Reference Materials for Confidence in NGS Applications

Marc Salit, NIST

Leader, Multiplexed Biomolecular Science Group

Multiplexed Biomolecular Science Group at NIST

- Development of "Measurement Infrastructure" for genome-scale measurements
  - reference materials
  - reference data
  - reference methods and procedures
- ***translation of genomics***
  - gene expression
    - controls, methods, metrics
  - genome-wide variant calling
    - genomic DNA standards
    - synthetic structures
  - microbial detection and identification
    - biothreat
    - food safety
    - human health
    - environmental monitoring

Some Reference Material Products

- NIST SRM 2374
  - *DNA Library for External RNA Controls*
  - collaborative with the External RNA Control Consortium
  - template library for mRNA mimics
  - useful for technical assessment of gene expression experiments
  - 1st reference material of its kind, certified for sequence
    - numerous alternative applications

- In development:
  - Control yeast strains for microbial detection/identification performance assessment
    - biothreat NA detection validation
  - Transcript isoform controls
    - for detection, quantitation of isoform enrichment
  - Reference Materials for whole genome sequencing
    - Genomic DNA Reference Material for confidence in variant calling
      - largely SNV, small indels
    - Artificial DNA structures for structural variant calling
      - large scale variants

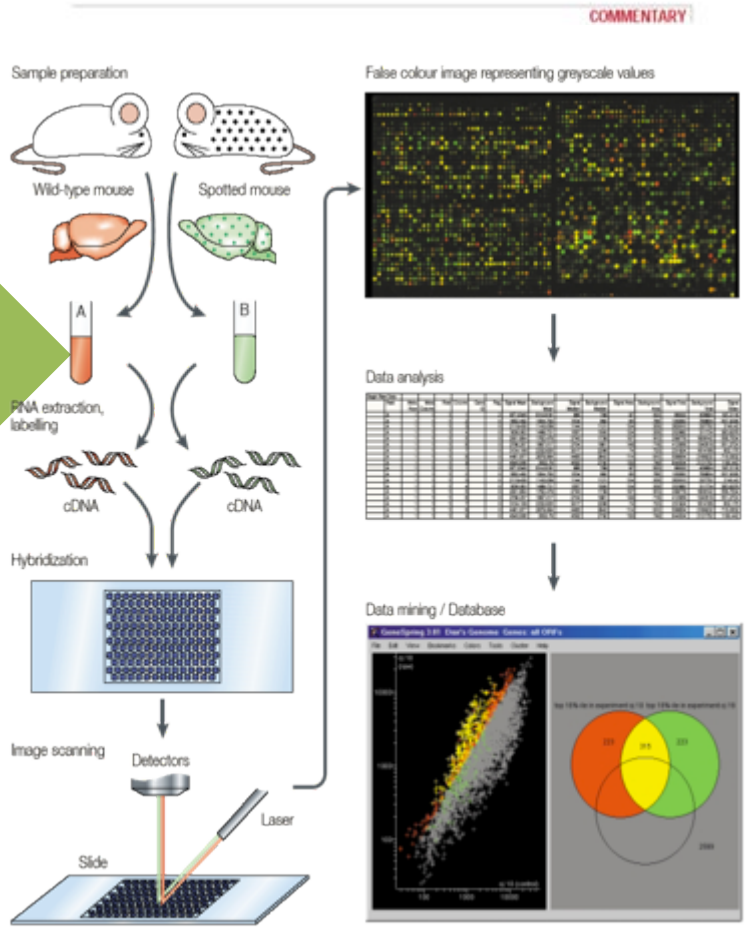
SRM 2374 – *DNA Sequence Library for External RNA Controls*

- NIST SRM containing 96 unique control sequences inserted in common plasmid DNA
  - engineered to be readily *in vitro* transcribed to make RNA controls
  - RNA controls intended to mimic mammalian mRNA
- Developed *sequence library* from submission by ERCC members, as well as synthesis
  - evaluated performance of RNA controls on variety of platforms
  - selected 96 well-performing sequences in collaborative study
    - from library of 176
- Array manufacturers modified products to include SRM sequences
- Prepared ~400 units of SRM
    - 96 tubes in each

External RNA Control Consortium (ERCC)

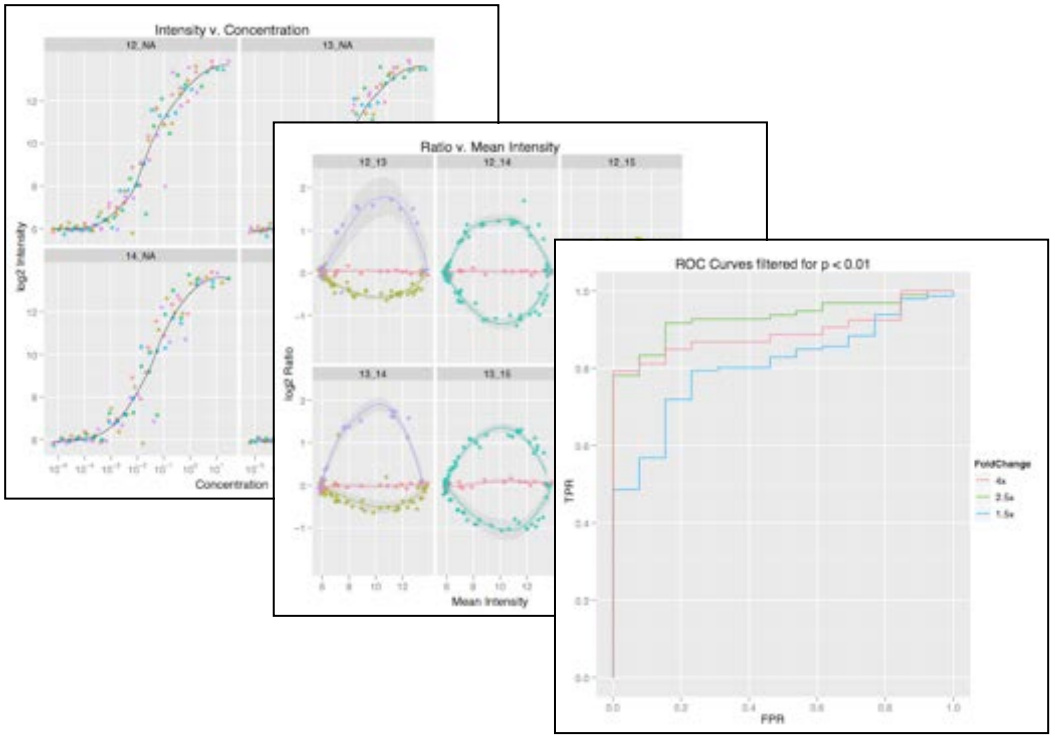
- Industry-initiated, NIST-hosted, stakeholder coupled
  - grew out of NIST workshop in 2003
    - initiated by Janet Warrington, VP Clinical Genomics at Affymetrix
  - all major microarray technology developers
  - other gene expression assay developers
- Open to all interested parties
- Voluntary
- More than 90 participants
  - Private, Public, Academic
- Model for future work in this area
  - sets up our work to be well-coupled to stakeholder needs
  - keeps us relevant and tied in
  - lets us SRMs that are readily adopted
- Standards will be 1st basis for comparability in this field

Spike-ins



Spike-in RNA Controls

- Applications of External RNA Controls
  - quantitative assessment of technical performance
    - RNA-Seq, Microarrays, qRT-PCR
    - Can be added to every sample
    - Multilevel performance assessment
- Some projects using/planning to use:
  - ENCODE/ModENCODE
  - ClinSeq
  - EDRN
  - TCGA
  - FDA SEQC



Alternative Application: Examine Systematic Sequencing Errors

- As sequencing costs drop, higher coverage can be used to minimize random sequencing errors
- Systematic sequencing errors (SSEs) are still significant at high coverage
- SSEs are especially problematic for detecting low fraction variants
  - E.g., tumor samples, pooled samples, mitochondrial heteroplasmy, metagenomics, RNA editing
- Spike-ins can help detect SSEs

Genomes in a Bottle:

Measurement assurance for genomic variation

NIST Reference Materials (RMs)

- Goal: Enable translation of whole genome sequencing to regulated clinical applications
- Prospective RMs
  - Human cell lines (1000 Genomes Project)
  - Synthetic spike-ins with variants

Collaborations

- CDC GeTRM
- ClinSeq (Leslie Biesecker)
  - Use NIST DNA spike-ins for bias detections
  - RNA editing
- 1000 Genomes Project (Lisa Brooks)
  - Highly sequenced trios as RMs