

MA Plot

Volcano Plot

Tran Thi Hong Minh

MA Plot

- In Microarray: normalize within arrays to control systematic biases.
- Minimizing systematic variations (dye coupling, hybridization efficiencies, technical biases) to discover true biological differences between two groups/conditions.
- MA plot compares two color channels, two arrays or two groups of arrays
- MA plot determines whether normalization is needed
- x axis: the difference between the logarithm of the signals
- y axis: the average of the logarithms of the signals

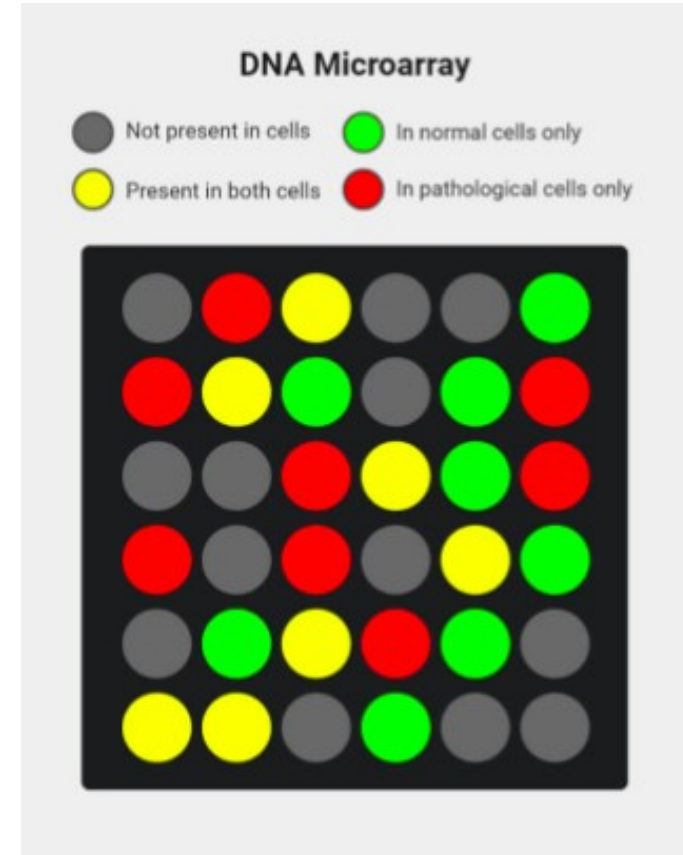


Image By Sagar Aryal, created using biorender.com

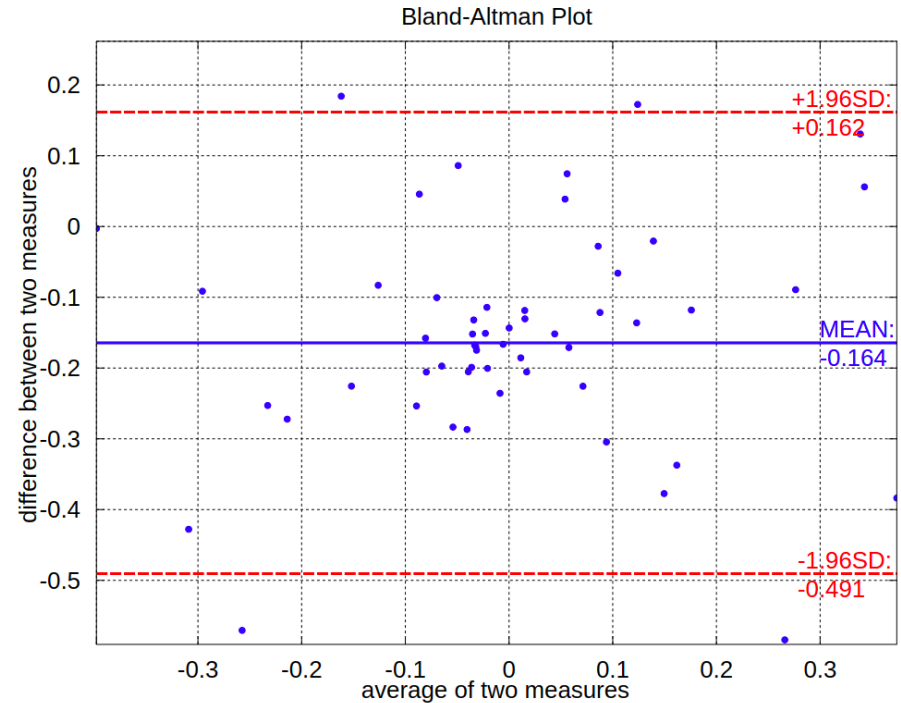
MA plot is an application of a Bland-Altman plot

Compare the disagreement, or differences,
between two quantitative measurements

The Bland-Altman plot measures the mean
differences along with limits of agreement.

These limits are calculated using the mean and
standard deviation of the differences between
the two measurements.

The Bland-Altman plot is not a measure of if the
differences between two measurements are
acceptable or not. The level of “acceptability” is
based on individual circumstances

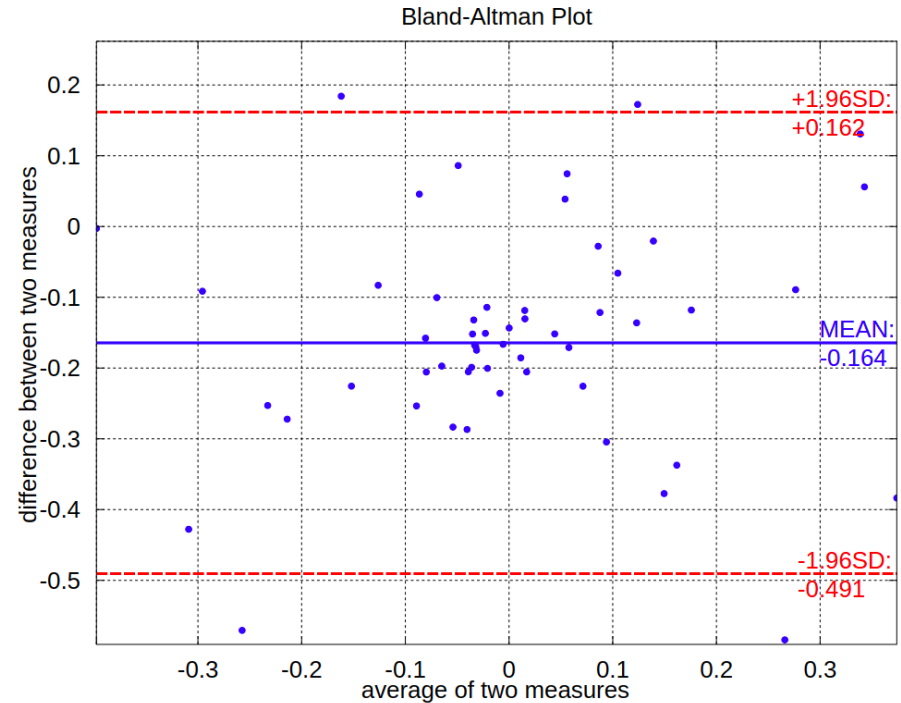


Y-axis: the difference between two paired measurements (A-B).

Blue line: the mean difference. For this plot, the mean difference is -0.164; this tells you that measurement procedure B, on average, differed from procedure A by -0.164.

Red Line: ± 1.96 standard deviations from the mean difference. Bland & Altman recommended that 95% of data points should fall within 2 standard deviations of the mean difference.

X axis: mean of the two measures $((A+B)/2)$.



	Condition 1	Condition 2
Gene 1	C_1	T_1
Gene 2	C_2	T_2
...	C_i	T_i

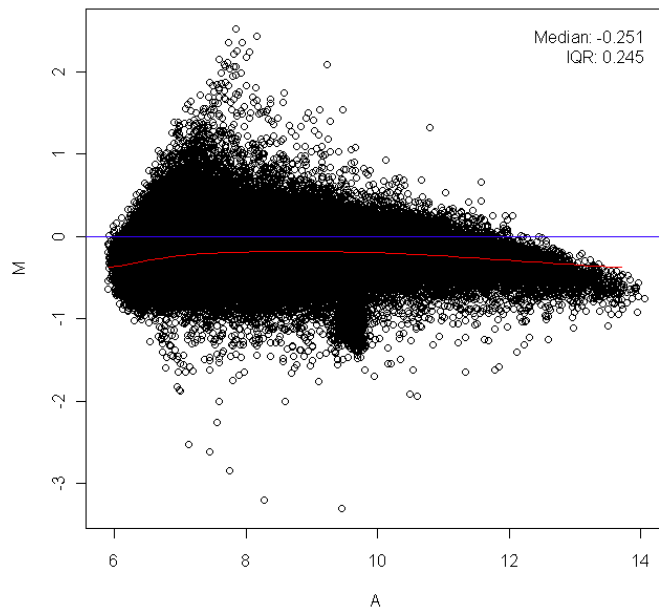
$$M_i = \log(T_i/C_i) \quad (\text{Log ratio})$$

$$A_i = [\log(T_i) + \log(C_i)]/2 \quad (\text{Average log intensity})$$

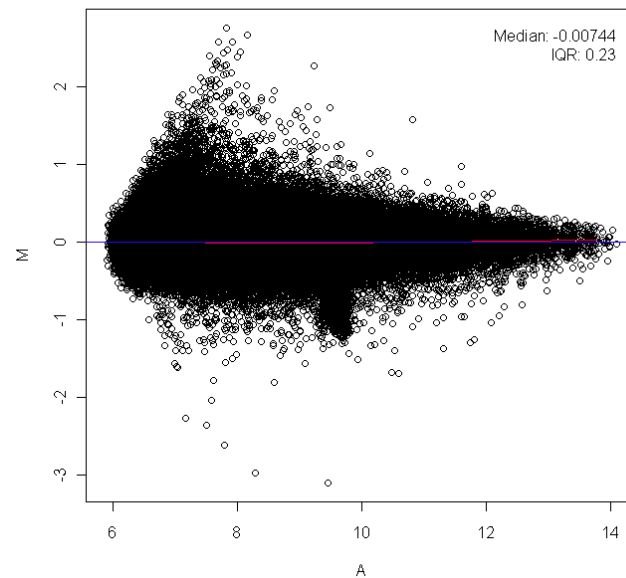
M: stands for
minus

; for add

Pre-Norm Dilutions Dataset (array 20B v 10A)



Post-Norm: Dilutions Dataset (array 20B v 10A)



	ensgene	symbol	baseMean	log2FoldChange	padj
1	ENSG000000000003	TSPAN6	747.1941954	-0.35070302	0.1630348
2	ENSG000000000005	TNMD	0.0000000	NA	NA
3	ENSG0000000000419	DPM1	520.1341601	0.20610777	0.1760317
4	ENSG0000000000457	SCYL3	322.6648439	0.02452695	0.9616942
5	ENSG0000000000460	C1orf112	87.6826252	-0.14714205	0.8158486
6	ENSG0000000000938	FGR	0.3191666	-1.73228897	NA



A



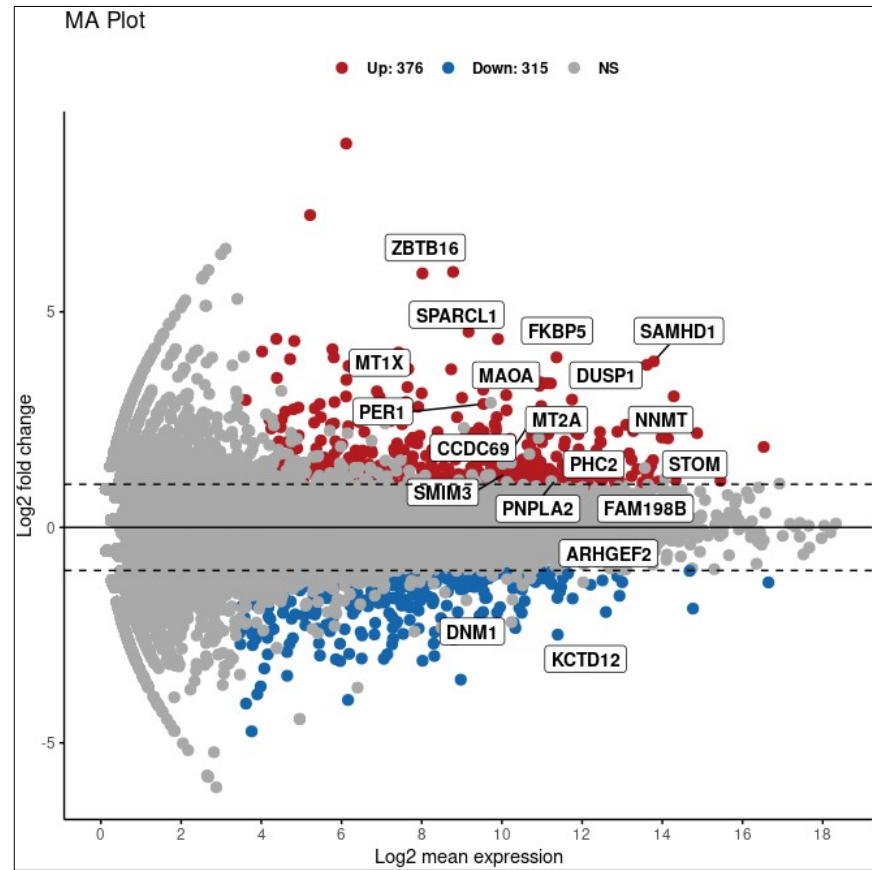
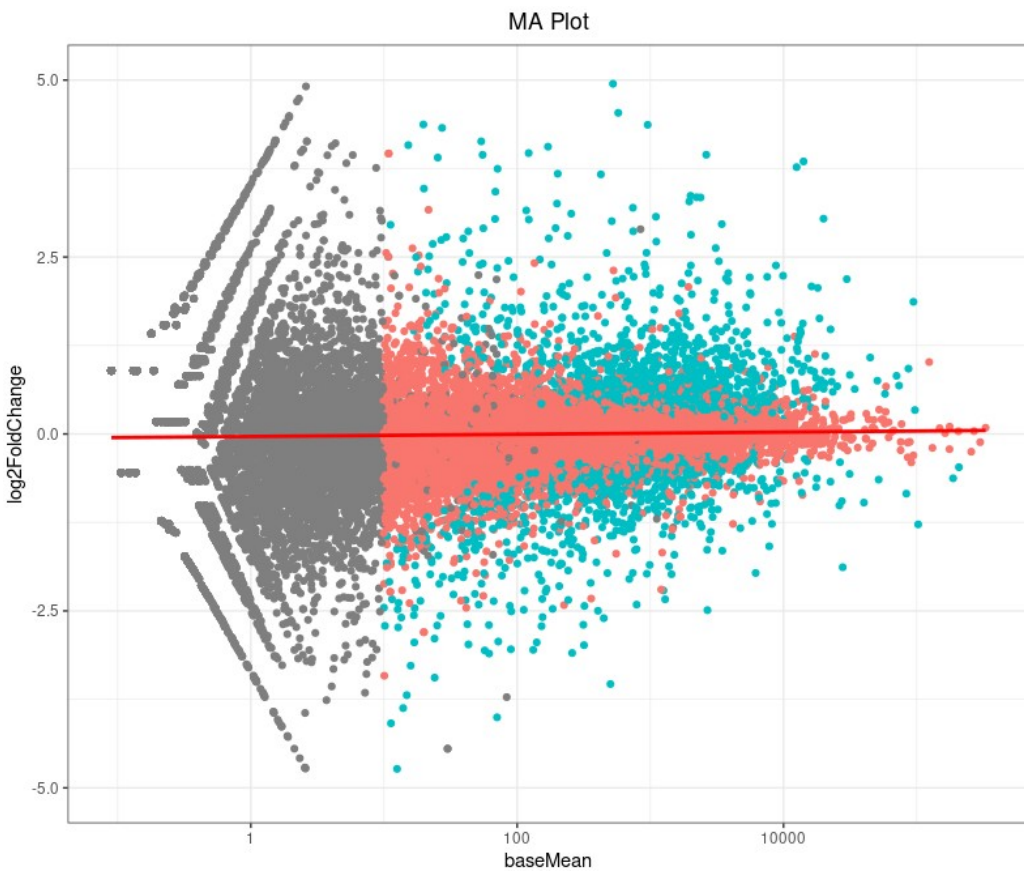
M



P-value after correct FDR

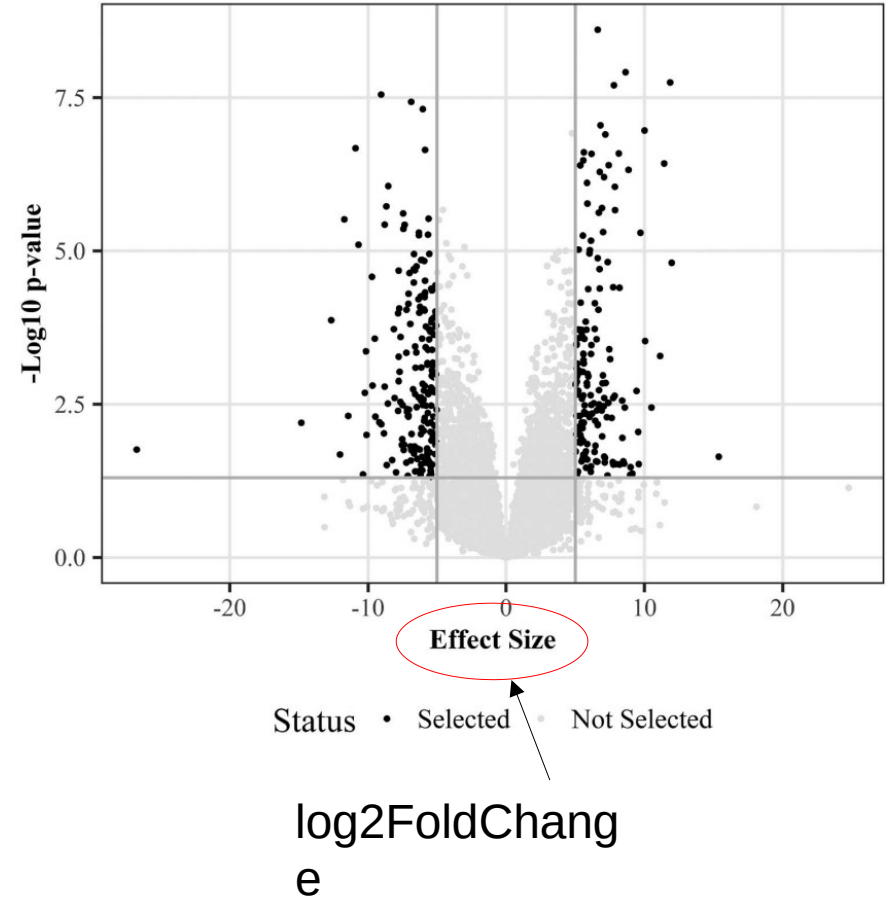
P-value is set to NA for some reasons:

- If within a row, all samples have zero counts, the baseMean column will be zero, and the log2 fold change estimates, p value and adjusted p value will all be set to NA
- If a row contains a sample with an extreme count outlier then the p value and adjusted p value will be set to NA
- If a row is filtered for having a low mean normalized count, then the adjusted p value will be set to NA

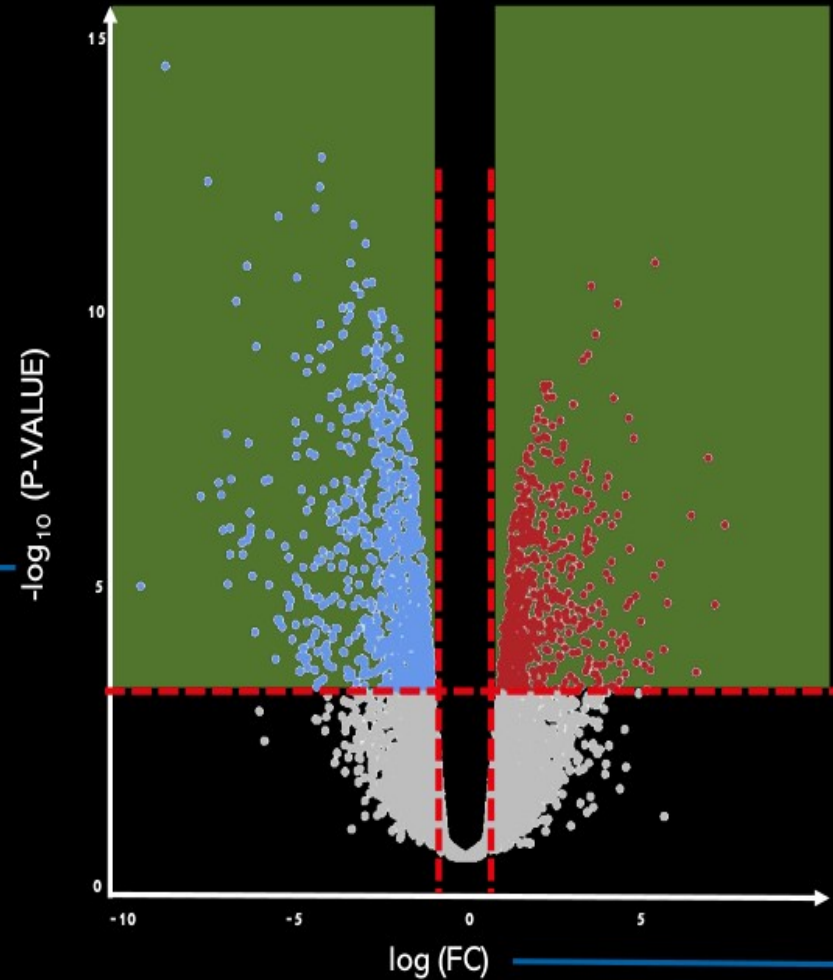


Volcano Plot

- The volcano plot is a scatter-plot of the statistical significance ($-\log_{10}$ p-values on Y-axis) against the magnitude of effect (estimated effect size or **log fold change** on X-axis)
- The intuition behind volcano plots is simple: it aims to select features that are not only significant but also carry the largest effect size.



This **function** will have high values when the P-value is low.



Fold Change is a measure describing how much a quantity changes between an original and a subsequent measurement.

	ensgene	symbol	baseMean	log2FoldChange	padj	sig	diff	delabel
40	ENSG000000003402	CFLAR	3368.7234	1.1624996	7.912546e-18	TRUE	Up	CFLAR
61	ENSG000000004799	PDK4	935.5593	2.5596753	3.787861e-02	TRUE	Up	PDK4
92	ENSG000000005249	PRKAR2B	301.3369	0.6602810	2.839846e-02	TRUE	Up	PRKAR2B
161	ENSG000000006788	MYH13	11.2024	2.9551957	4.107394e-02	TRUE	Up	MYH13
205	ENSG000000008130	NADK	1713.6320	0.8301413	1.656331e-06	TRUE	Up	NADK
209	ENSG000000008256	CYTH3	3740.9244	1.1968352	3.768491e-07	TRUE	Up	CYTH3

