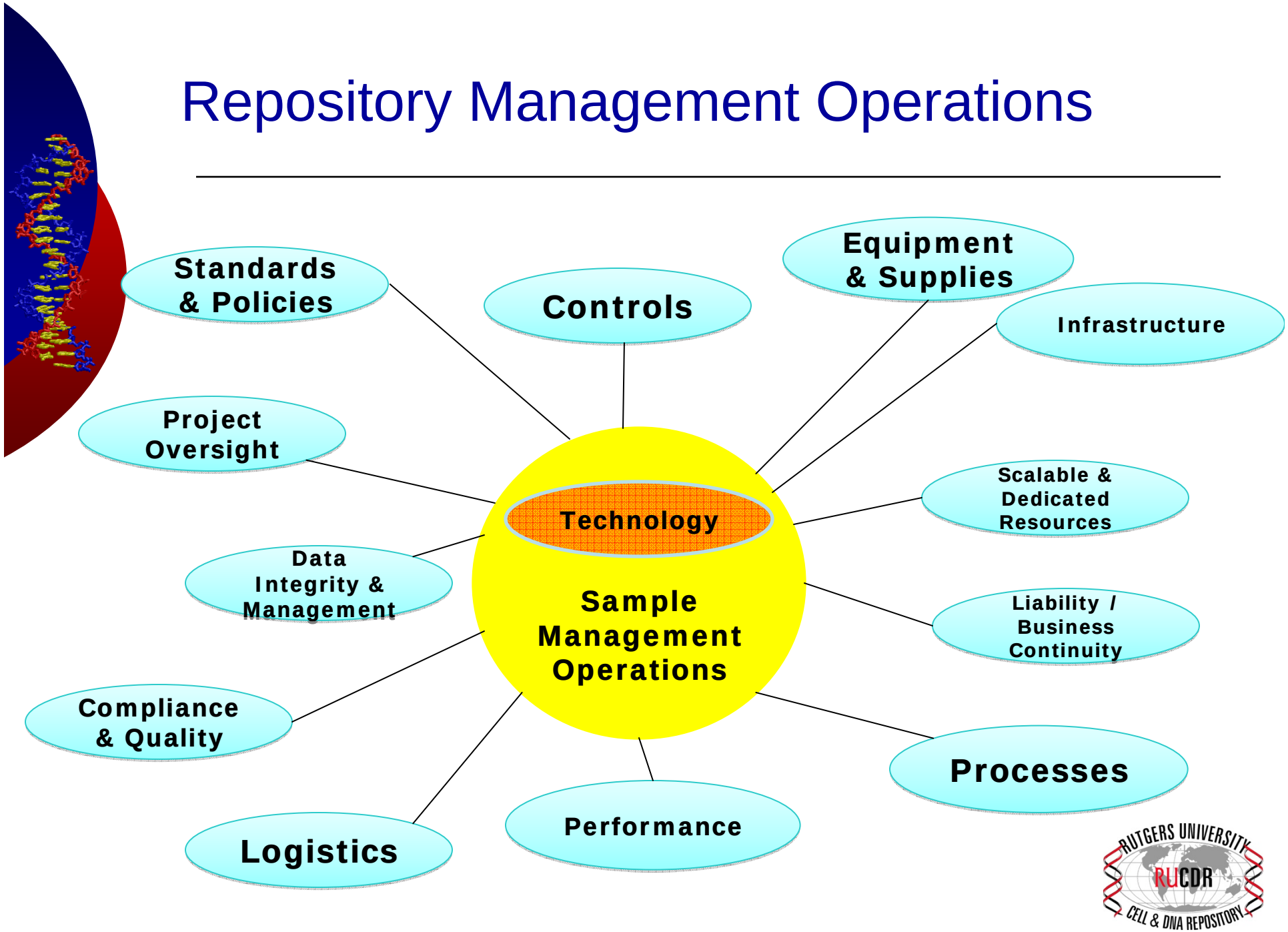
- Sample acquisition
- Processing
- Storage
- Distribution
- Analysis



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

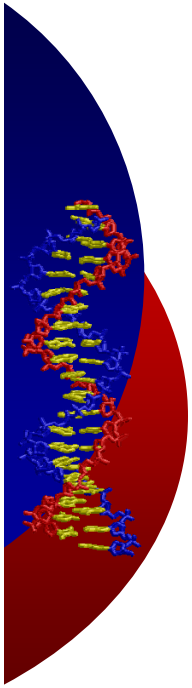


Courtesy of BST 2009

Biological Storage...Defined

What is the difference between a “Stored” sample and a “Biobank “Sample?

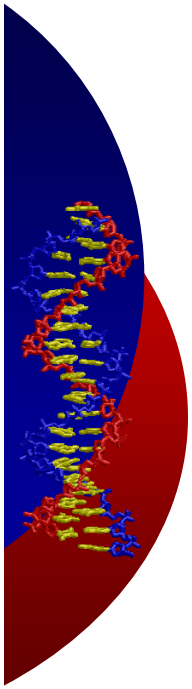
- A banked sample is proactively acquired for future testing or analysis
- A banked sample is often sent to multiple different recipients
- **A BANKED SAMPLE SHOULD NEVER BE DEPLETED**



Some Challenges for Genetics Repositories

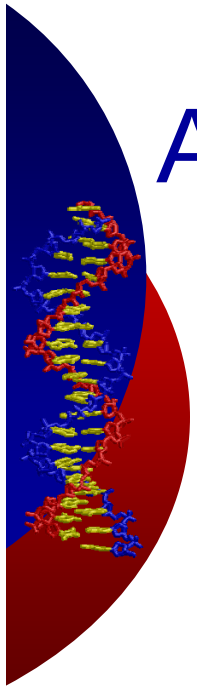
- Most DNA are genotyped (e.g., SNPs) soon after collection and provided to several labs who may compare data.
  - Errors are revealed quickly!
- Samples must be of high quality and uniform concentration
  - Requirement of high throughput assays
- Must accommodate up to a 5-fold daily variation in number of samples received (labor, space and supplies issues)

MUST BE ABLE TO MANAGE BANDWITH!



Sources of errors…

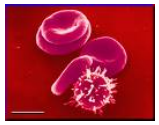
- Sample identity errors are often revealed by lack of Mendelian relationship between samples.
  - Non-paternity, non-maternity (adopted)
  - Mislabeling in field (most common error)
    - Mixing samples from two individuals (especially common when collecting family samples at the same time)
  - Repository 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



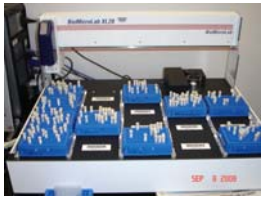
Sample

DNA / RNA Extraction

Nucleic Acid Amplification

STARLIMS

Sample Archival

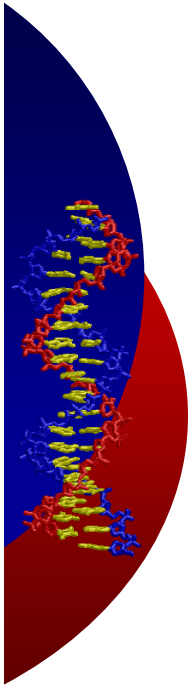


Sample Distribution

Qualitative & Quantitative Analyses

Comprehensive Tracking

RUTGERS UNIVERSITY RUCDR CELL & DNA REPOSITORY

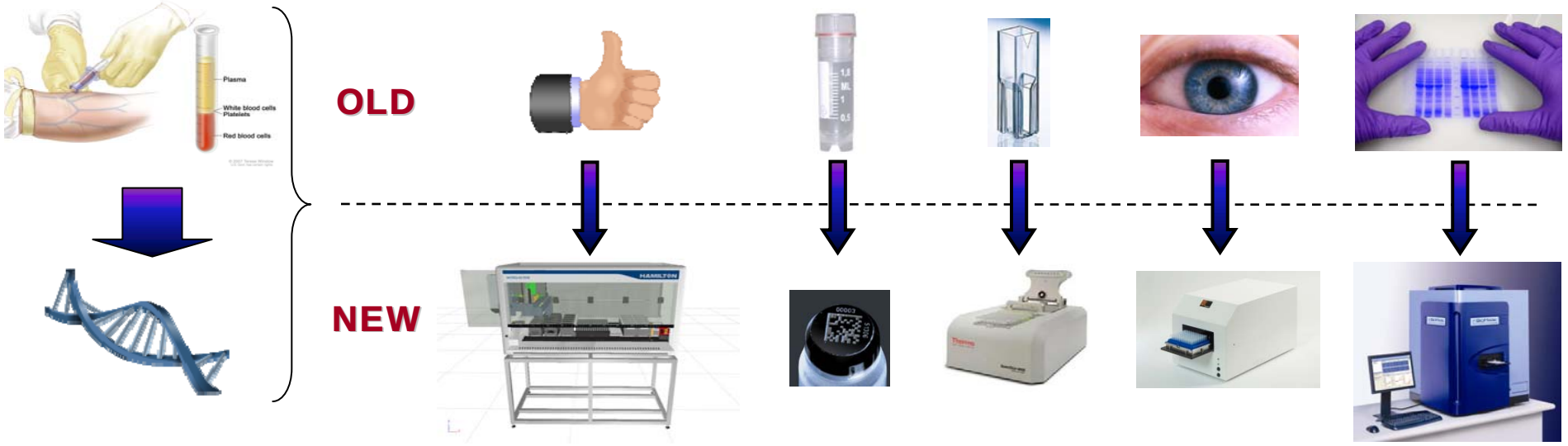


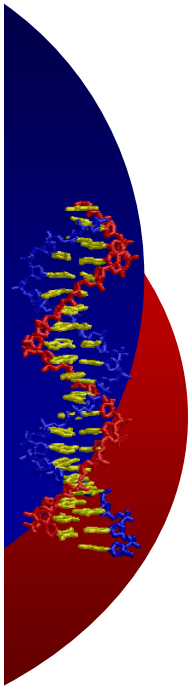
Integrated QC Processes

- 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…

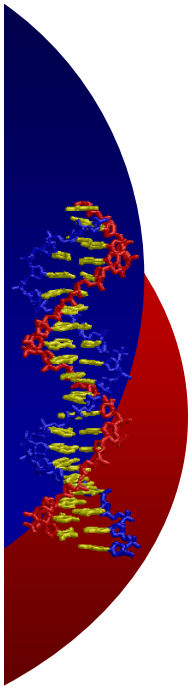
- **Qualitative**
- **Quantitative**
- **Functional**





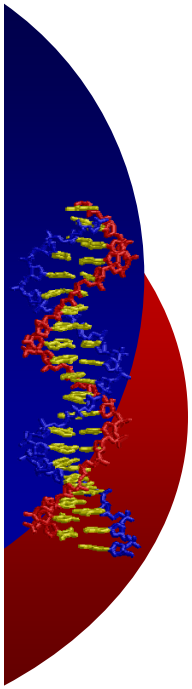
Quality Control / Quality Assurance

- 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



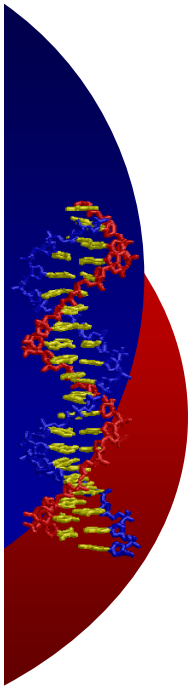
Analytical Quality Control

- 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?



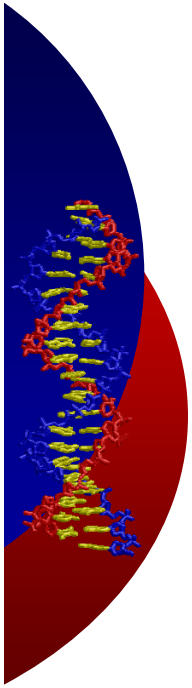
Analytical Quality Control II

- 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”



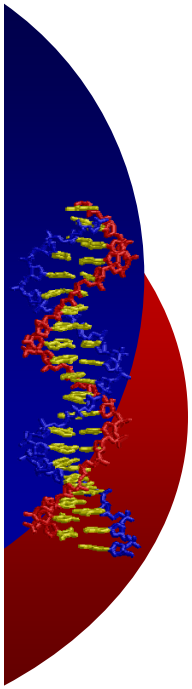
Analytical Quality Control III

- 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?



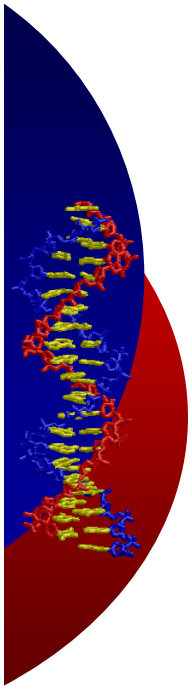
Functional Quality Control

- 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



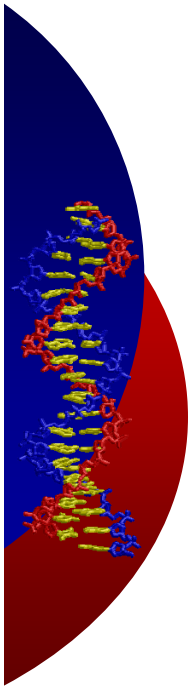
Functional Quality Control II

- 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?

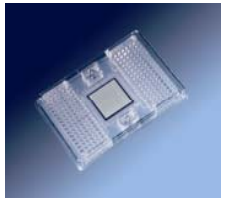


RUCDR ID™ 96 SNP Panel

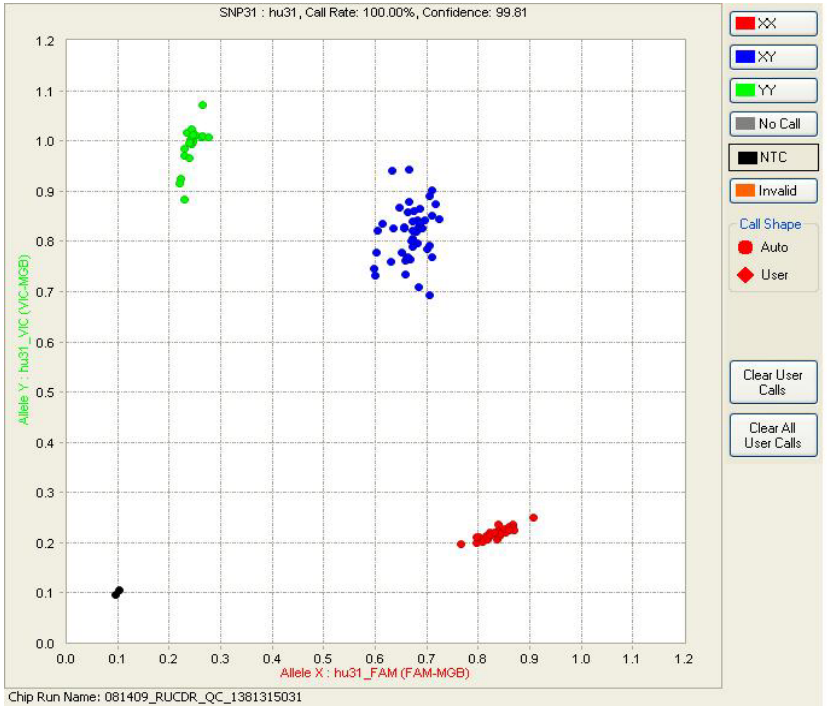
RUID	dbSNP	Category	RUID	dbSNP	Category	RUID	dbSNP	Category	RUID	dbSNP	Category
hu1	rs1471939	I	hu29	rs4746136	A,I	hu54	rs9319336	I	hu79	Rs1336071	A
hu2	rs4666200	I	hu30	rs4821004	A,I	hu55	rs1019029	I	hu80	Rs740598	P
hu3	rs7554936	I	hu31	rs13218440	I	hu56	rs1358856	P	hu81	Rs12997453	I
hu4	rs9530435	A,I	hu32	rs1523537	I	hu57	rs279844	P	hu82	Rs2352476	P
hu5	rs6104567	I	hu33	rs1058083	I	hu58	rs1823718	I	hu83	Rs1554472	A,I
hu7	rs2272998	P	hu34	rs1344870	P	hu59	rs2503107	I	hu84	rs10007810	I
hu9	rs560681	A,I	hu35	rs7704770	I	hu61	rs10236187	A,I	hu85	rs1760921	I
hu11	rs6591147	A,I	hu36	rs1410059	A,I	hu62	rs1513181	I	hu86	rs1040045	I
hu12	rs321198	P	hu37	rs5768007	I	hu63	rs7657799	A,I	hu87	rs10496971	A,I
hu13	rs3784230	I	hu38	rs260690	A,I	hu64	rs2504853	A,I	hu88	rs7803075	A,I
hu14	rs870347	I	hu39	rs13400937	I	hu65	rs772262	I	hu89	rs987640	P
hu15	rs2946788	I	hu41	rs4918842	A,I	hu66	rs3737576	I	hu90	rs6444724	I
hu16	rs4891825	I	hu42	rs9809104	I	hu67	rs7520386	P	hu91	rs10092491	I
hu17	rs10108270	A,I	hu43	rs2073383	P	hu68	rs445251	P	hu92	rs735612	A
hu18	rs2397060	A,I	hu44	rs1821380	P	hu69	rs10488710	A	hu93	rs985492	A,I
hu20	rs7229946	A	hu45	rs279844	P	hu70	rs722869	P	hu94	rs338882	I
hu21	rs13182883	A,I	hu46	rs952718	P	hu71	rs1109037	A	hu95	rs9951171	P
hu22	rs1876482	P	hu47	rs447818	P	hu72	rs3780962	I	hu96	rs3907047	A,I
hu23	rs315791	A,I	hu48	rs13134862	P	hu73	rs7997709	I	hu98Y	rs1865680	G
hu24	rs7205345	P	hu49	rs4463276	I	hu74	rs4670767	I	hu103X	rs525869	G
hu25	rs798443	A,I	hu50	rs9845457	I	hu75	rs9522149	I	hu106Y	rs2058276	G
hu26	rs4717865	A,I	hu51	rs3943253	I	hu76	rs4908343	A,I	hu107X	rs2040962	G
hu27	rs2416791	A,I	hu52	rs6548616	A,I	hu77	rs6451722	I	hu109X	rs530501	G
hu28	rs2125345	A,I	hu53	rs731257	A,I	hu78	rs12629908	I	hu111Y	rs2032624	G

High-throughput Allelic Discrimination

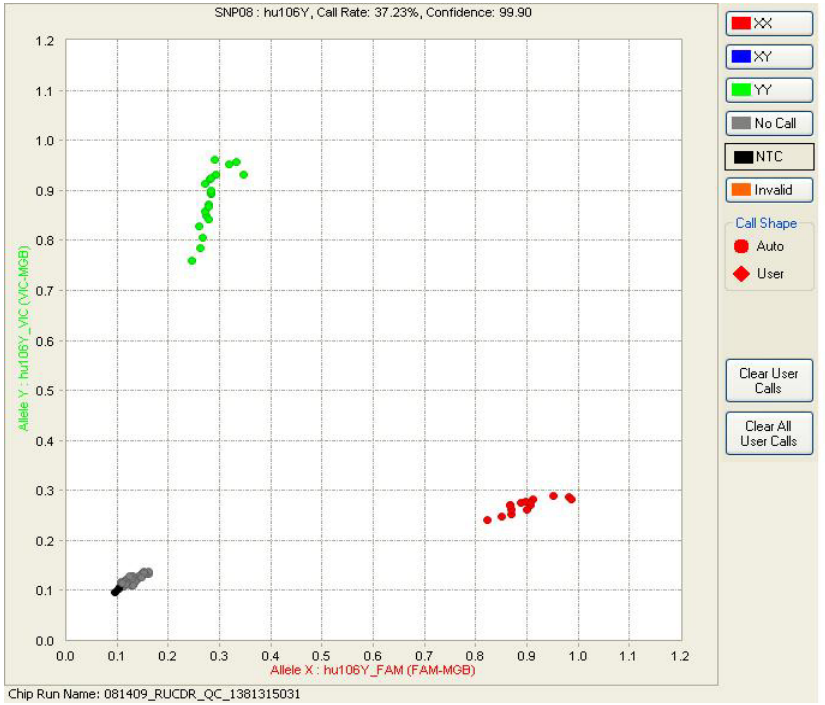
96.96 Array



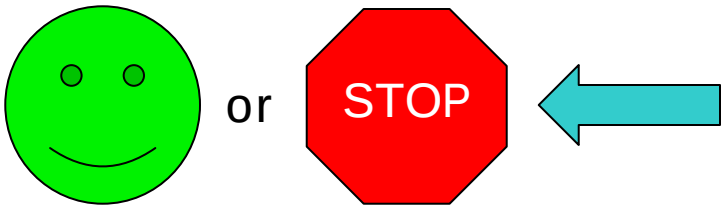
Polymorphic

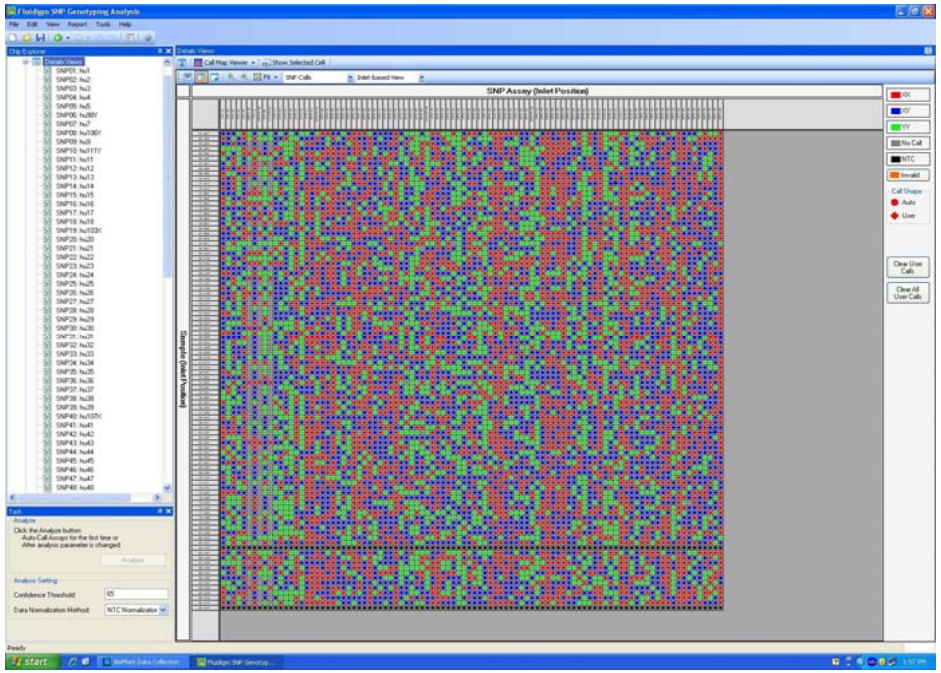


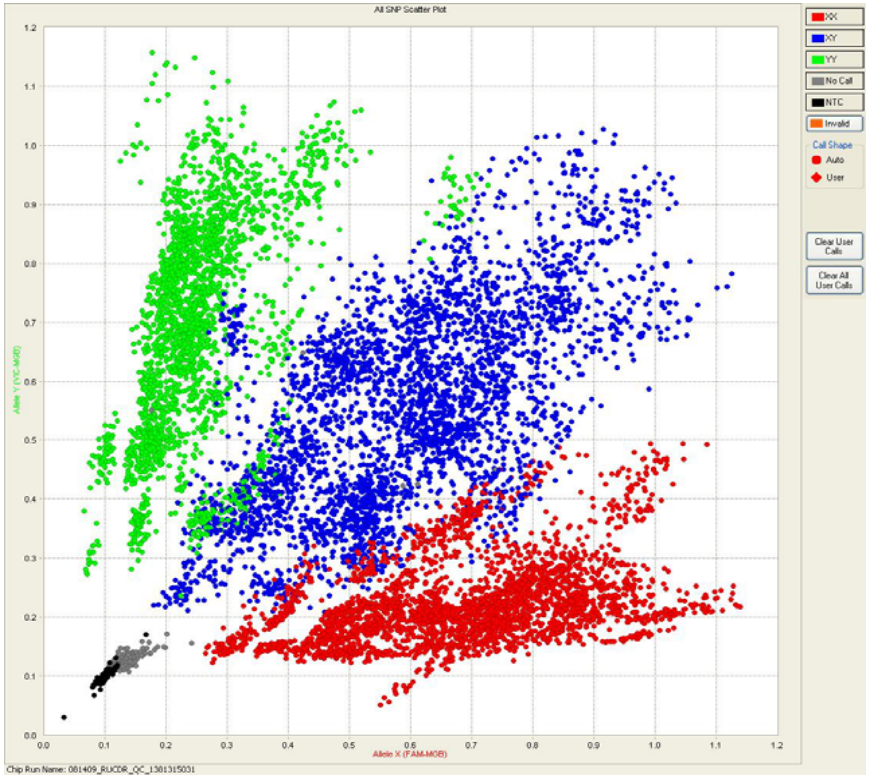
Gender

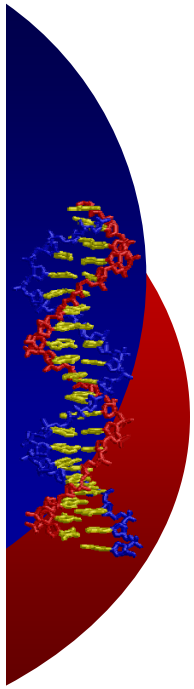


Data ⇨ Analysis ⇨ Interpretation





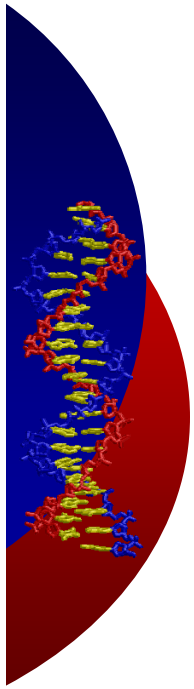




RUCDR ID™ Data Resource

- 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



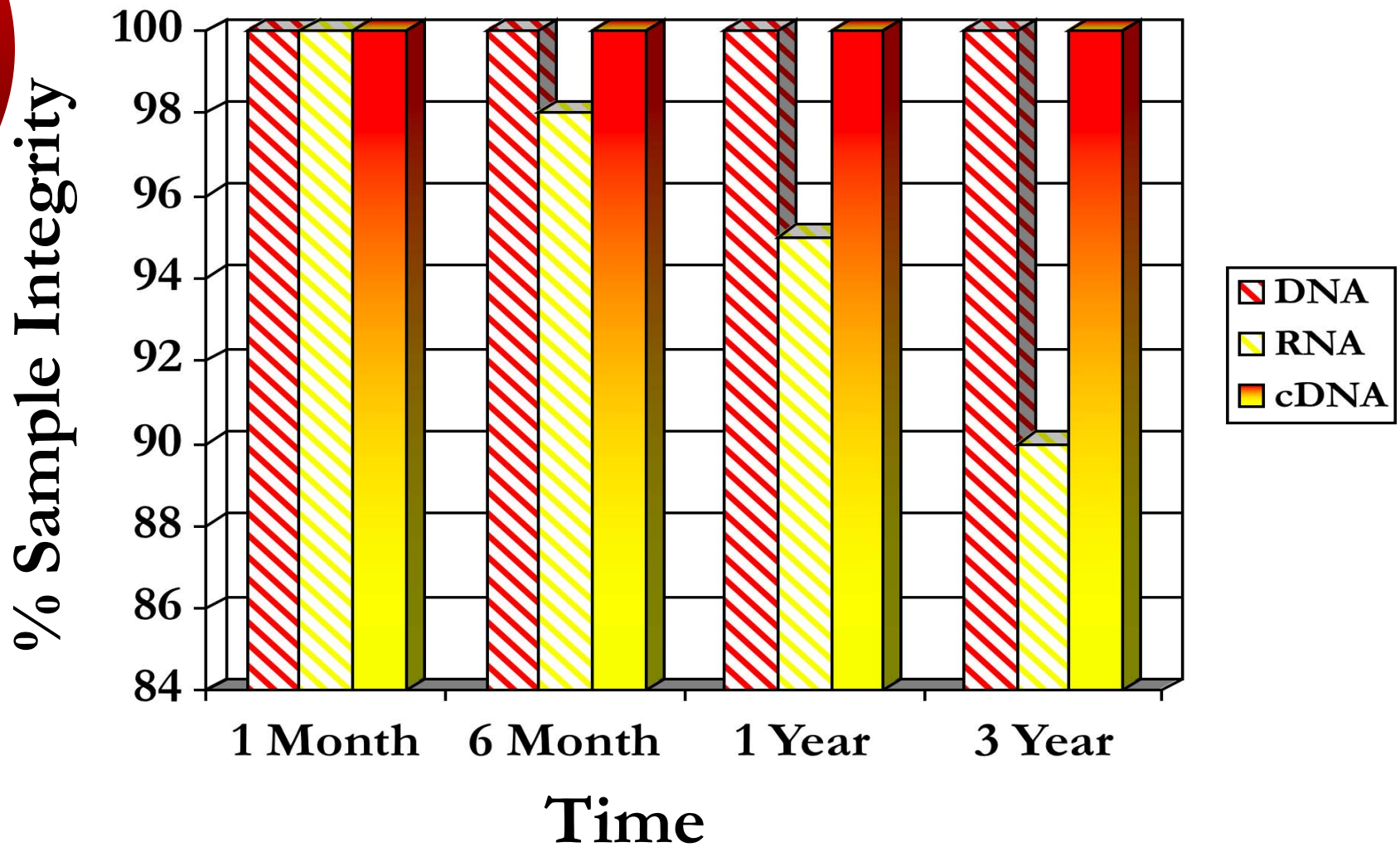
Important Metrics for RNA Quality Control

- 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?

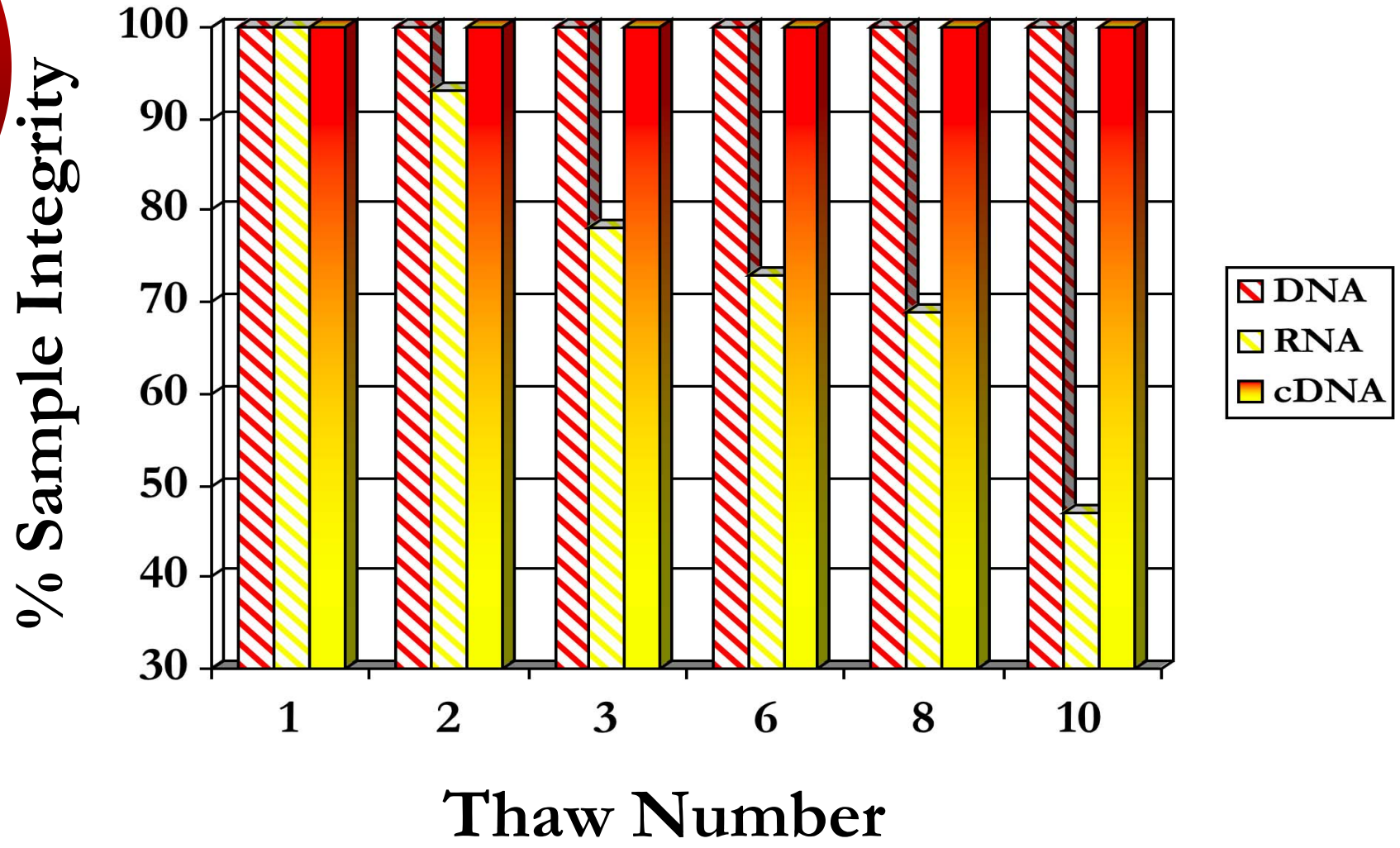
- Is RNA the best biorepository source for expression studies?
  - Amplified cDNA for distribution (NuGEN Inc.)

Metrics for RNA Quality:

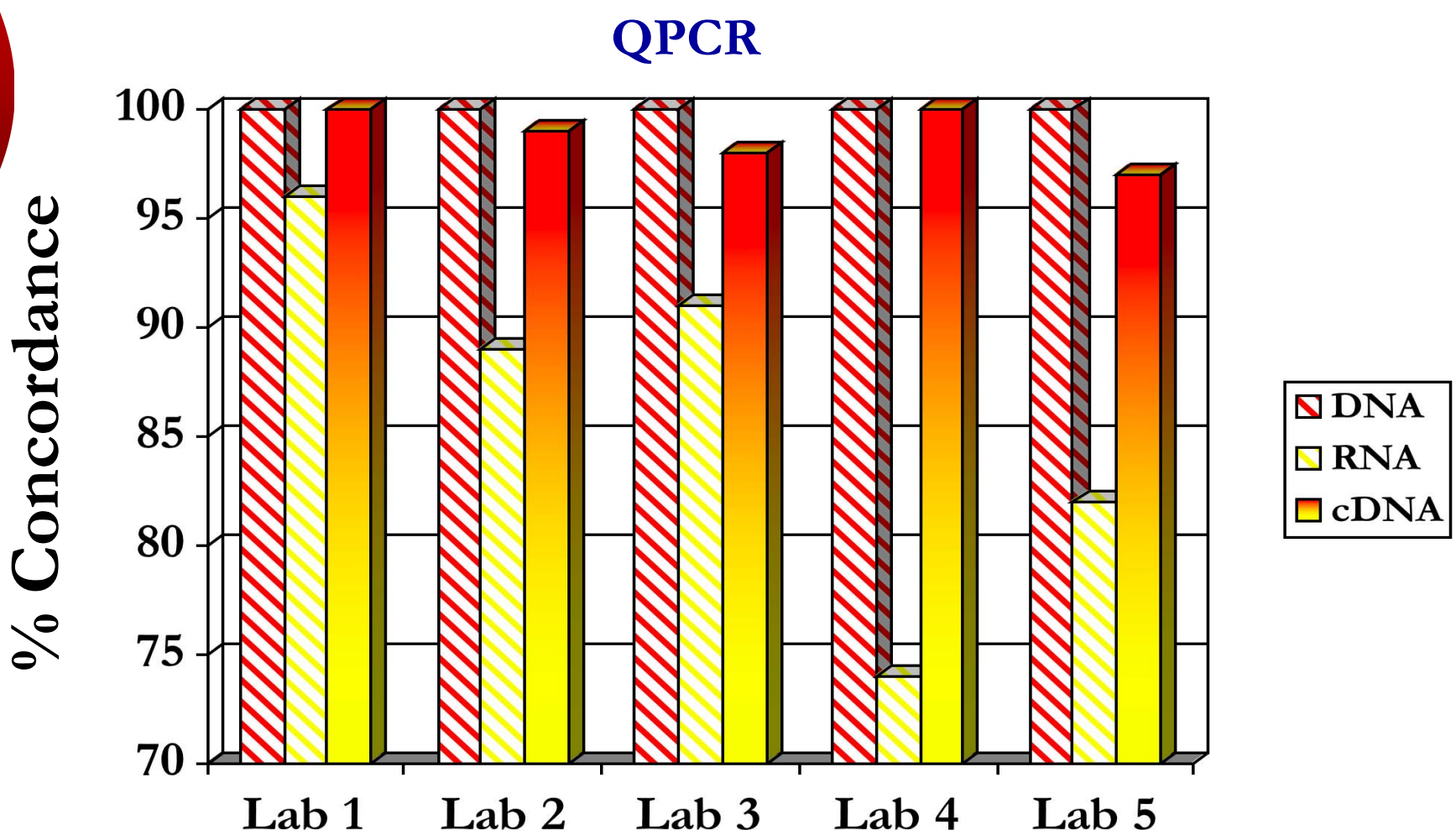
Sample Stability

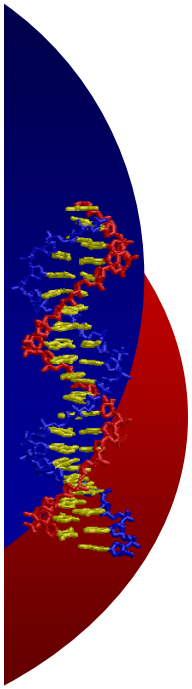


Metrics for RNA Quality: Degradation as a Function of Use



Metrics for RNA Quality: Temporal and Technical Variation

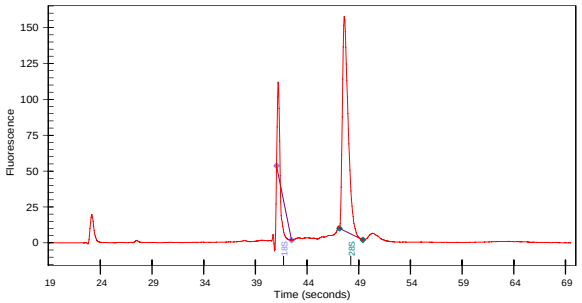


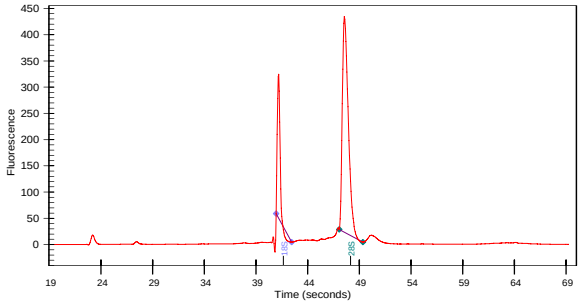


Quality Control – RNA Integrity

- 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

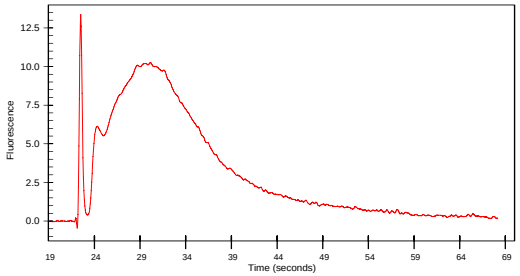
Raising the Bar for Sample Validation: A Standard Approach

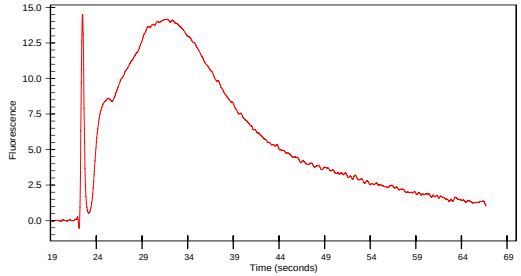




cDNA yield = 8.5 ug

cDNA yield = 7.4 ug





Genes Present = 49%

B-actin 3'/5' ratio = 9.6

GAPDH 3'/5' ratio = 1.6

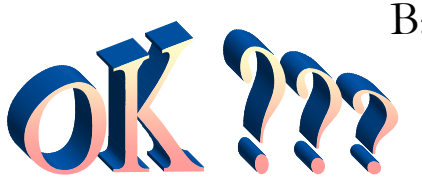
Background = 53

Genes Present = 51%

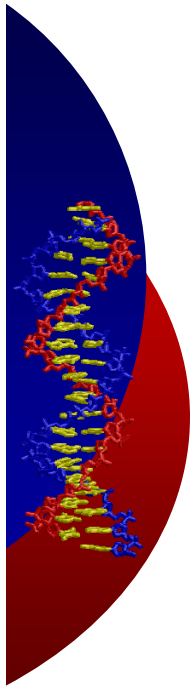
B-actin 3'/5' ratio = 5.4

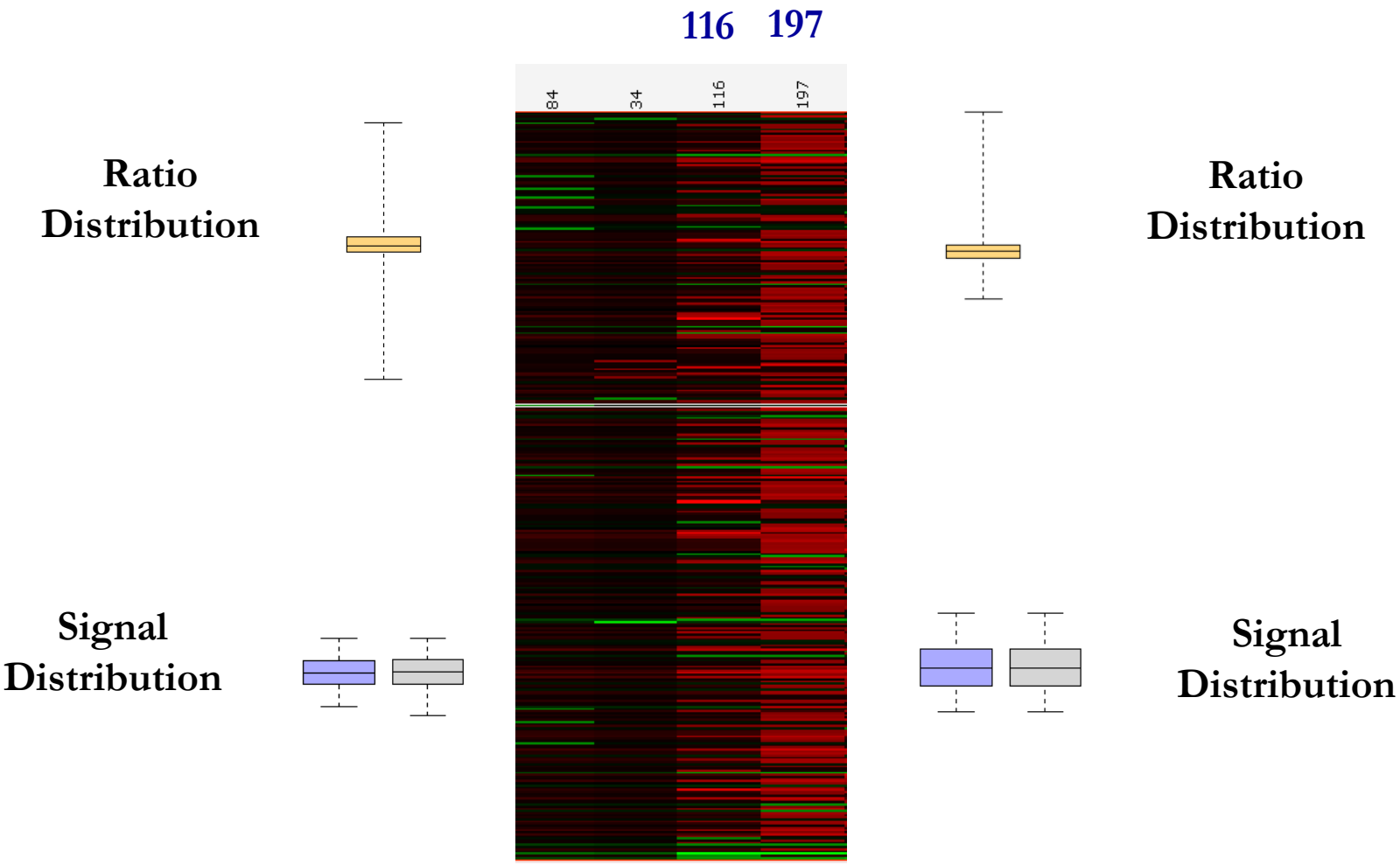
GAPDH 3'/5' ratio = 1.8

Background = 52

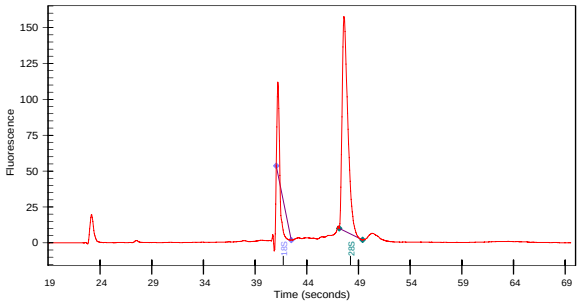


Reasons for Concern: Biology or Technology?

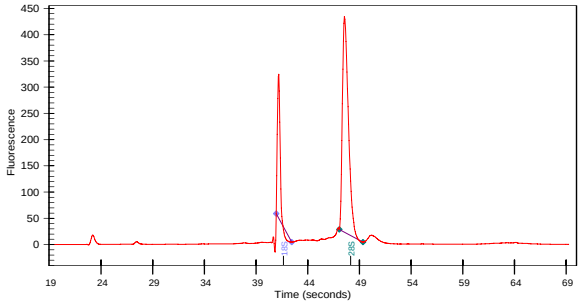




cDNA distribution (focused)= Sample Quality Assessment

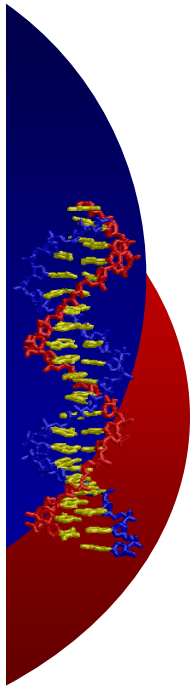


cDNA yield = 8.5 ug



cDNA yield = 7.4 ug

Gene 1a (3')	Pass	High Expresser	Pass
Gene 1b (M)	Pass		Pass
Gene 1c (5')	Pass		Pass
Gene 2a (3')	Pass	Medium Expresser	Pass
Gene 2b (M)	Pass		Pass/Fail
Gene 2c (5')	Pass		Fail
Gene 3a (3')	Pass	Low Expresser	Fail
Gene 3b (M)	Pass		Fail
Gene 3c (5')	Pass		Fail



Independent Consistent Expression Discriminator (ICED)

Step 1: Assigning Weights

$$W_1(g) = \frac{\frac{1}{m}\sum_{i=1,m}\left|g_{2i} - \mu_{1,n}(g)\right|}{\sigma_{1,n}(g)} \qquad W_2(g) = \frac{\frac{1}{n}\sum_{j=1,n}\left|g_{1j} - \mu_{2,m}(g)\right|}{\sigma_{2,m}(g)}$$

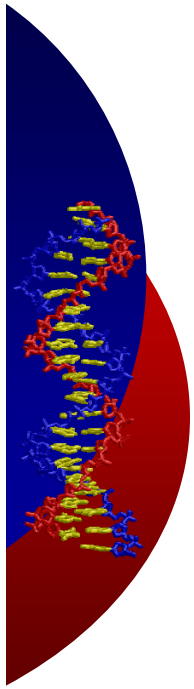
Step 2: Assigning Votes

$$V_1(g) = W_2(g) \bullet \left|g_x - \mu_{2TR,m}(g)\right|$$

$$V_2(g) = W_1(g) \bullet \left|g_x - \mu_{1TR,n}(g)\right|$$

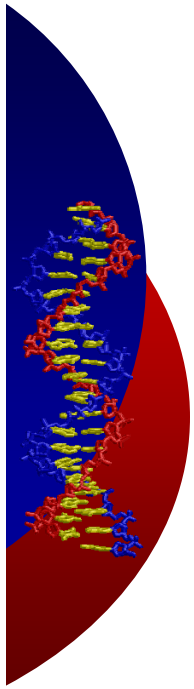
Step 3: Counting Votes

$$P(x) = \frac{q \bullet \sum_{i=1,p} V_1(g_i) - p \bullet \sum_{i=1,q} V_2(g_i)}{q \bullet \sum_{i=1,p} V_1(g_i) + p \bullet \sum_{i=1,q} V_2(g_i)}$$



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

RUTGERS UNIVERSITY
RUCDR
CELL & DNA REPOSITORY

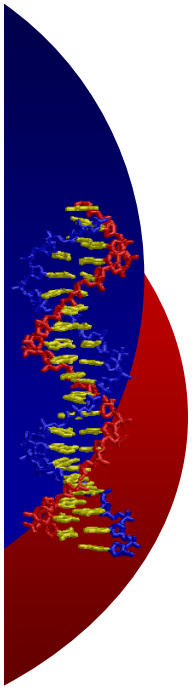
Biobanking: Is there a real need?

- 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”**

Come visit us…



RUTGERS UNIVERSITY
RUCDR
CELL & DNA REPOSITORY

Applying Technological & Business Infrastructure to Complex Disease Research

http://www.RUCDR.org

brooks@biology.rutgers.edu