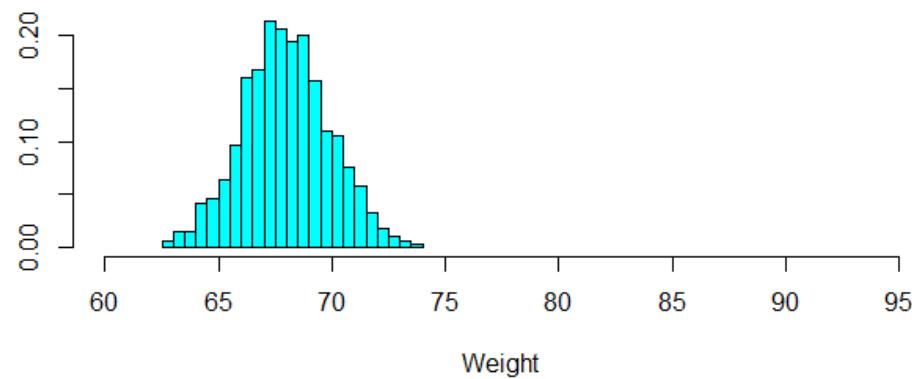
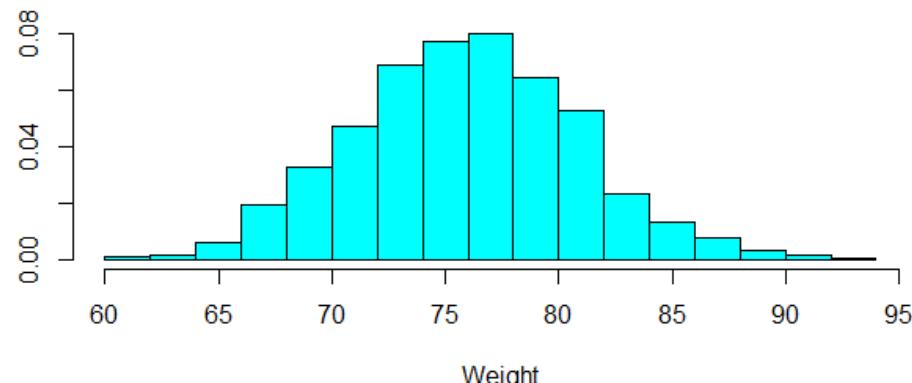


# 第14回バイオインフォマティクス 研究会

マイクロアレイデータの正規化について

2014/07/28

# データの正規化



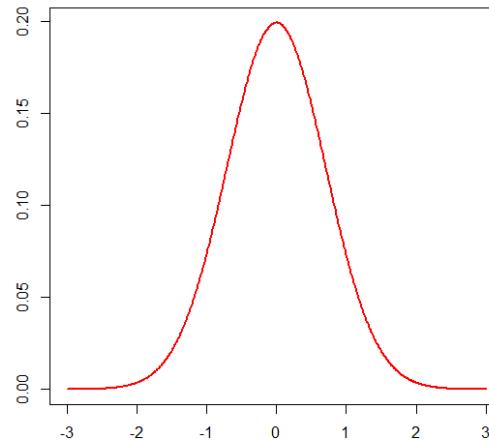
# Z-Score化

$$z_i = \frac{x_i - \bar{x}}{v}$$

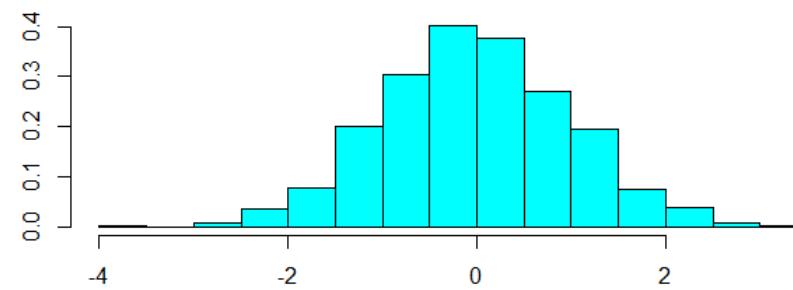
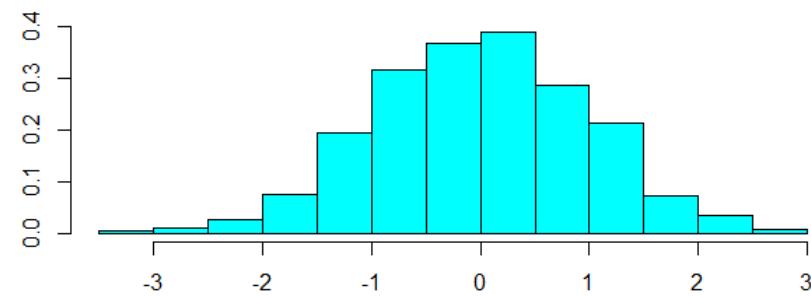
平均:  $\bar{x} = \frac{\sum_{j=1}^N x_j}{N}$

標準偏差:  $v^2 = \frac{\sum_{j=1}^N (x_j - \bar{x})^2}{N - 1}$

平均0、標準偏差1の標準正規分布に補正



# Z-Score化



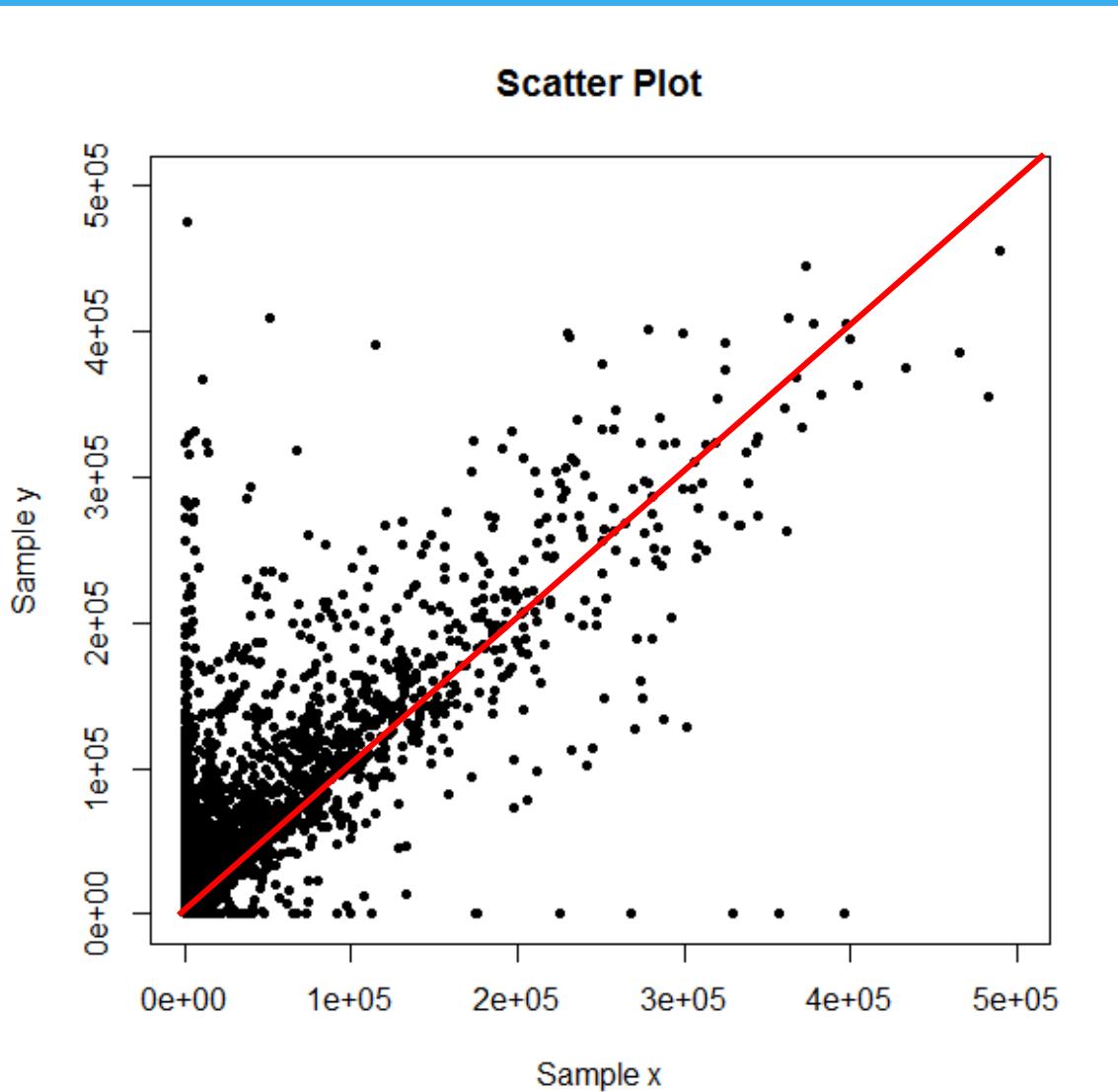
# マイクロアレイデータの前処理

- \* るべき分布を想定し、それに合うようにデータ値を補正する。

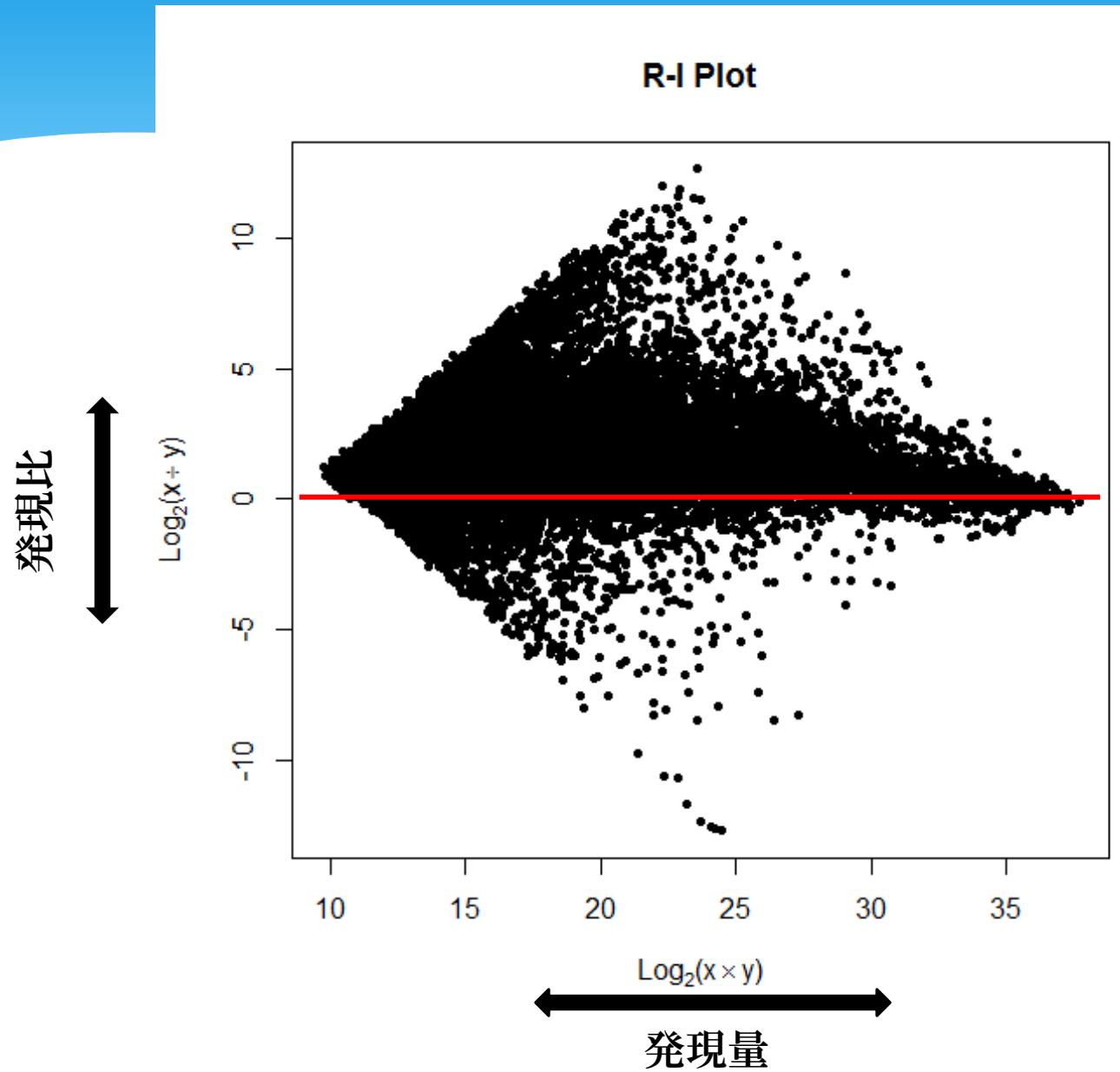
# DATA

- \* ERR-α ノックアウトマウス GSM1412850
- \* Brd1ノックアウトマウス GSM1184703
- \* Agilent-028005 SurePrint G3 Mouse GE 8x60K

# Scatter Plot (正規化前)



# R-I Plot (正規化前)



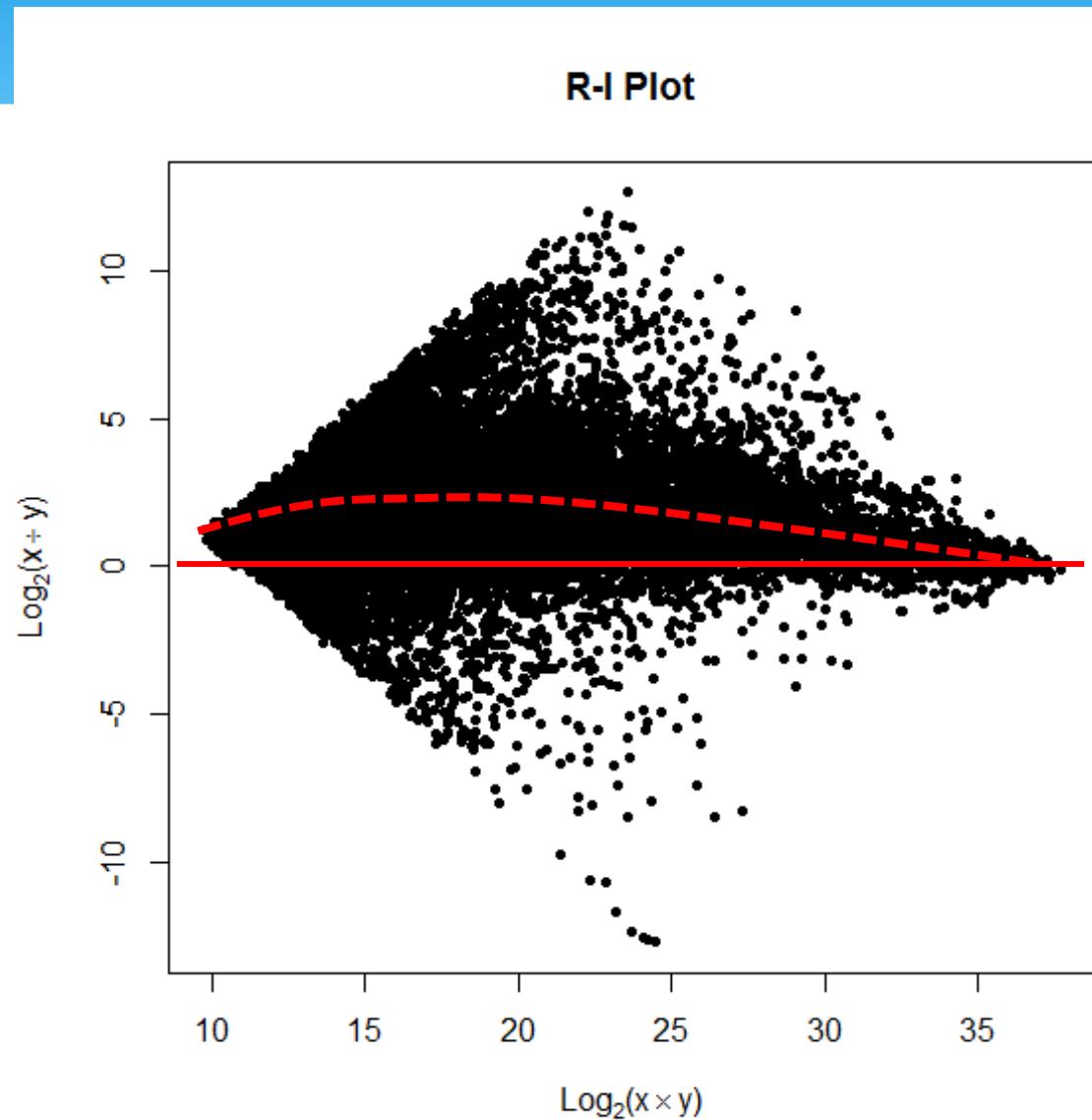
# 正規化

- \* LOWESS (Locally Weighted Scatter Plot Smoothing)
- \* Quantile normalization

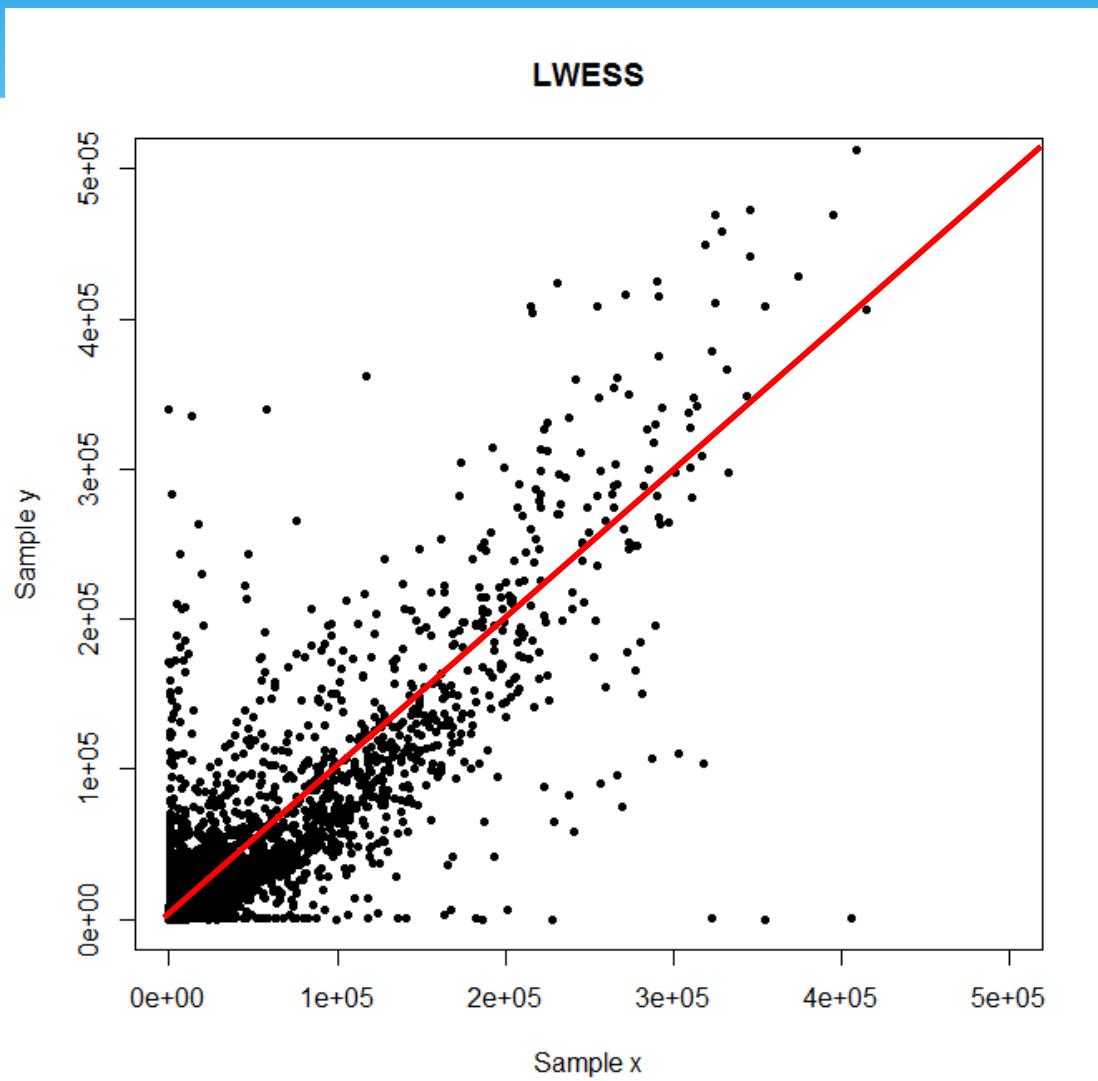
# LOWESS

- \* 発現量の少ない遺伝子ほど発現量がばらつく傾向がある
- \* ほとんどの遺伝子の発現は変動していない

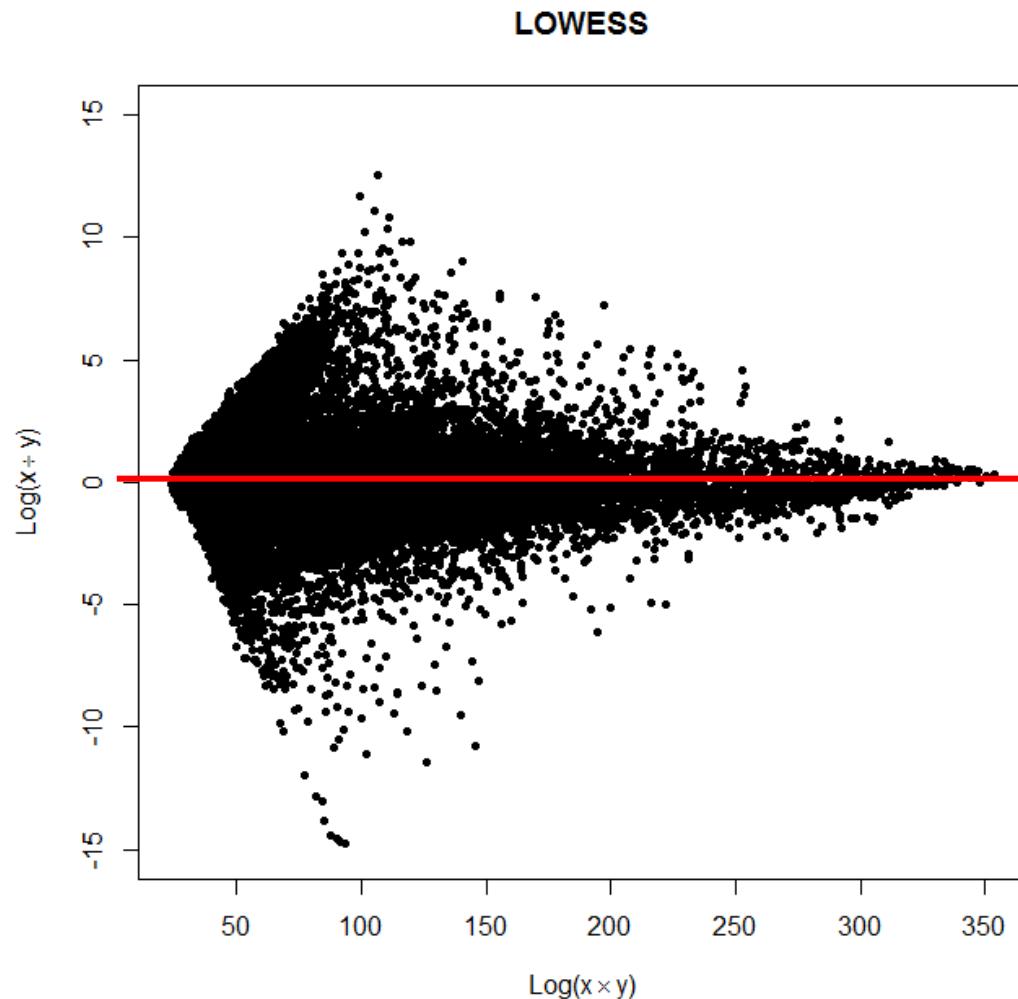
# R-I Plot (正規化前)



# Scatter Plot (LOWESS)



# R-I Plot (LOWESS)



# quantile normalization

- \* 遺伝子を発現量順に並べたとき、同じ順位の遺伝子は同じ発現量を示す。

# quantile normalization

gene	Sample x	gene	Sample y
gene1	13.0	gene1	14.0
gene2	51.0	gene2	55.0
gene3	34.0	gene3	87.0
gene4	21.0	gene4	66.0
gene5	15.0	gene5	12.0
...	...	...	...

# quantile normalization

gene	Sample x	gene	Sample y	平均
gene1	13.0	gene5	12.0	12.5
gene5	15.0	gene1	14.0	14.5
gene4	21.0	gene2	55.0	38.0
gene3	34.0	gene4	66.0	50.0
gene2	51.0	gene3	87.0	69.0
...	...	...	...	...

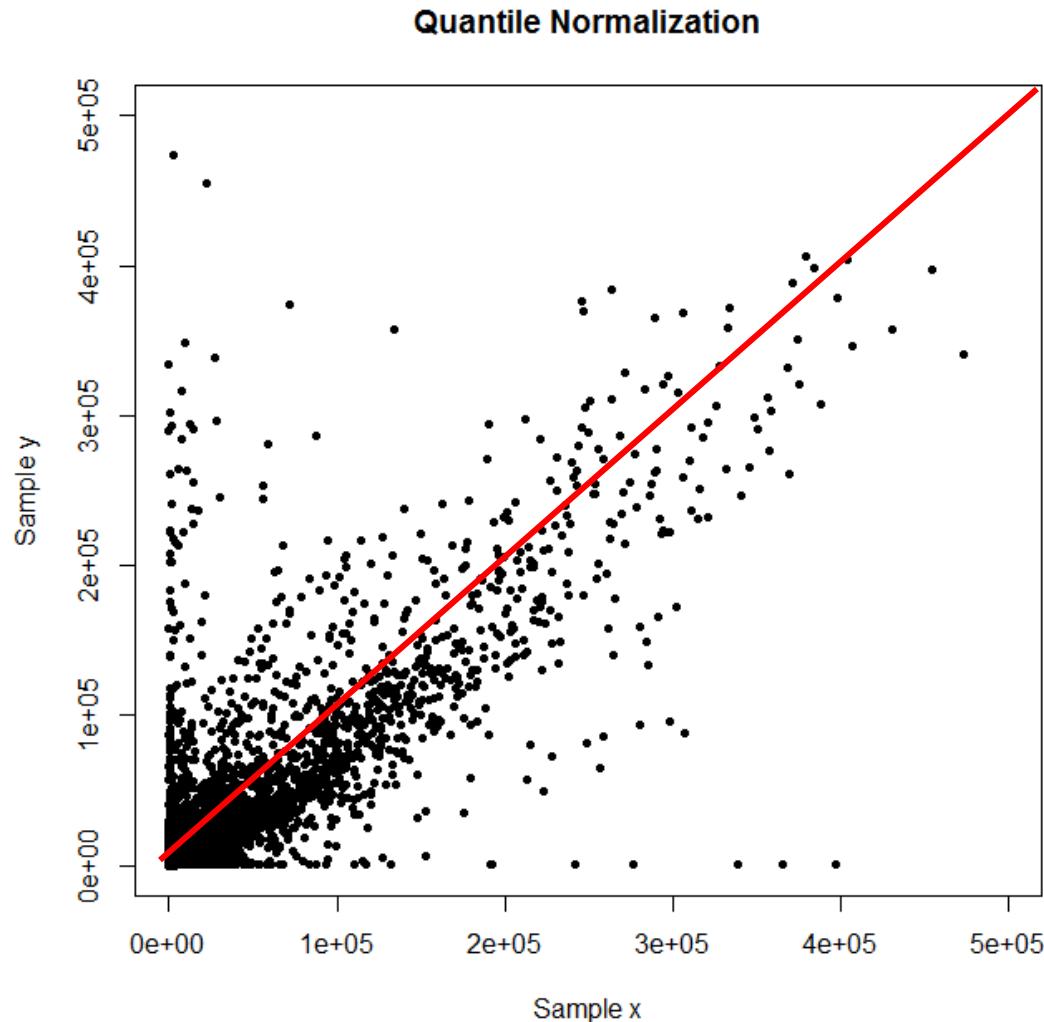
順位で並び替え、平均値を求める

# quantile normalization

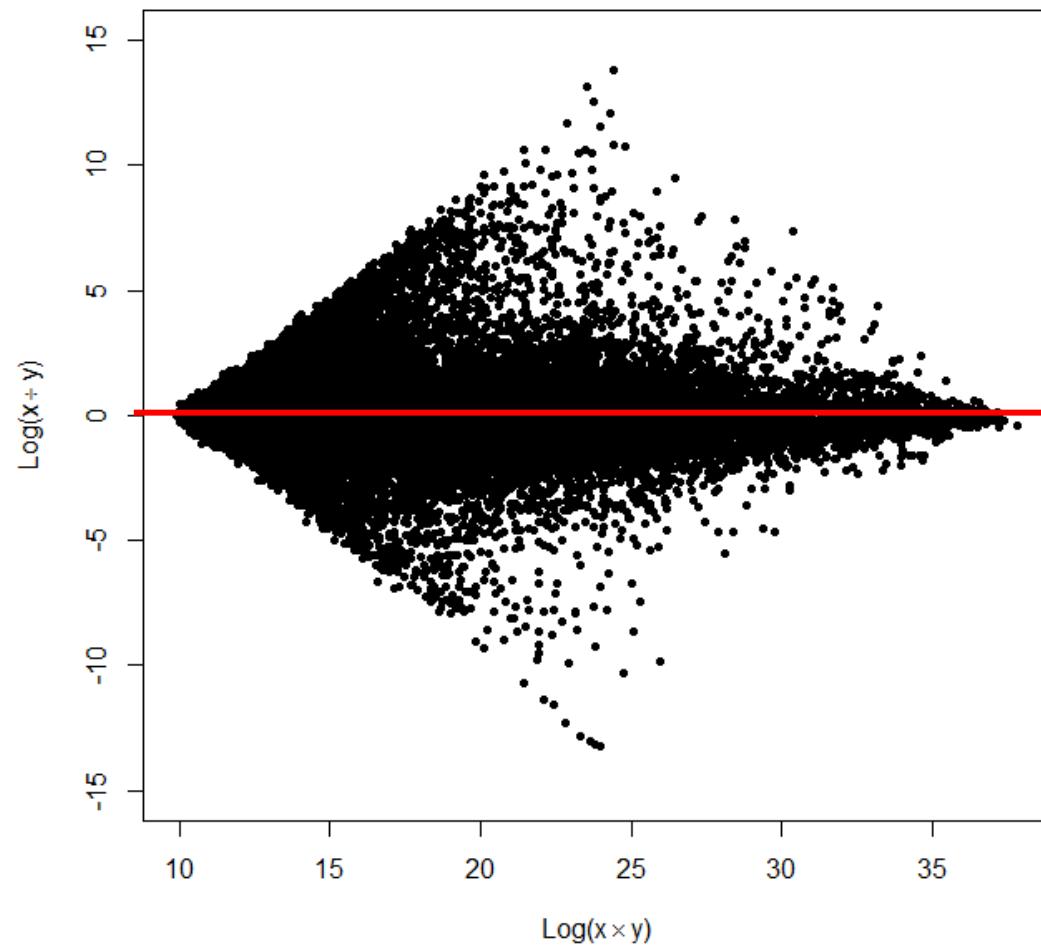
gene	Sample x	gene	Sample y
gene1	12.5	gene1	14.5
gene2	69.0	gene2	38.0
gene3	50.0	gene3	69.0
gene4	38.0	gene4	50.0
gene5	14.5	gene5	12.5
...	...	...	...

各遺伝子に値を割り当てる

# Scatter Plot (quantile normalization)



# R-I Plot (quantile normalization)



# 二群間比較

- \* Sample y / Sample x > 10

LOWESS	1015
quantile normalization	1234
共通	1011

# まとめ

- \* 別の実験で取得されたマイクロアレイデータを取得し、LOWESS、Quantile Normalization、二つの方法で正規化した。
- \* 正規化後のデータを用いて二群間比較を行った。
- \* 正規化法によって結果が異なることを確認した。