Luminex Analysis:

Calibration Curves

Missingness

Normalization

DAVID L. TABB, PH.D.

MOLECULAR BIOLOGY AND HUMAN GENETICS

Outline

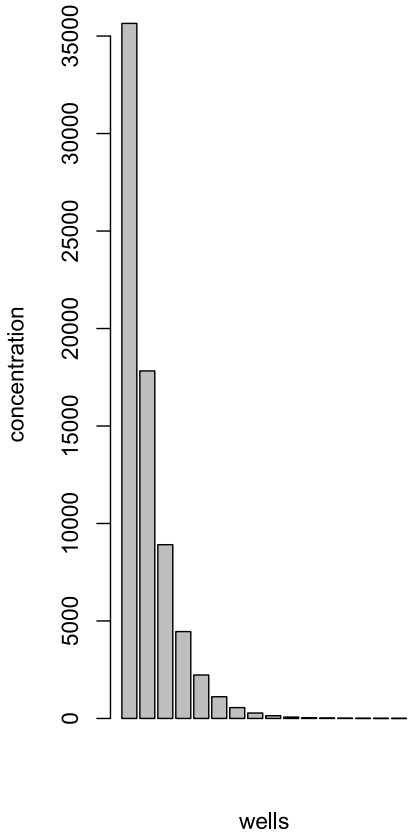
- Calibration, LOD, and LOQ: mapping fluorescent intensity to analyte concentration
- Missingness and Imputation: strategies for handling missing measurements
- Normalization: decreasing the undue influence of extreme measurements

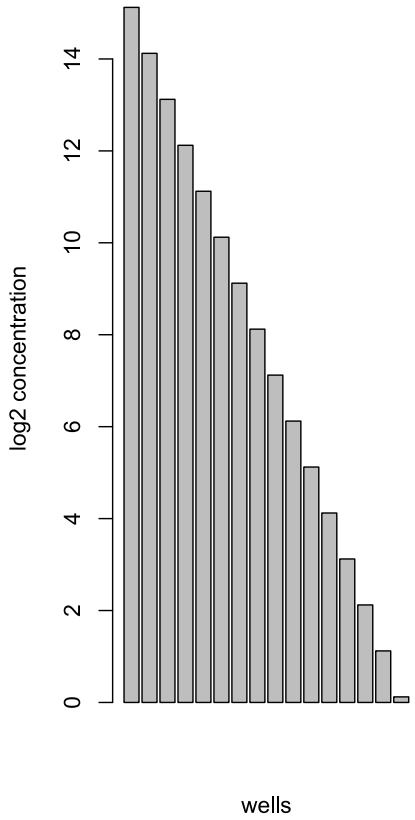
Why calibrate?

- We must infer analyte concentration from measured fluorescent intensity.
- Samples from different plates need to be compared on a common footing.
- We need to know the variability of measurement with respect to concentration.
- *Biologists care about analyte concentrations, not fluorescence.*

Typical strategy

- Reserve 16 wells of each 96-well plate.
- Spike large known amount in first well.
- Halve conc. for each successive well.
- Reserve other wells for blanks.

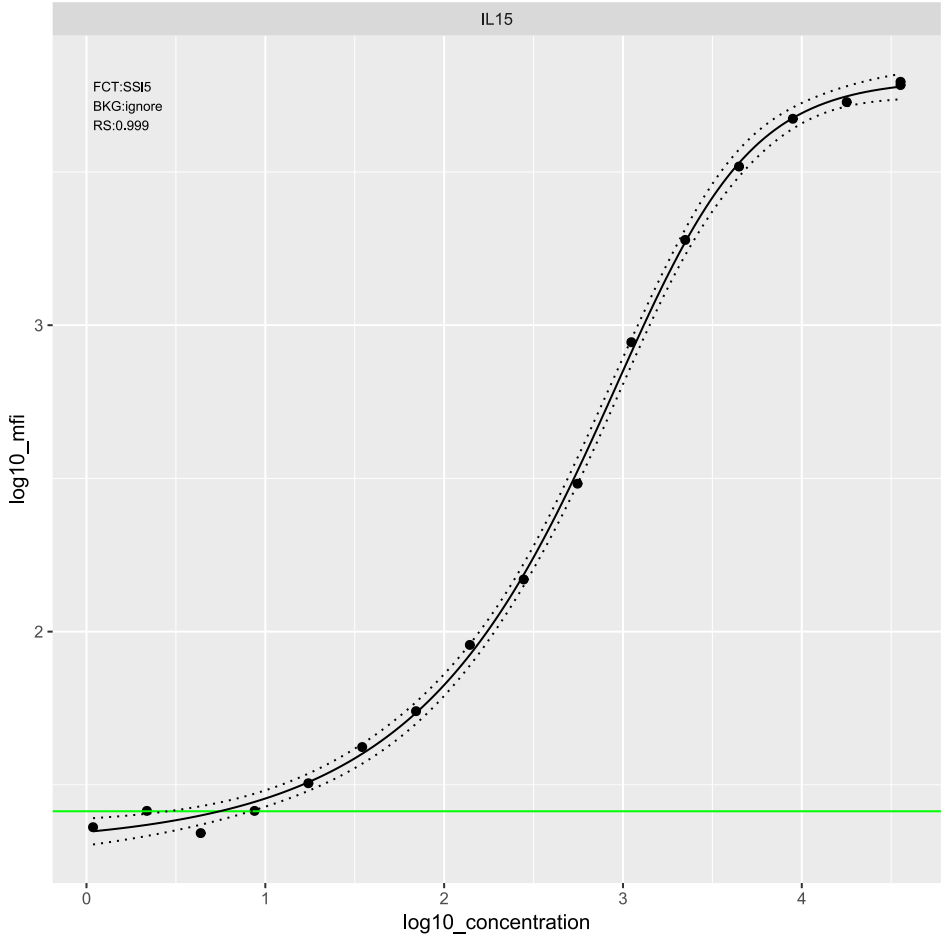




Data from drLumi R package, “ecdata” IL-15

Constructing a curve

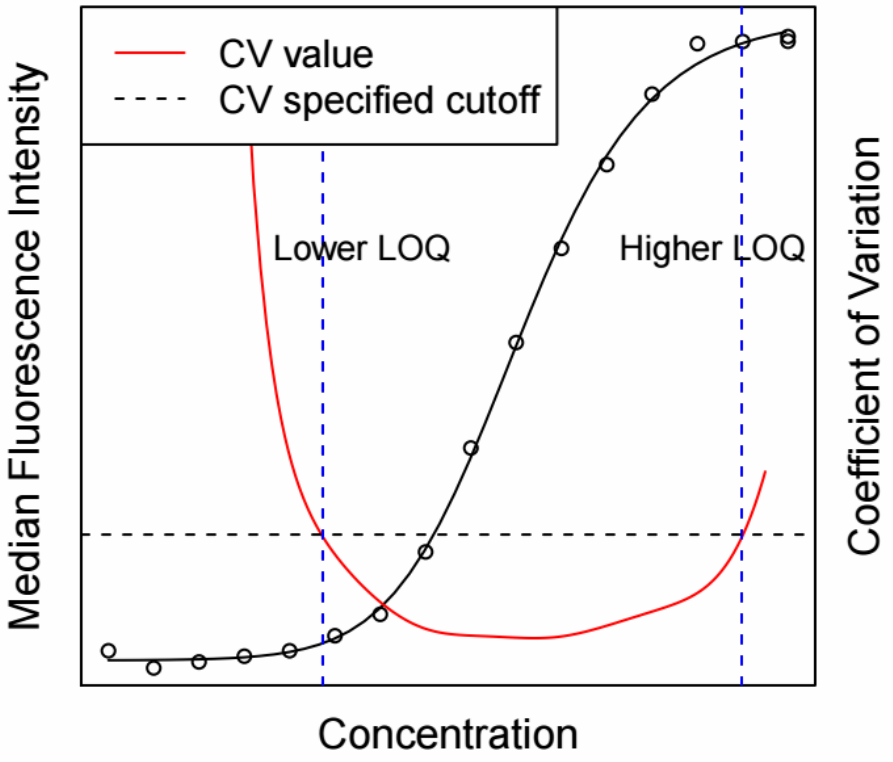
- Report median fluorescent intensity by conc. on log-log plot.
- Fit a logistic function to intensities.
- Dotted lines are confidence bounds.
- Green is blank intensity.



Logistic function is SSl5 for IL15, ignoring blanks

Determine limits of quantitation (LOQ)

- Intensity must exceed blank level for *detection*.
- Logistic model yields estimated coefficient of variation (CV) for given concentration.
- Quantifiable range is where CV is low enough.



From manual for drLumi package

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Interpolation = good, Extrapolation = bad

- The logistic model allows us to estimate concentration from measured intensity.
- If intensity falls within LOQ, it falls between standard measurements: *interpolation*.
- Some intensity values are more extreme than LOQ, allowing only *extrapolation*.
- Extrapolation leads to *Out of Range* reports: "OOR <" or "OOR >"

http://foldoc.org/extrapolate

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Why are concentrations missing from our data?

- Some cytokines are at extremely low levels or poorly detected: “OOR <”.
- We might have a sample too concentrated in this analyte for our method: “OOR >”.
- A technician may fail to include the beads for an analyte or get a bad batch.
- A sample may be limited in quantity.

Different types of missingness

- MCAR: Missing completely at random

  A bit of lint blocks light from a microarray.

- MAR: Missing at random (as a function of other recorded characteristics)

  One sex is less willing to report weight.

- MNAR: Missing not at random

  Rich people are less likely to report income.

Strategy 1: Listwise deletion

- If a cytokine is observed in only half of your samples, it may be better to remove it from consideration.
- If a sample is lacking many measurements that other samples report, it may be better to remove that sample from consideration.

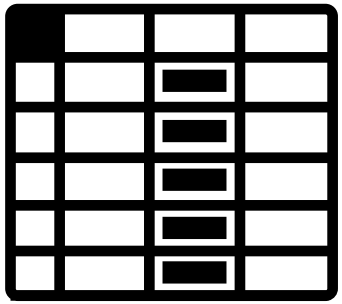
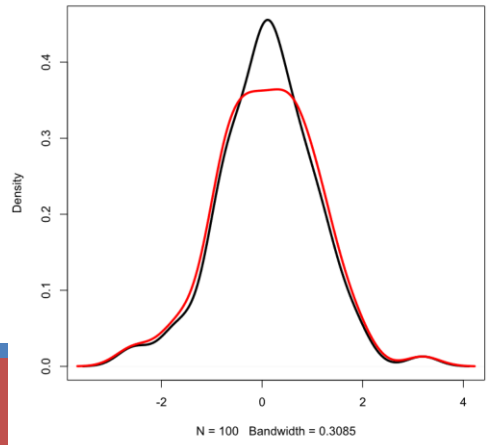


Image by FreePik.com
http://www.stat.columbia.edu/~gelman/arm/missing.pdf

Strategy 2: Impute mean or min

- Compute the mean of all observations for this analyte; overwrite missing cells.
- Compute the minimum of all observations for this analyte; overwrite missing cells.
- *Either option will severely distort the distribution of data, for example reducing the standard deviation!*



Red=before; black=after; 10% of data imputed

Strategy 3: Model to impute

- Simple random imputation: pick another observed value for this analyte to fill the gap.
- Linear model: fit a regression that can estimate analyte *Z* based on measures of analytes *A* thru *Y*. Relies upon correlation.
- Multivariate imputation: use a multivariate normal or t distribution-based model to fill in the gaps for multiple analytes at once.

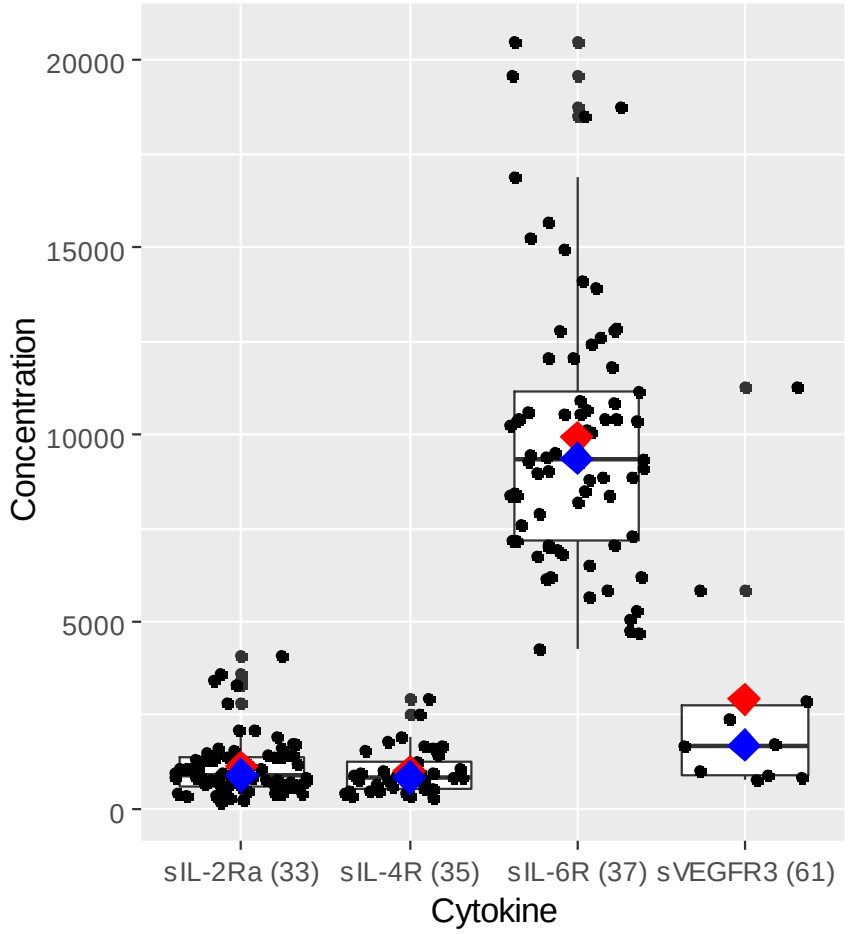
Image from starwars.com

Strategy 4: Use intensities directly

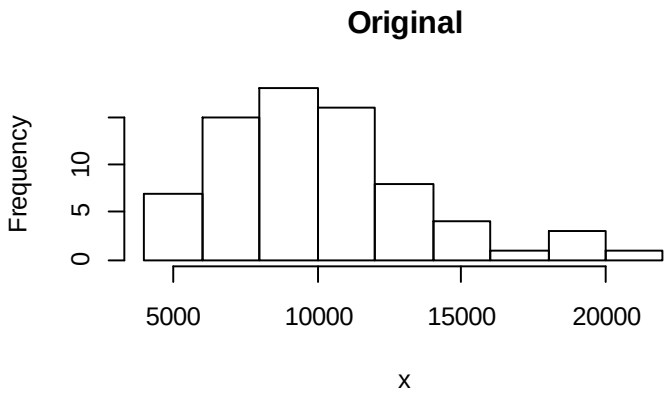
- We often want to know "does this analyte *differ* between case and cohort" rather than "*by how much* does this analyte differ."
- We have fluorescence values for all, but only some map to concentrations.
- Testing intensity rather than concentration is one way to side-step missingness.
- Comparing among different plates is unlikely to be safe without a batch parameter.

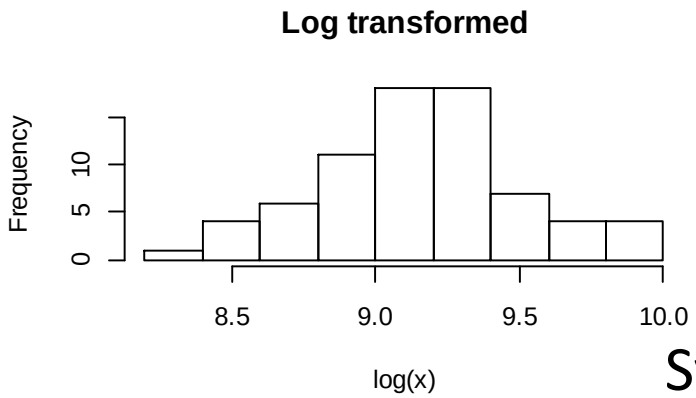
A long way from normal

- Plots mask “OOR <” values: 36/73 for sIL-4R, and 63/73 for sVEGFR3.
- Medians (blue) and arithmetic means (red) differ, warning of non-normality.
- We will use sIL-6R because of intensity; no “OOR <”



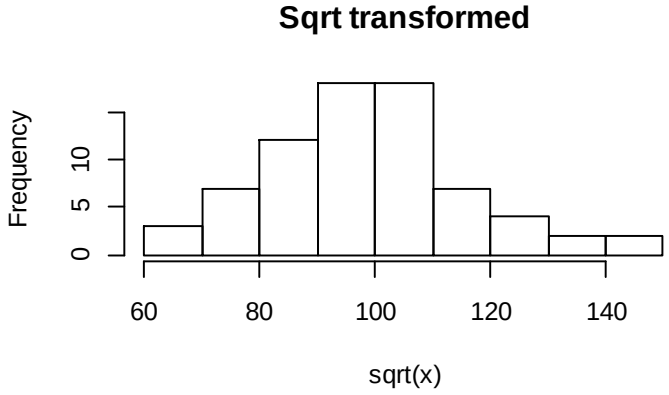
Simple functions

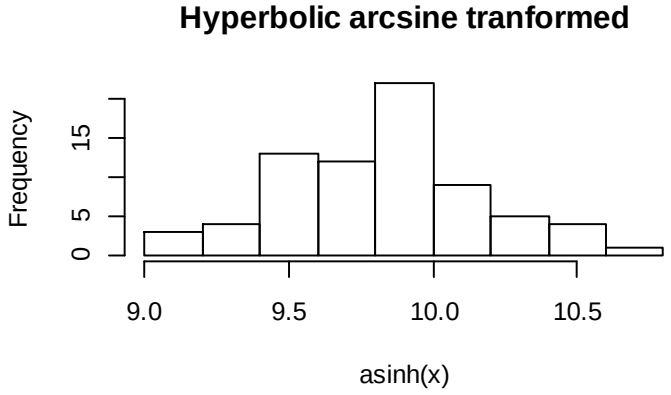




Symmetry

"Thick tails"

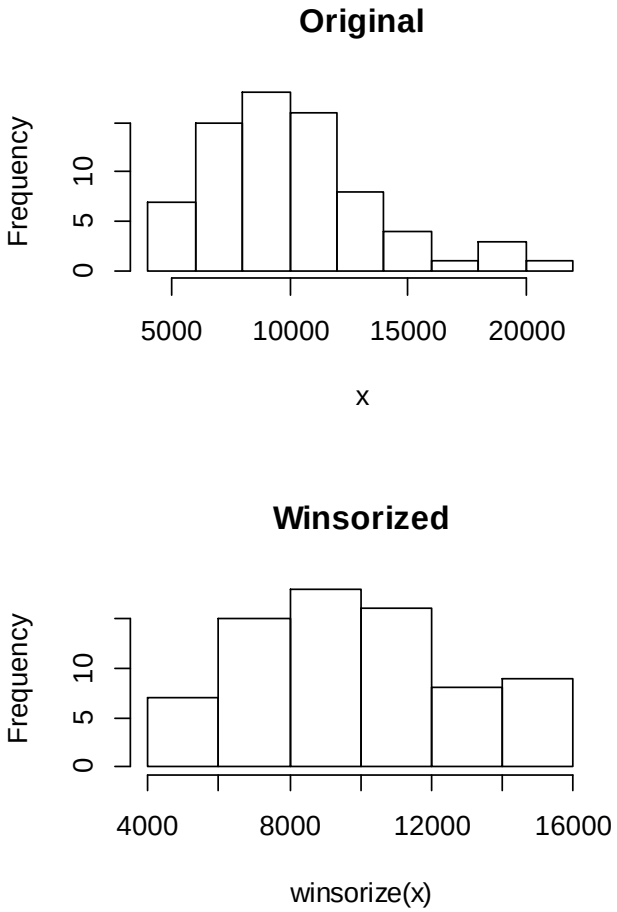




http://www.biostathandbook.com/transformation.html

Winsorization via *robustHD*

- Sample means are sensitive to extreme values / outliers.
- Winsorization pulls points from far left and right closer to main mass of data.
- The rank order of data points may be preserved, *or* ties may be created.



https://blogs.sas.com/content/iml/2017/02/08/winsorization-good-bad-and-ugly.html

Takeaway Messages

- Several wells of every plate supply the grist to determine concentrations from intensity.
- Concentration estimates are only reliable within the limits of quantitation (LOQ).
- Missingness is a fact of life. Understanding why values are missing is key to imputation.
- Data transforms help reduce outlier impacts and encourage normal residuals.