

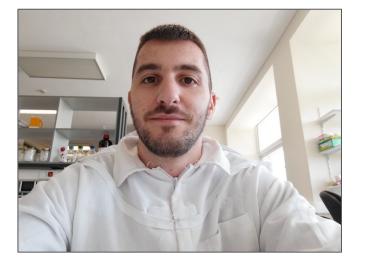
After iGEM Academic Publishing Workshops

Analysis and Visualisation of Gene Expression Data by Dr Jacob Beal



Chair - Academia & Research Network

Speaker - iGEM



Anastasios Galanis After iGEM



Jacob Beal Raytheon BBN Technologies iGEM Engineering Committee



Data Analysis and Visualization

Content:

- Relating data to biology
- Graphical Presentation



RELATING DATA TO

BIOLOGY



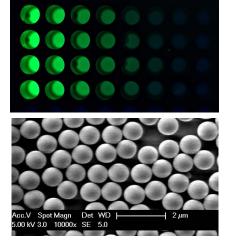


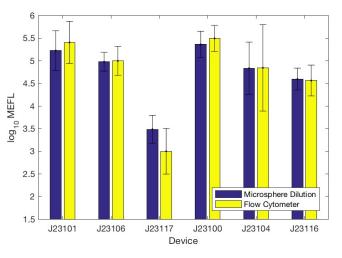


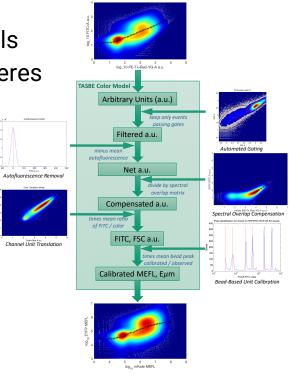
Avoid Relative and Arbitrary Units

Example: calibration of fluorescence & OD measurements

- Flow Cytometry: NIST-certified beads, WT, color controls
- Plate reader: fluorescein, Texas Red, cell-like microspheres
- Result: directly comparable MEFL units







[Roederer, '02; Wang et al., '08; NIST/ISAC, '12; Beal et al., '12; Castillo-Hair et al., '16; Beal et al., '18, Beal et al., '20, Beal et al., '21]



Think Carefully About Relating Measurement to Biology

Example: estimating cell and molecule counts from plate reader data

- Raw readings include background, even after calibration
- To estimates counts, subtract the background:

$$Cells = (OD - media) * \frac{particles}{OD}$$

$$\frac{MEFL}{Cell} = (a.u. -WT a.u.) * \frac{MEFL}{a.u.} * Cells$$



Use Geometric Statistics for Gene Expression

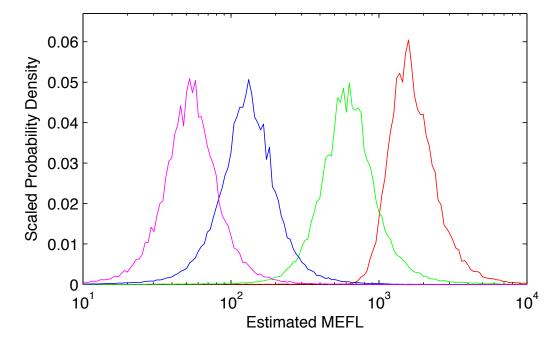
Why geometric stats?

Complex catalytic reactions → multiply many rates:

$$R_{express} = R_1 R_2 R_3 R_4 R_5 \dots$$

→ converge to log-normal!

Gamma distribution bursting also implies geometric stats



Take Advantage of Process Controls

Experimental Controls:

Academia & Research Network

AFTER IGEM

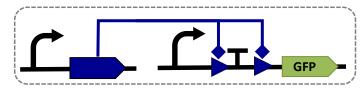
- Is my hypothesis true?
- One control per factor under study
- Best when new data
- Control very close to experiment conditions

Process Controls:

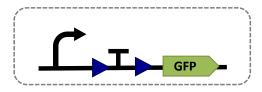
- Should I trust the data?
- One control per assumption in study
- Best when known value
- Control should have minimal relation to experiment conditions



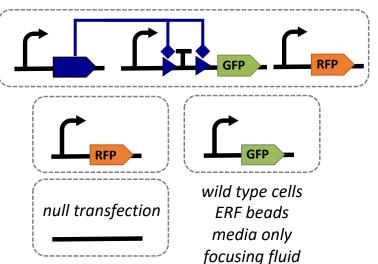
Example of Experimental vs. Process Controls



Experimental Controls:



Process Controls:



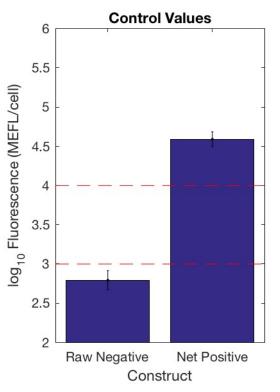


Sanity Check Your Control Values

- Compare to calibrants to ensure instrument linear range
- Compare positive to max number of proteins per cell:
 - E. coli: 2e6 Yeast: 6e7 Human: 2e9
- Negative control should be much smaller than positive
- Problems with values indicate likely process failure

Example: E. coli negative <1e3, positive >1e4

http://book.bionumbers.org/how-many-proteins-are-in-a-cell/



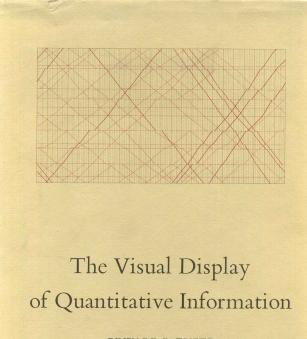


GRAPHICAL PRESENTATION





Presenting data is just as important as collecting it!



EDWARD R. TUFTE

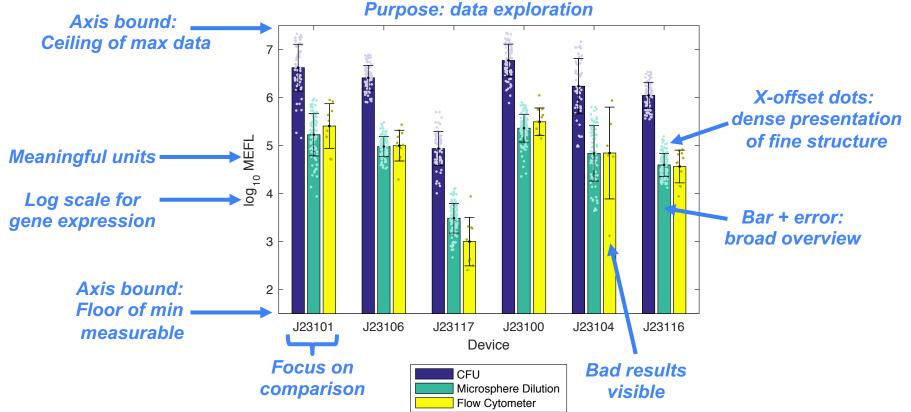


Core Principles of Data Presentation

- Show the data
- Focus the viewer on data comparison, not graphic design
- Avoid distorting the data Principled choice of axis bounds, scale Don't hide bad results
- Present many numbers densely and coherently
- Show both broad overview and fine structure of data
- Serve a clear purpose: e.g., description, exemplification, exploration
- Integrate graphics with statistical and prose descriptions



Applying the Core Principles



Example from iGEM 2018 interlab publication: [Beal et al.,'20]



Tell stories in your captions

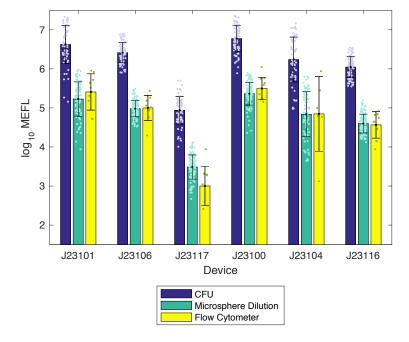


This is a horse.

What about it?



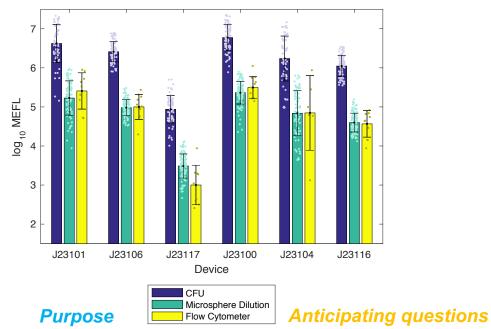
Tell stories in your captions



Fluerescence per cell after 6 hours of growth.



Tell stories in your captions



Fluorescence per cell after 6 hours of growth, comparing calibrated flow cytometry to estimates using cell count from CFU and Key results microsphere dilution protocols (LUDOX/water is not shown as the units it produces are not comparable). Microsphere dilution produces values extremely close to the ground truth provided by calibrated flow cytometry, whereas the CFU protocol produces values more than an order of magnitude different, suggesting that CFU calibration greatly underestimates the number of cells in the sample. Bars show geometric mean and standard deviation. Team count per condition provided in Supplementary Data 3 Teams Per Condition.

QUESTIONS?







After iGEM Academic Publishing Workshops

THANK YOU!