log2 transformation

Mikhail Dozmorov

Fall 2017

Expression Summaries: cDNA arrays

· For custom spotted arrays, the quantity used for analysis is most often the
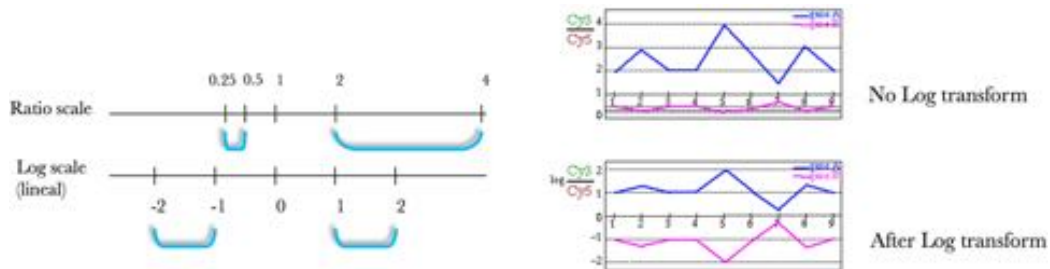
$$log_2(\frac{SampleSignal}{ReferenceSignal})$$

· This ratio may or may not include background subtraction

$$log_2(\frac{SampleSignal - SampleBackground}{ReferenceSignal - ReferenceBackground})$$

Problems with fold change

- Fold changes, or ratios, can be larger than 1 (2-fold increase), or smaller than one (0.5).
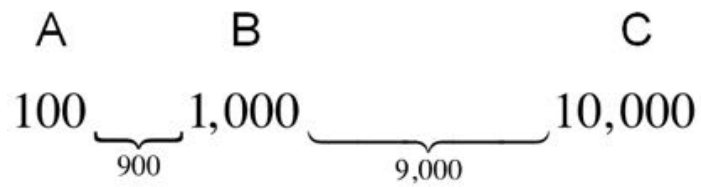- Not symmetric around 1

Why log2

- $log_2(\frac{x}{y}) = log_2(x) - log_2(y)$ - fold change converted to difference
- log ratios are symmetric around 0
- $log_2(1) = 0$
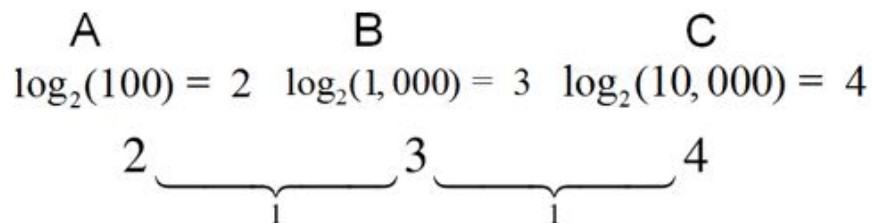- $log_2(2) = 1$
- $log_2(0.5) = -1$

log2 transformation

$$A \qquad B \qquad\qquad C$$

$$100 \underbrace{}_{900} 1{,}000 \underbrace{}_{9{,}000} 10{,}000$$

Condition	Fold change	Difference
B vs. A	10	900
C vs. B	10	9000
C vs. A	100	9900

log2 transformation

$$A \qquad\qquad B \qquad\qquad\qquad C$$

$$\log_2(100) = 2 \quad \log_2(1{,}000) = 3 \quad \log_2(10{,}000) = 4$$

$$2 \underbrace{}_{1} 3 \underbrace{}_{1} 4$$

- Note that on a log scale,
- The differences are 1.

log2-transformation of raw intensities

- The fold change distribution has a fat right tail
- The log2-transformed fold changes are linear