LabKey User Conference 2016 Workshop

Quality Control

Josh Eckels, LabKey

10/7/2016

What Does "Quality Control" Mean?

- Understanding data
  - Norms, trends, and outliers
- Disallowing bad data
- Agreeing on workflows and standards
- Sharing data appropriately
- Communicating confidence

Pain Points

- Knowing what to monitor
- Knowing how frequently to monitor
- Making sense of variability
- When to ignore and when to intervene
- Lack of automation
- Remembering to review results

10/7/2016

What Can a Tool Do to Help?

- Automate data flow and analysis
- Capture context and metadata
- Provide errors and warnings at time of import
- Provide proactive notifications

Validating Data at Import

- Required
- Range
  - Min, max numeric values
- Regular expressions
- Lookup values

Programmatic QC – Trigger Scripts

- Written in JavaScript
- Run as rows are inserted/updated/deleted
- Works for API or UI access for most data types
- Provide warnings and errors
- Row or field level feedback

Programmatic QC – Transform Scripts

- Specifically for assay data
- Operates on full file
- Warnings or errors
- Can validate and/or transform data
- Shell out to external scripts

QC Status – Capturing in LabKey Server

- Datasets (part of study module)
  - Built-in QCState column
  - Study defines allowable values
  - All data is public, but easily filtered
- Flag column (part of experiment/assay module)
  - Free-form text
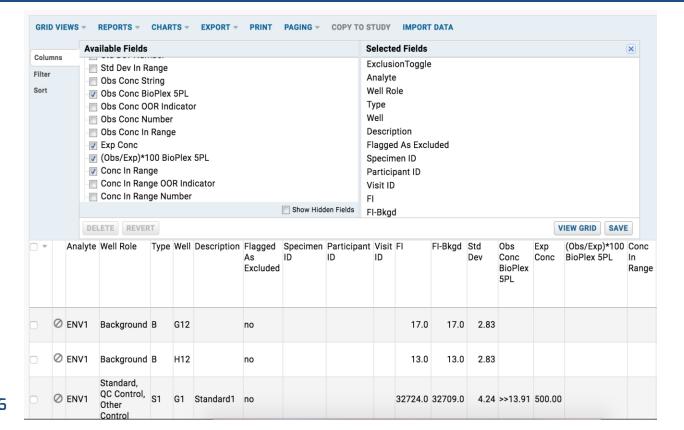 - Available on runs, materials, data files, etc

Out of Range Indicators

- Extra data column with "OOR" suffix
- Typically text ("<", "<<", ">" etc)
- Supplements "real" data column
- Flags value as being outside limits of detection, limits of quantitation, etc

Out of Range - Screenshots

Missing Value Indicators

- Similar to Out of Range
- Setting in designer on administrator-defined types
- Modeled on SAS
- Discrete list of allowable indicators at folder level

Data Sandboxing

- Allows private area for "raw" data
- Copy/move to shared location after performing QC
- Copy to study (assays)
  - Link individual data rows with a study dataset
  - Can copy to multiple studies
- Linked schemas
  - Allow custom SQL query to expose subset of data
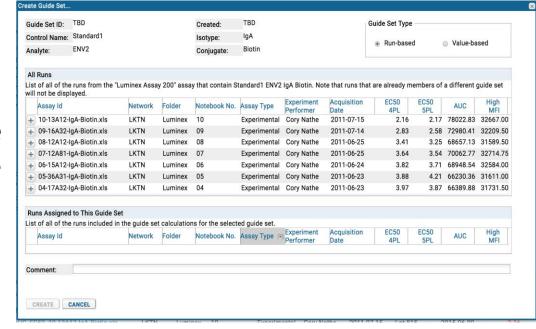 - Live view over source data, which may have different permissions

Luminex

- First built-in "deep dive" into QC analytics
- Plate-based assay (96-well)
- Multiplexed
  - 5-40 analytes per run
- Dilution assay
  - Titrated controls
  - Curve fits, EC50s, etc
  - Unknowns calculated against controls

Luminex − Guide Sets

- Guide sets
  - Establish the normal range for a metric
  - Defined at an analyte level, with date range
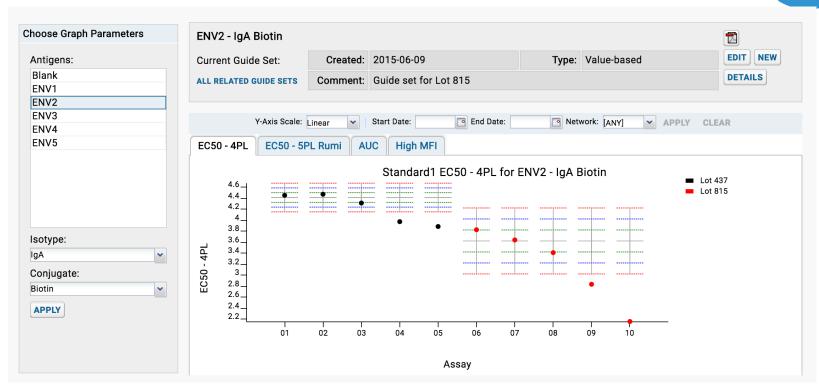 - Calculate range or import from external source

Luminex – Levey-Jennings Plots

- Visualize metric performance over time
- Based on mean and +/- 1, 2, 3 standard deviation
- Identify outliers
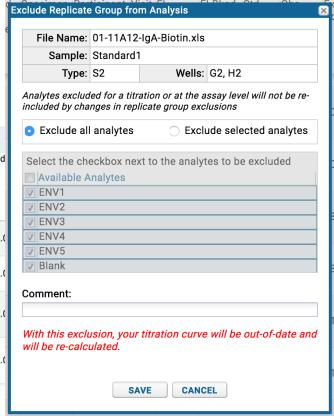- Generated on a per-guideset basis

Luminex – Levey-Jennings Plots

Luminex – Exclusions

- ## Mark data as invalid
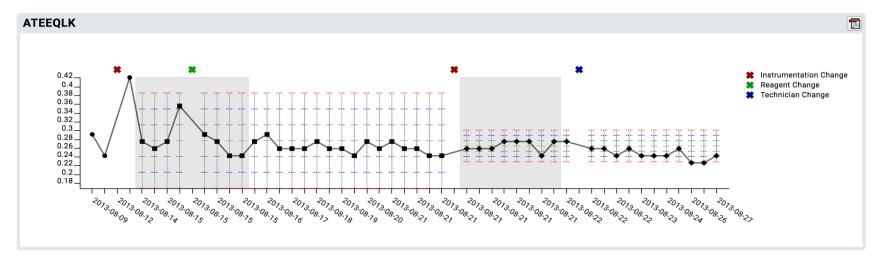- ## Well or analyte level

Panorama

- Targeted mass spectrometry (SRM, MRM, PRM)
- Specific folder type for QC analysis
- Inspired by Statistical Process Control (SProCoP)
  - Bereman et al
- Implemented in collaboration with MacCoss lab (UW) and Panorama Partners
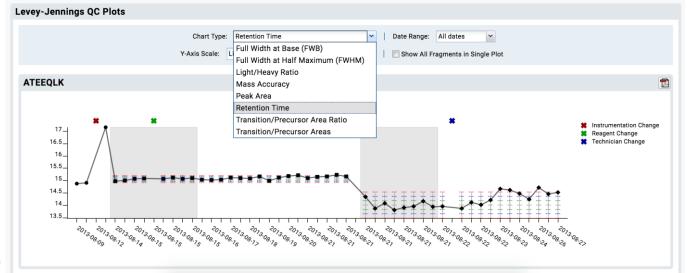
Panorama – Guide Sets

- Simplified to only specify date range
- Create directly on plots

Panorama - Metrics

- Growing list of values to monitor
- Customizable in the future

Panorama − Pareto Plots

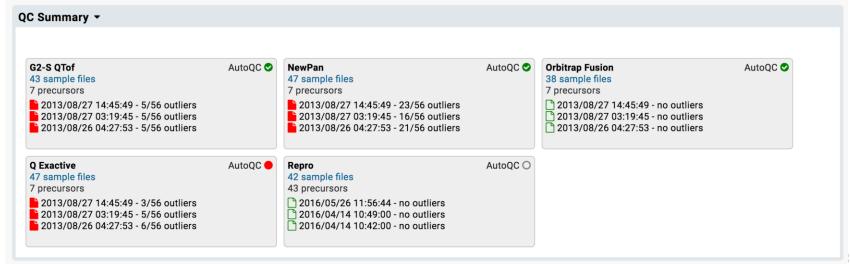- Identify most commonly problematic metrics

Panorama – AutoQC

- Automate workflow
- Monitors for newly acquired QC data files
- Analyze using Skyline
- Import into QC folder

Panorama – Dashboards

- Shows status for multiple instruments at once
- Indicates if AutoQC is running

Panorama – Future Plans

- More plots
  - CUSUM, mR
- More metrics
- Proactive alerts via email, etc

LabKey User Conference 2016 Workshop

Questions and Discussion

10/7/2016