

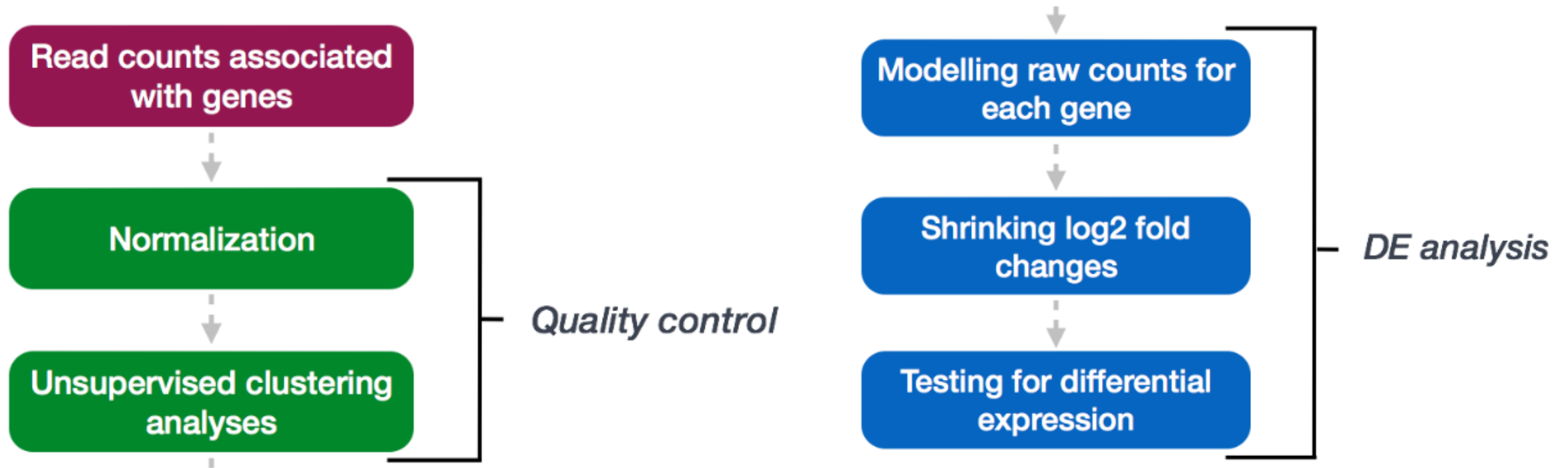
Bulk RNA Sequencing :

CD CrF(creeping fat), CD MAT(mesenteric adipose tissue), control(UC, colorectal cancer) MAT의
DEG 분석 결과

DESeq2

Estimate **variance-mean dependence** in **count data** from high-throughput **sequencing assays** and **test for differential expression** based on a **model** using the **negative binomial distribution**.

DESeq2 workflow



DESeq2 Normalization

To normalize for **sequencing depth** and **RNA composition**, DESeq2 uses the **median of ratios method**.

Step 1 : creates a pseudo-reference sample (row-wise geometric mean)

gene	sampleA	sampleB	pseudo-reference sample
EF2A	1489	906	$\text{sqrt}(1489 * 906) = \mathbf{1161.5}$
ABCD1	22	13	$\text{sqrt}(22 * 13) = \mathbf{17.7}$
...

Step 2 : calculates ratio of each sample to the reference

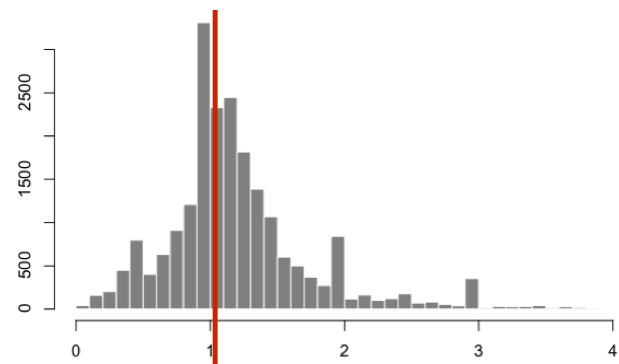
gene	sampleA	sampleB	pseudo-reference sample	ratio of sampleA/ref	ratio of sampleB/ref
EF2A	1489	906	1161.5	$1489/1161.5 = \mathbf{1.28}$	$906/1161.5 = \mathbf{0.78}$
ABCD1	22	13	16.9	$22/16.9 = \mathbf{1.30}$	$13/16.9 = \mathbf{0.77}$

DESeq2 Normalization

To normalize for sequencing depth and RNA composition, DESeq2 uses the **median of ratios method**.

Step 3 : calculate the normalization factor for each sample (size factor)

sample 1 / pseudo-reference sample

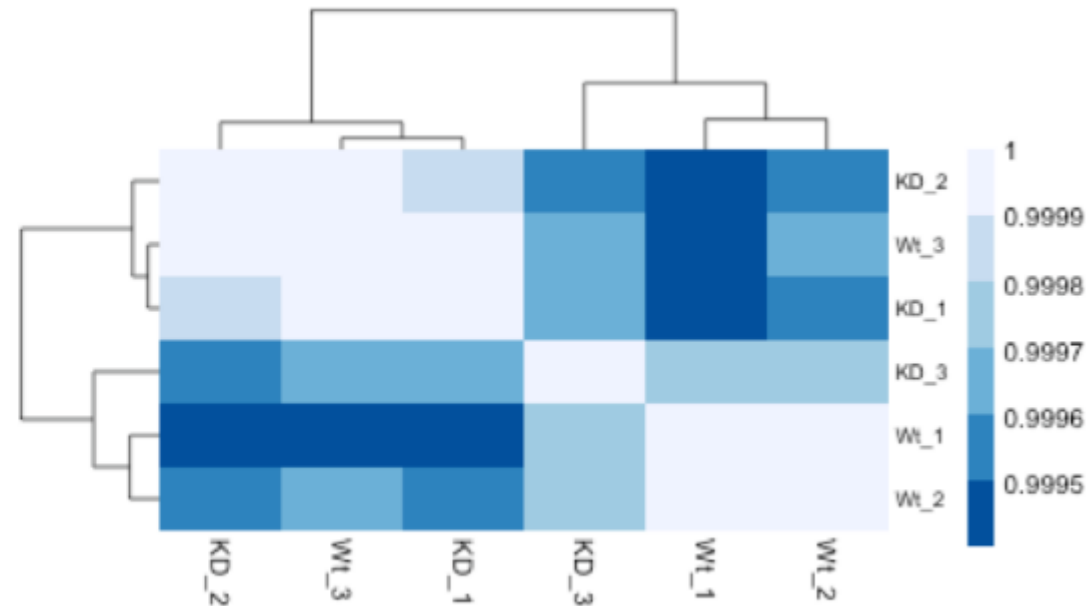
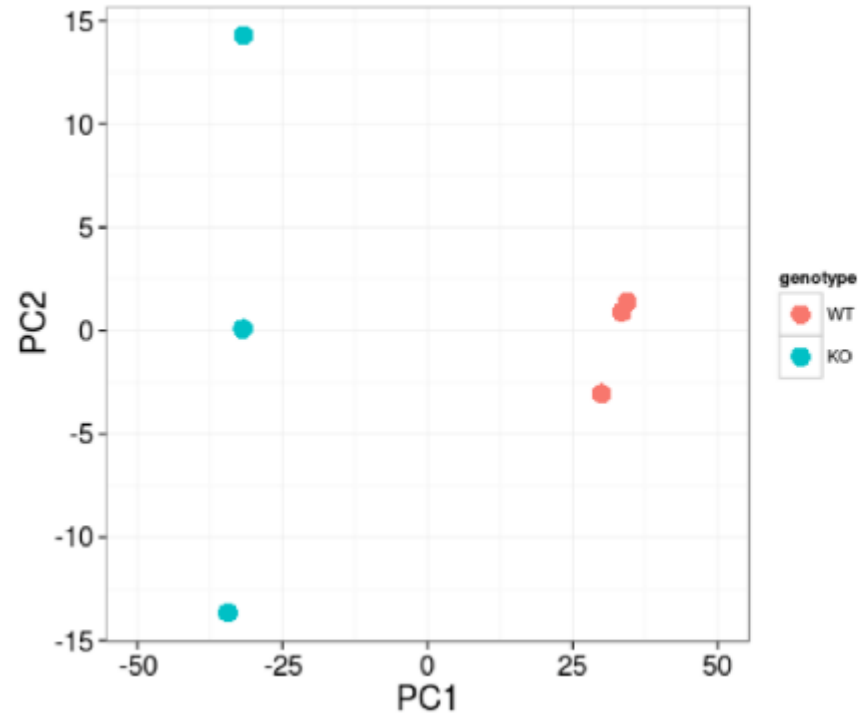


Step 4: calculate the normalized count values using the normalization factor

gene	sampleA	sampleB
EF2A	1489 / 1.3 = 1145.39	906 / 0.77 = 1176.62
ABCD1	22 / 1.3 = 16.92	13 / 0.77 = 16.88
...

DESeq2 Unsupervised clustering analyses

Sample-level QC



- To explore the **similarity** of our **samples**, we will be performing sample-level QC **using Principal Component Analysis (PCA) and hierarchical clustering Heatmap**.
- Performing sample-level QC can also **identify any sample outliers**, which may need to be explored further to determine whether they **need to be removed prior to DE analysis**.

DESeq2 Unsupervised clustering analyses

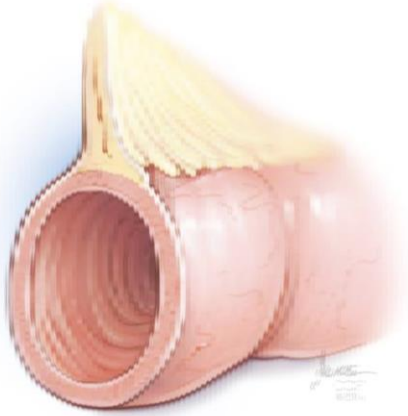
Gene-level QC

	SRR1039508	SRR1039509	SRR1039512	SRR1039513	SRR1039516	
ENSG000000000003	67	44	87	40	1138	Genes with extreme count outlier
ENSG000000000005	0	0	0	0	0	
ENSG000000000419	467	515	621	365	587	Genes with zero counts
ENSG000000000457	260	211	263	164	245	
ENSG000000000460	2	5	1	0	1	Genes with low mean normalized counts ('Independent filtering')

- Prior to DEG(differential expression genes) analysis it is beneficial to omit genes that have little or no chance of being detected as differentially expressed.
- This will increase the power to detect DEG(differential expression genes).

DESeq2 Fat tissue of Crohn's disease

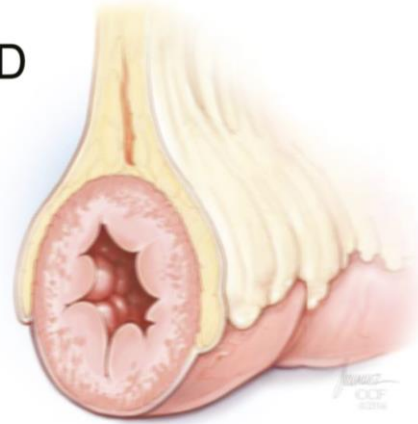
A



Healthy bowel

No inflammation
No fibrosis
No creeping fat

D



Crohn's disease

Inflammation

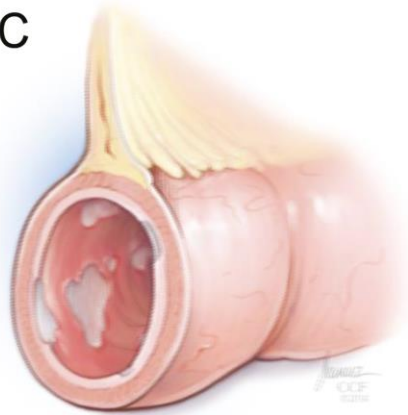
Fibrosis

Creeping fat

36 patients(23 CD, 13 UC & CRC)

- CD – CF : 23 sample
- CD – MAT : 23 sample
- Control(UC & CRC) : 13 sample

C



Ulcerative colitis

Inflammation

Fibrosis

No creeping fat

DESeq2 – DEG(Differentially expressed gene) analysis

Gene quality control

	Gene	Sample
Input data	58,204	59
(TPM \geq 0.1) \geq 30 & (raw read count \geq 6) \geq 30	20,109	59
GTF file(Human release 38)	465,027	
Chromosome 1~22, X, Y, M(mitochondria) 중 chromosome X, Y 제거	446,162	
Exon, gene, transcript 중 Exon, transcript 제거	55,290	
Protein coding, lncRNA(long non coding RNA), miRNA, rRNA 등의 40개의 type 중 Protein coding, lncRNA 남기고 제거	34,183	
Raw read count, TPM 조건 충족 gene & GTF file gene	17,268	59
Final data	17,268	59

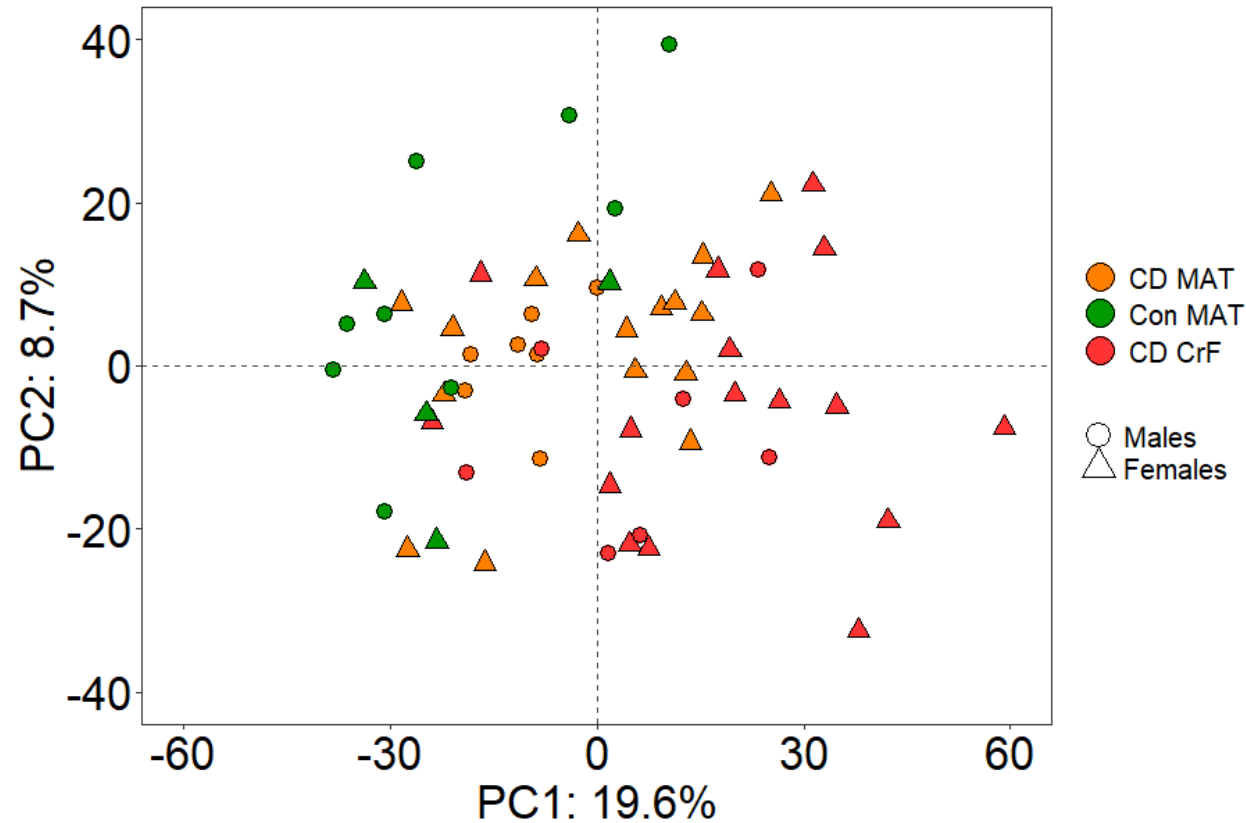
Study Design formula :

design= ~sex + sample type

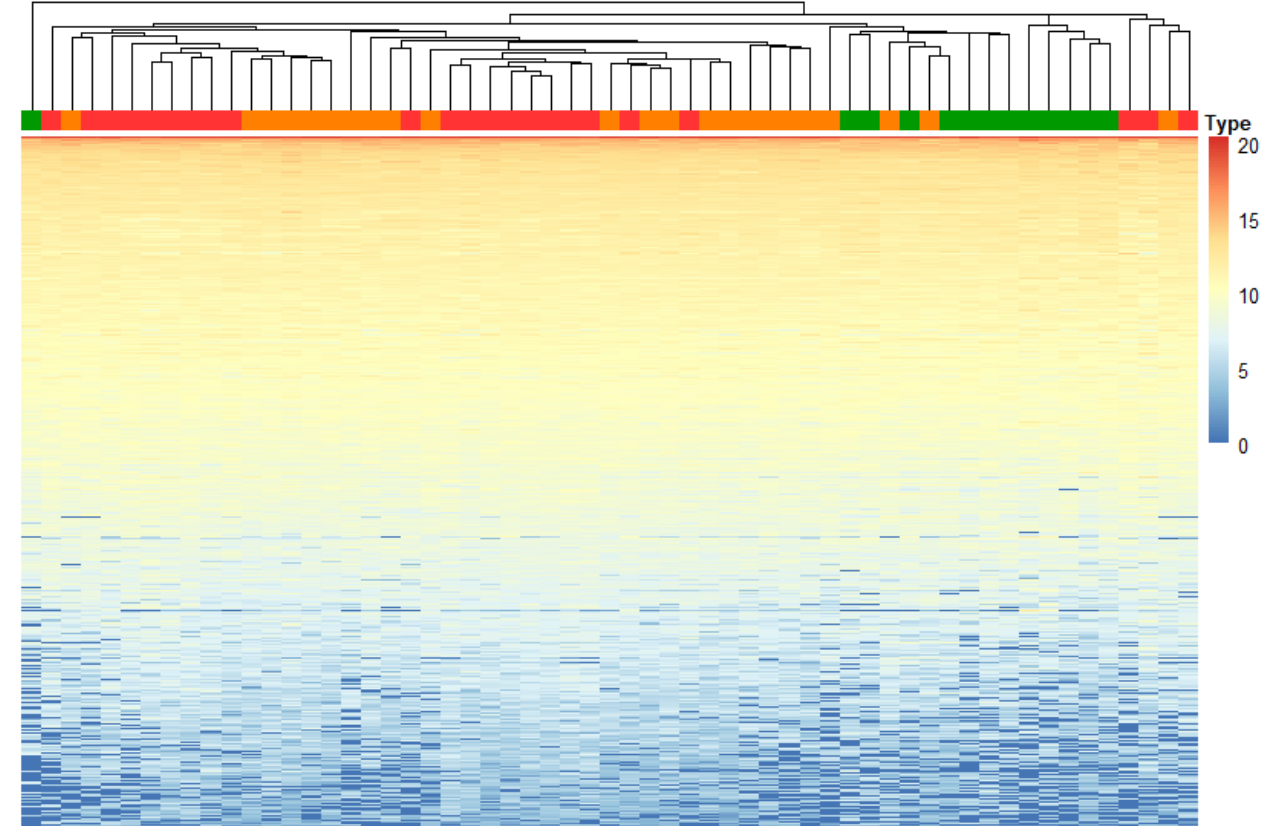
CD CrF vs Control MAT, CD MAT vs Control, CD CrF vs CD MAT

DESeq2 – DEG(Differentially expressed gene) analysis

59 sample, 17,268 gene



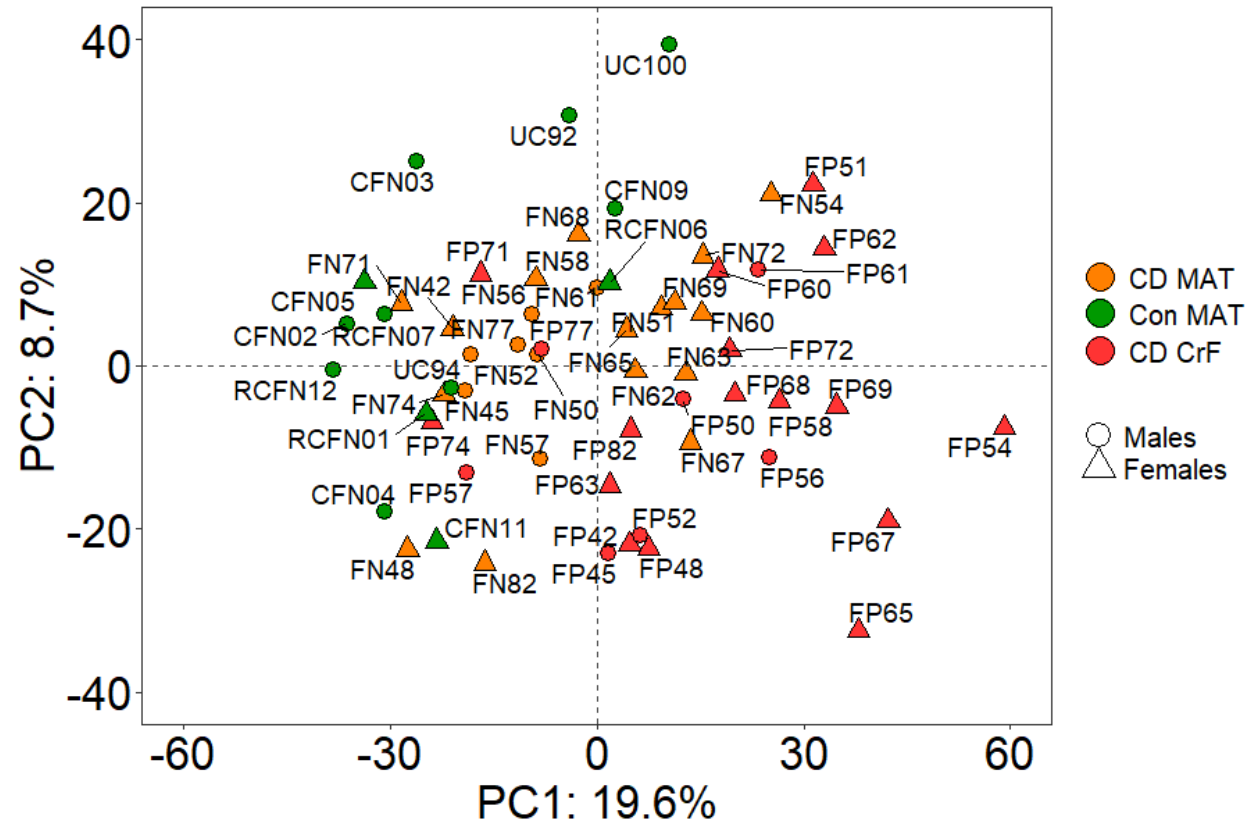
➤ PCA(principal component analysis)



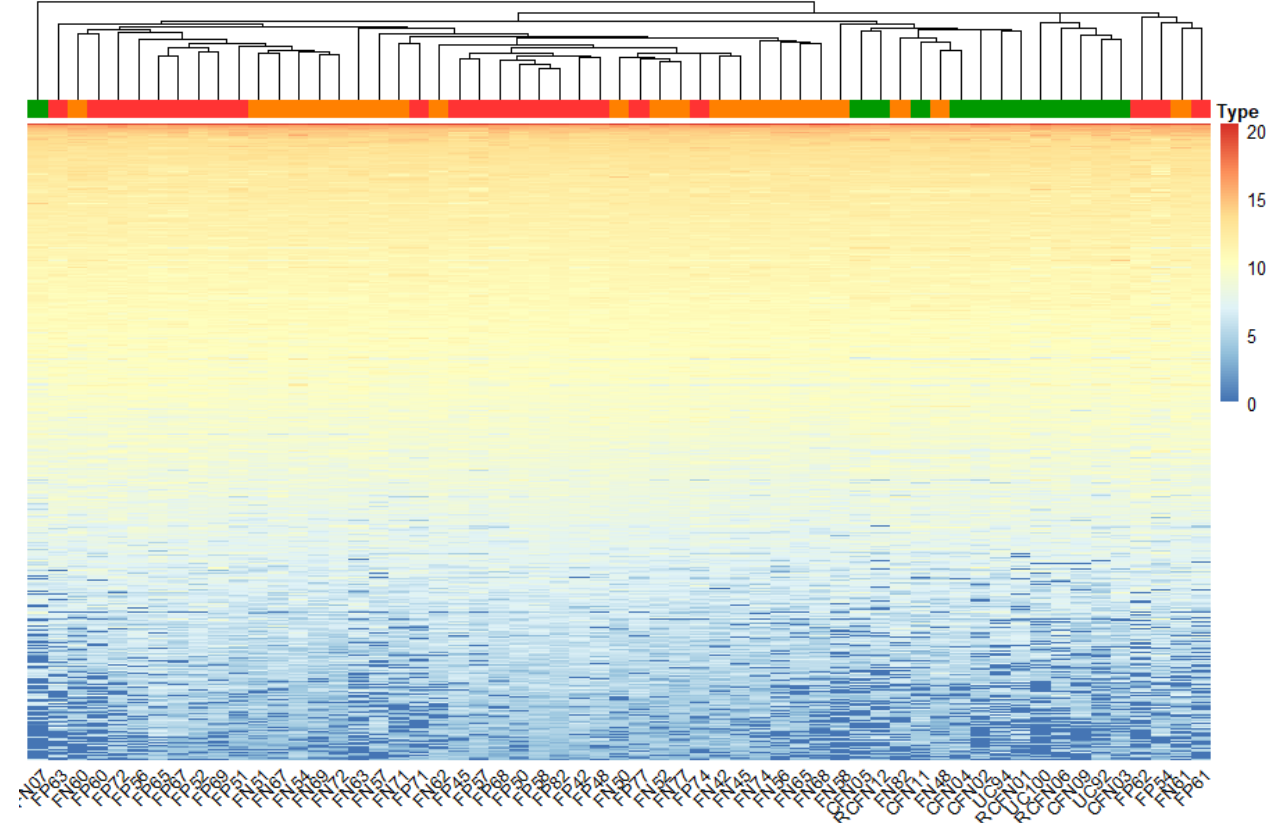
➤ Hierarchical clustering heatmap

DESeq2 – DEG(Differentially expressed gene) analysis

59 sample, 17,268 gene



➤ PCA(principal component analysis)

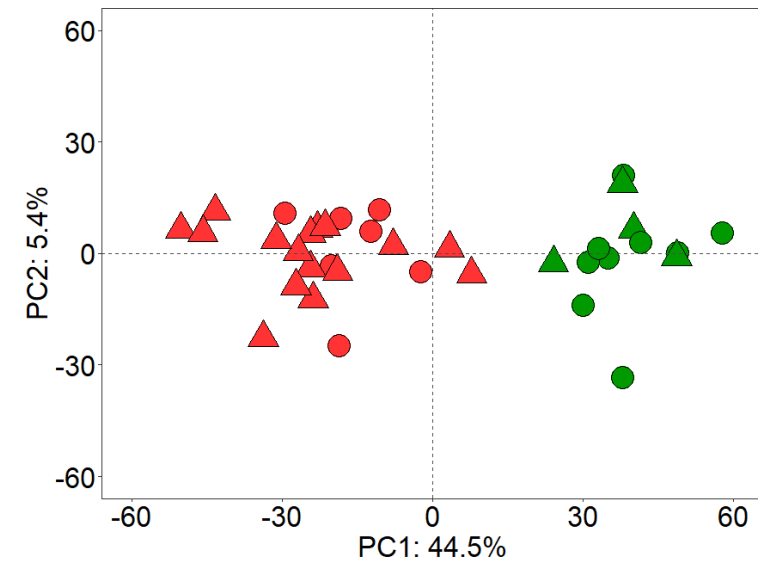


➤ Hierarchical clustering heatmap

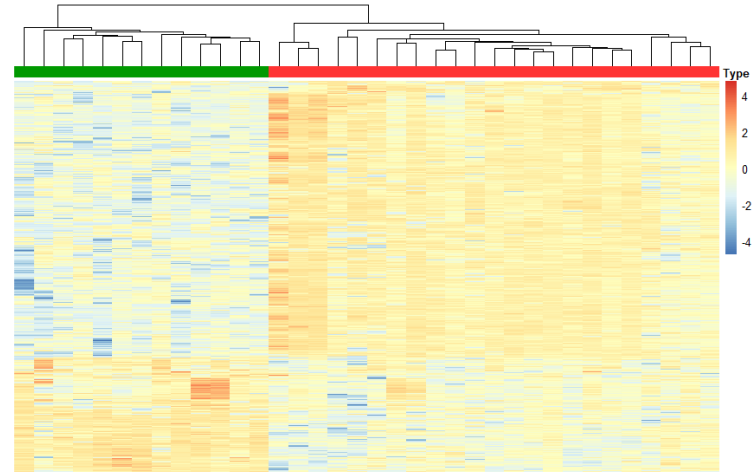
DESeq2 - DEG analysis : CD CrF(23) vs control MAT(13)

Adjusted p-value < 0.05, | Log₂ fold change | > 1.5

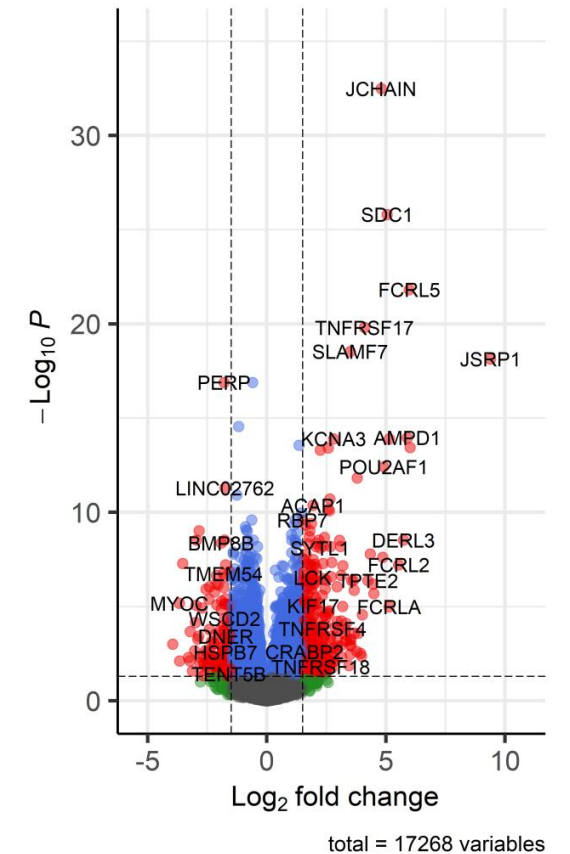
DEG : **463 gene** (Down-regulation : 136 gene, Up-regulation: 327 gene)



➤ PCA
(principal component analysis)



➤ Hierarchical clustering heatmap

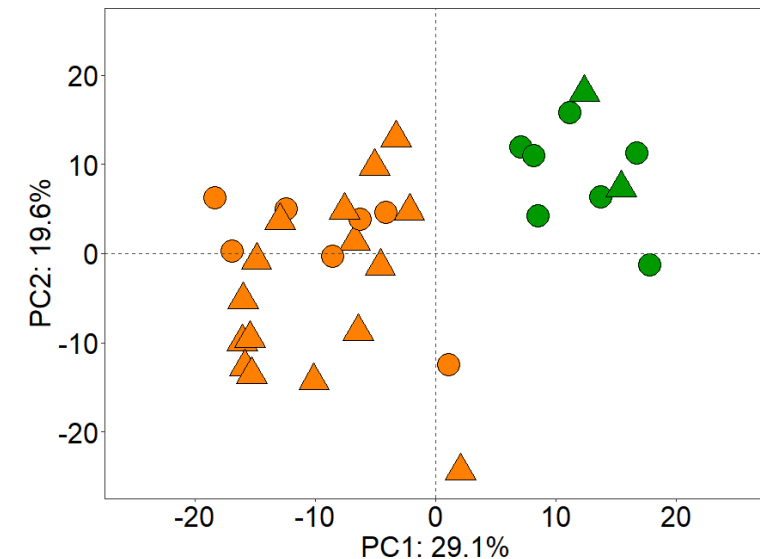


➤ Volcano plot

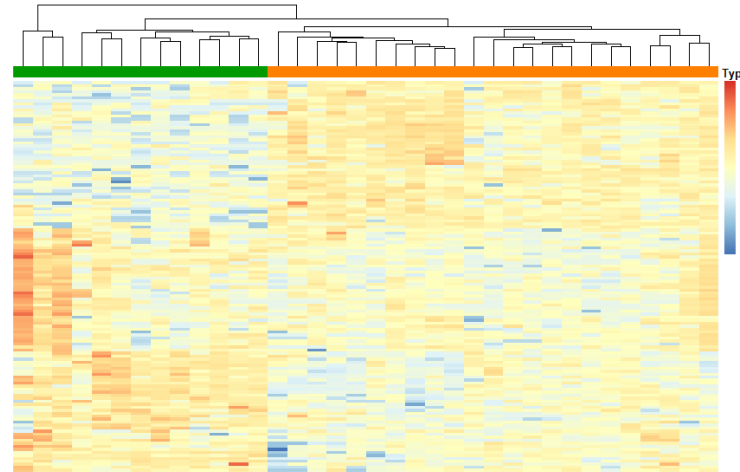
DESeq2 - DEG analysis : CD MAT(23) vs control MAT(13)

Adjusted p-value < 0.05, | Log₂ fold change | > 1.5

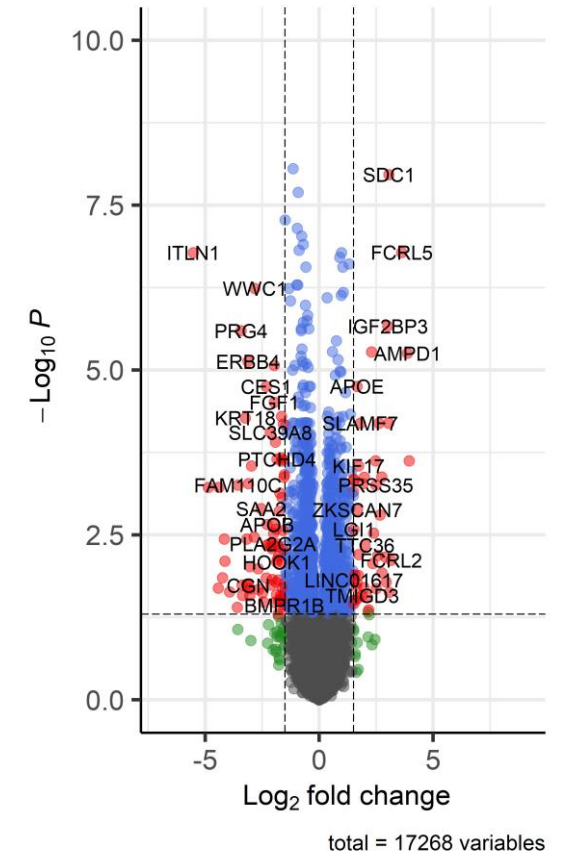
DEG : **130 gene** (Down-regulation : 81 gene, Up-regulation : 49 gene)



➤ PCA
(principal component analysis)



➤ Hierarchical clustering heatmap

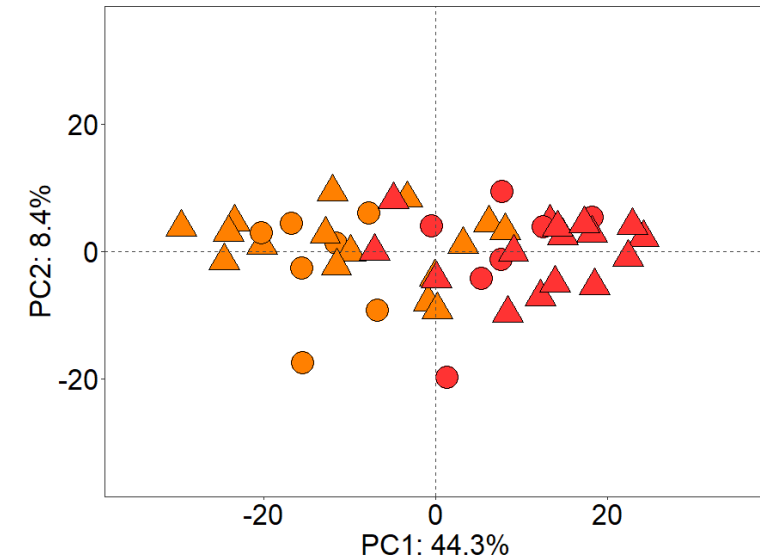


➤ Volcano plot

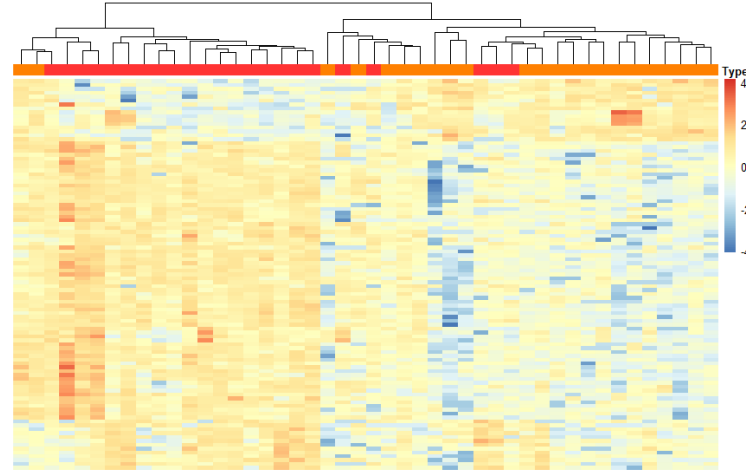
DESeq2 - DEG analysis : CD CrF(23) vs CD MAT(23)

Adjusted p-value < 0.05, | Log₂ fold change | > 1.5

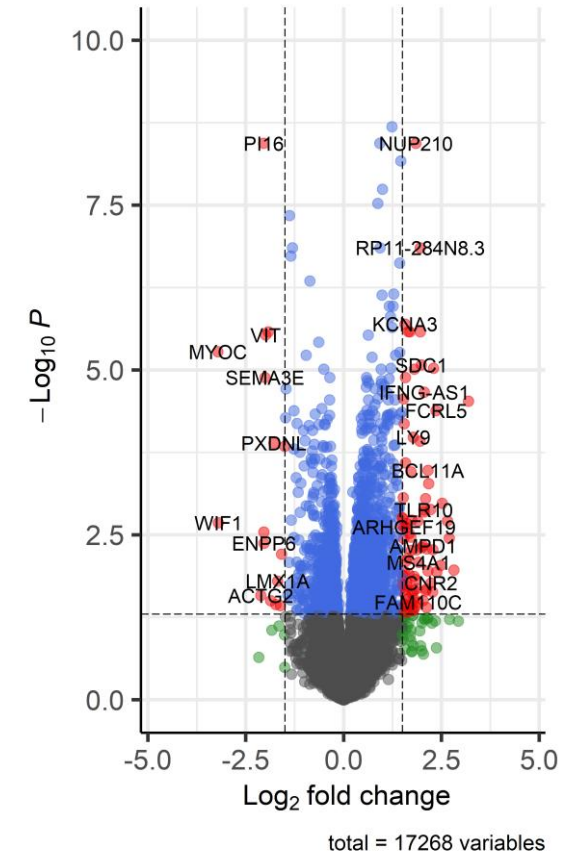
DEG : **101 gene** (Down-regulation : 16 gene, Up-regulation : 85 gene)



➤ PCA
(principal component analysis)

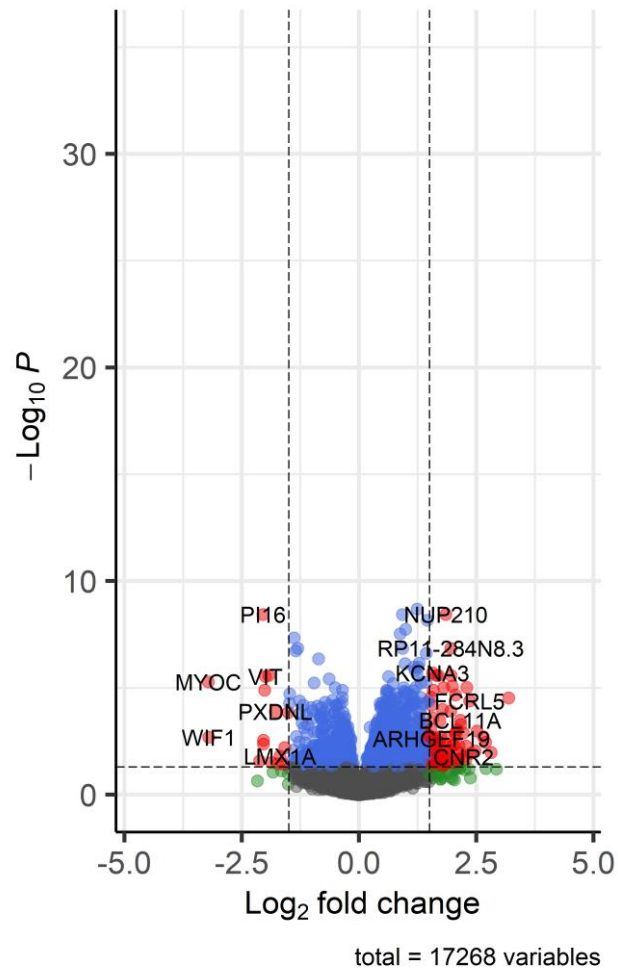


➤ Hierarchical clustering heatmap

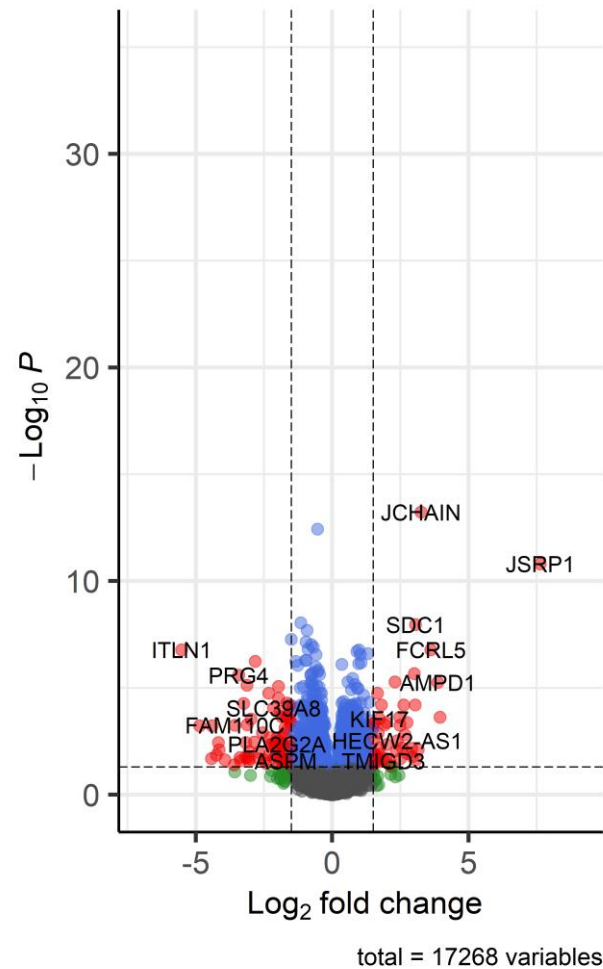


➤ Volcano plot

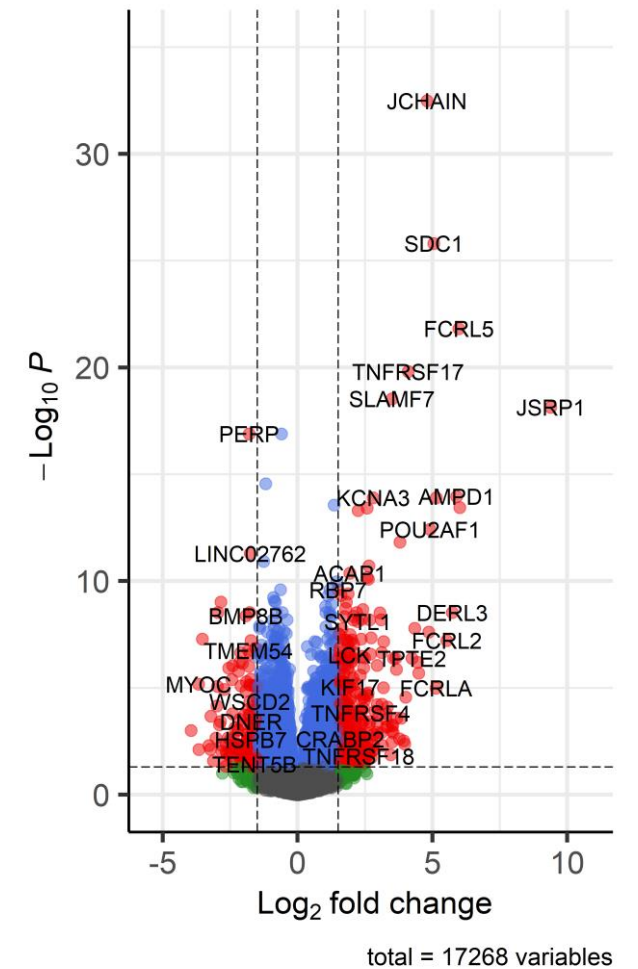
DESeq2 - DEG analysis volcano plot



➤ CD CrF(23) vs CD MAT(23)
DEG : 101 gene

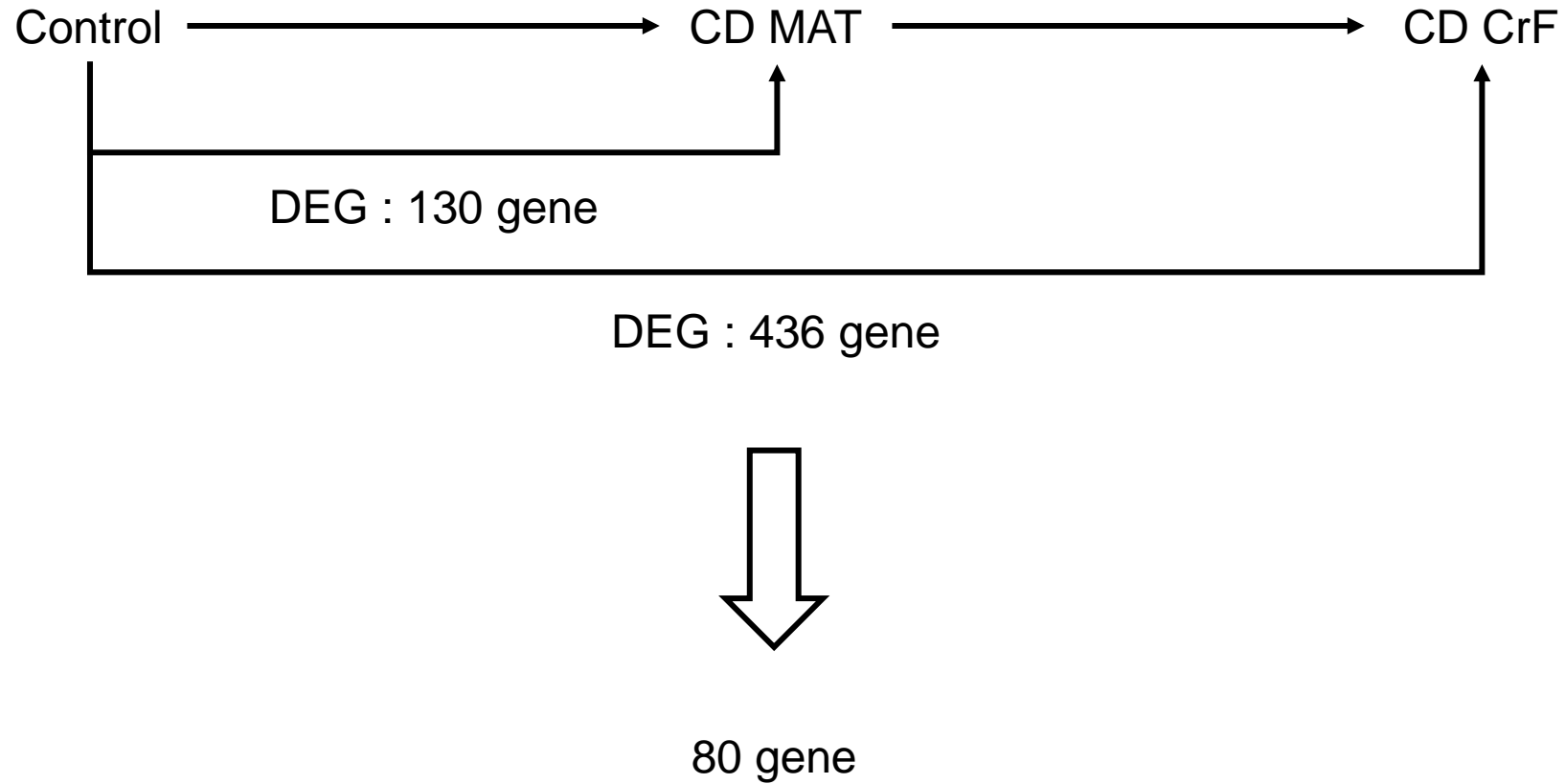


➤ CD MAT(23) vs control MAT(13)
DEG : 130 gene



➤ CD CrF(23) vs control MAT(13)
DEG : 463 gene

DESeq2 - DEG analysis conclusion

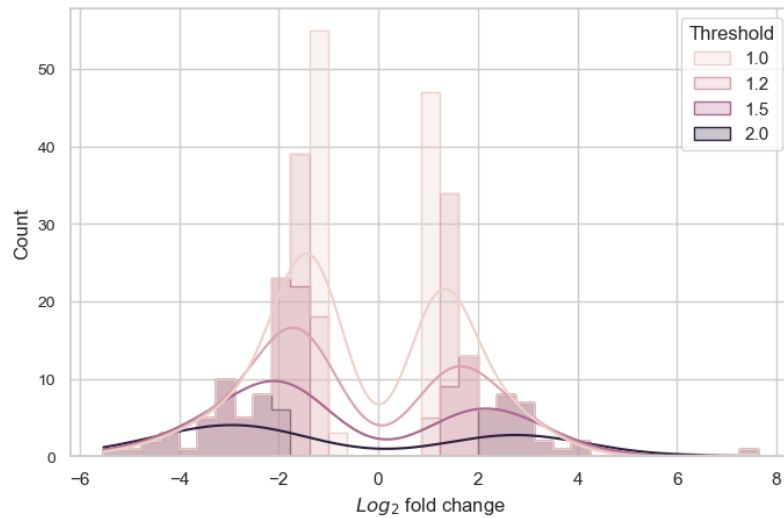


DESeq2 - DEG analysis : various Log₂ fold change threshold

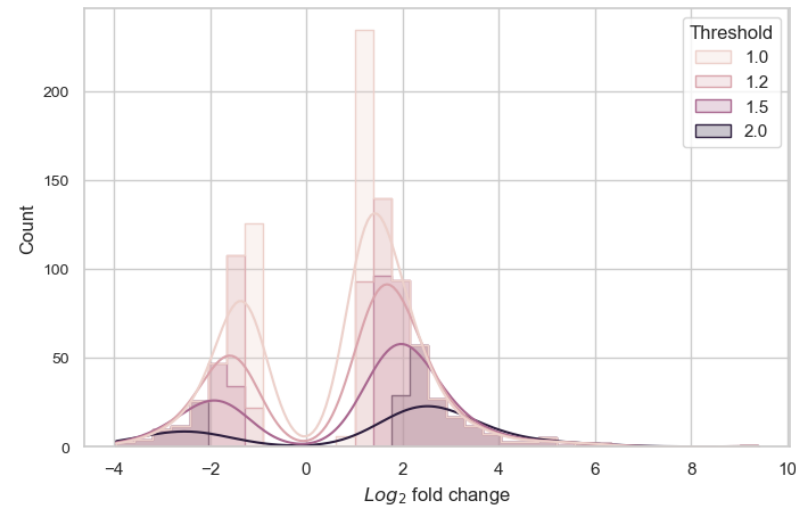
	CD MAT vs control (up/down)	CD CrF vs control (up/down)	CD CrF vs CD MAT (up/down)
Log ₂ fold change > 1.0	277(121/156)	948(612/336)	297(236/61)
Log ₂ fold change > 1.2	195(79/116)	696(464/232)	188(155/33)
Log ₂ fold change > 1.5	130(49/81)	463(327/136)	101(85/16)
Log ₂ fold change > 2.0	69(27/42)	221(166/55)	33(26/7)

Chengen Wang et al., *ScienceDirect*, 2020
Xuanwen Bao et al., *Cancer Immunol Immunother*, 2021
Lianghe Yu et al., *Frontiers in Immunology*, 2022
https://www.researchgate.net/post/What_is_an_ideal_threshold_for_log2Fold_Change, 2023-08-02

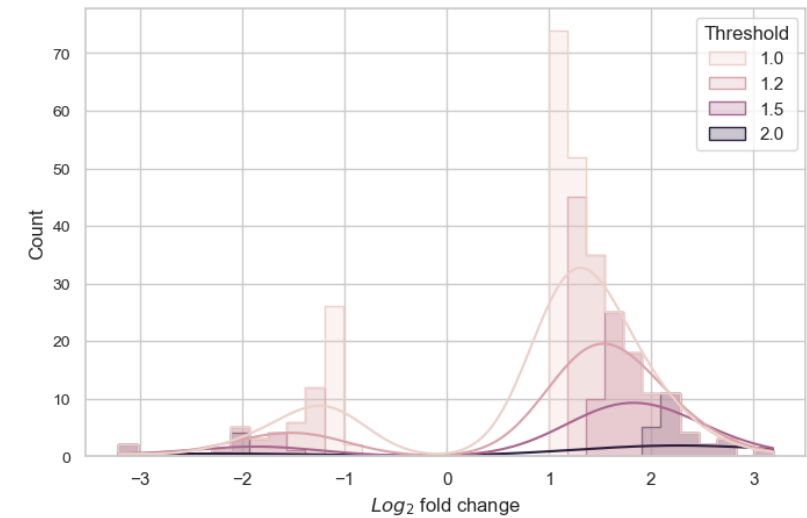
DESeq2 - DEG analysis : various Log_2 fold change threshold



➤ CD vs Control

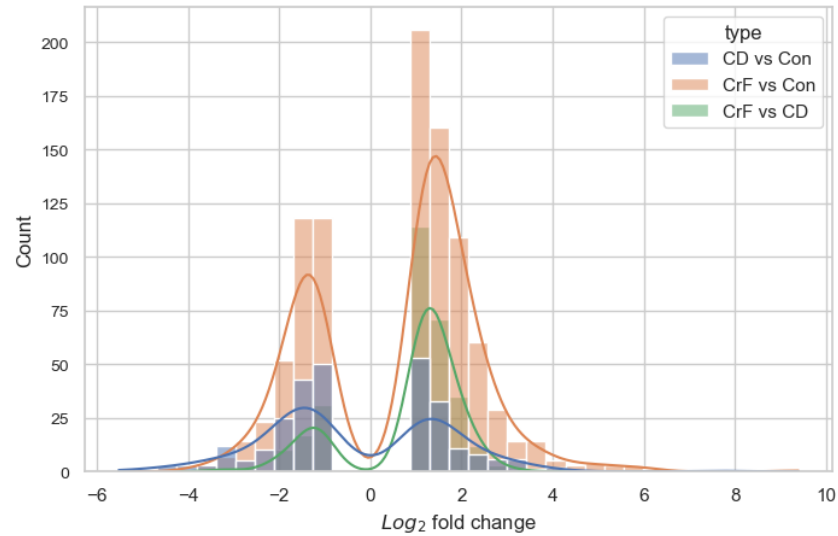


➤ CrF vs Control

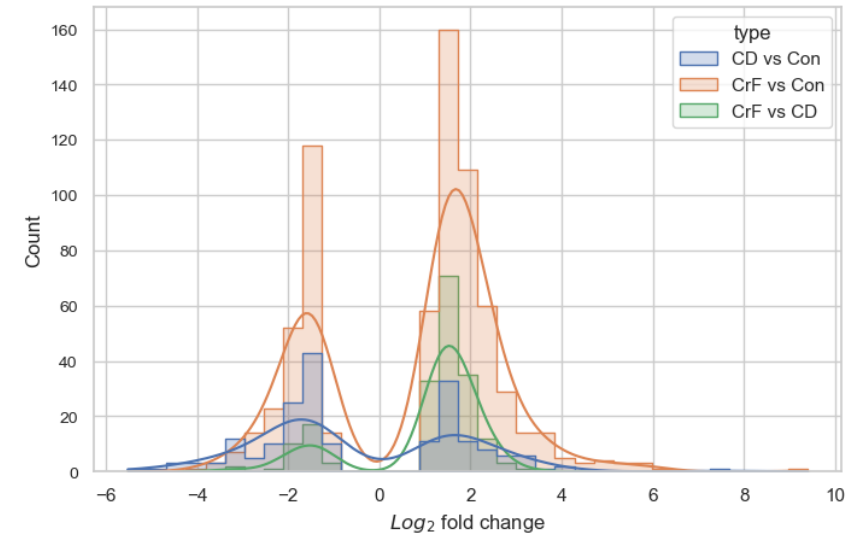


➤ CrF vs CD

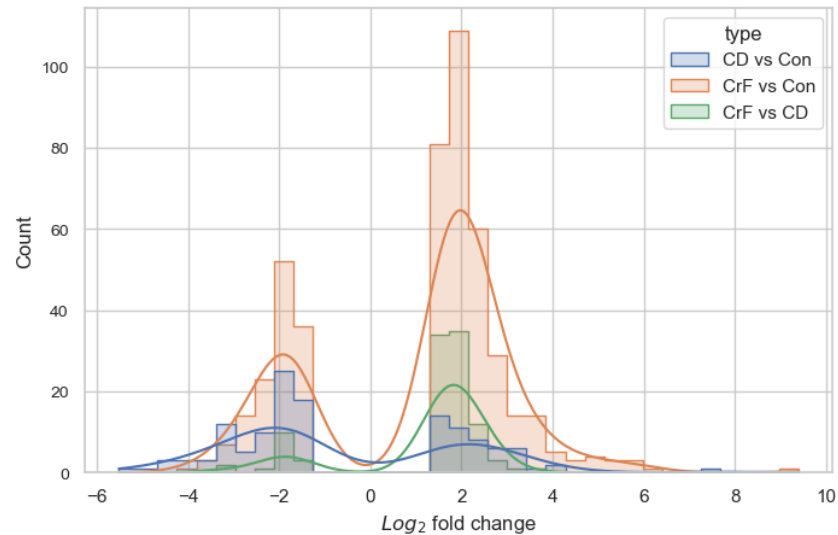
DESeq2 - DEG analysis : various Log_2 fold change threshold



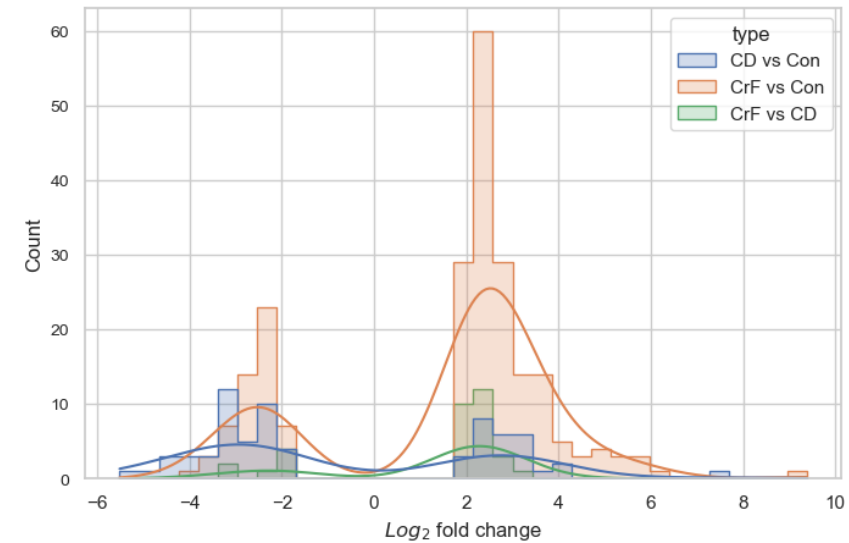
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

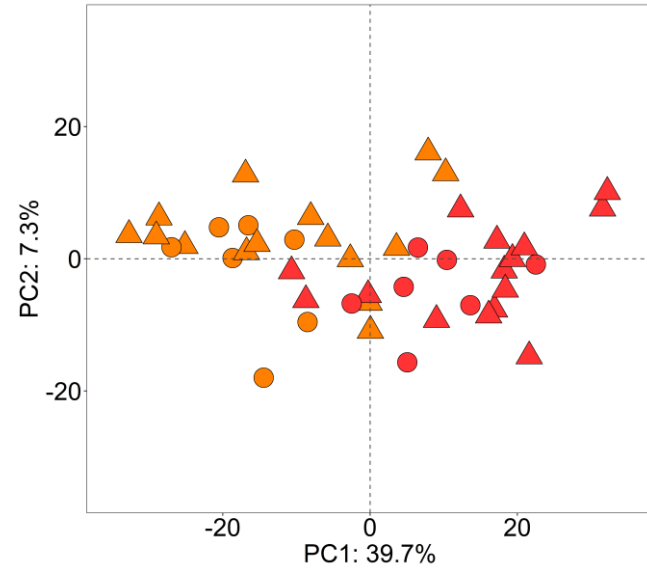


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

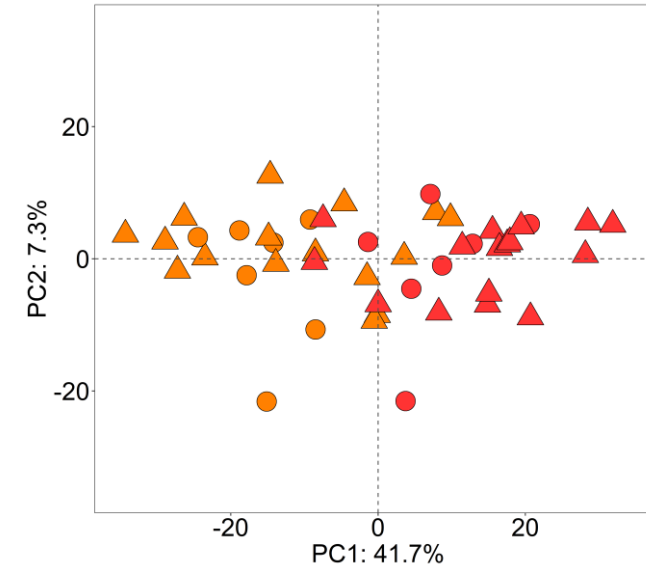


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

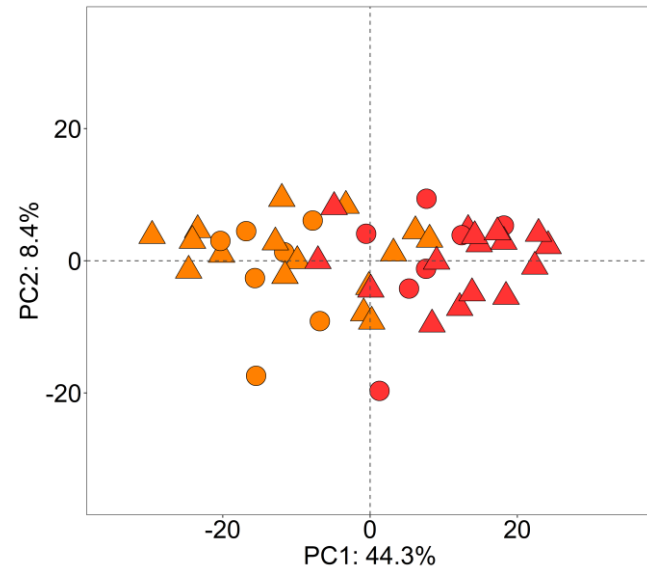
DESeq2 - DEG analysis : CD CrF(23) vs CD MAT(23) PCA



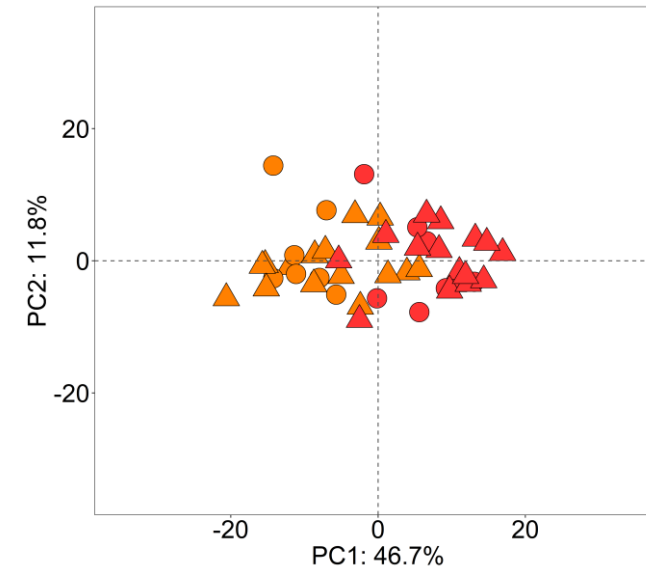
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

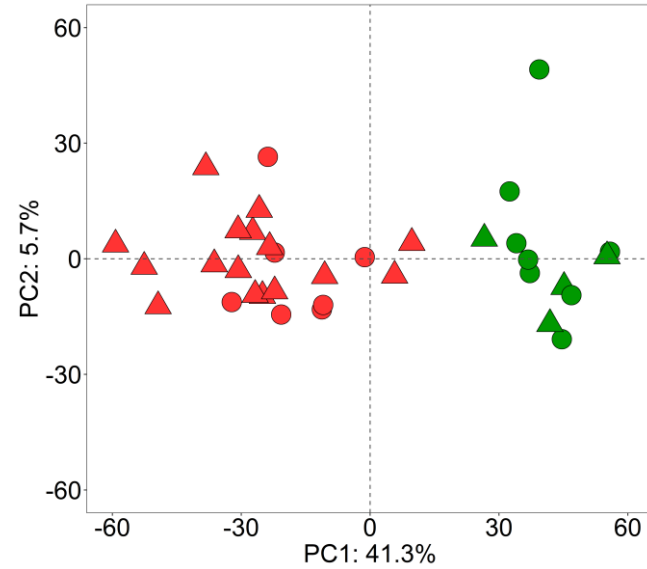


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

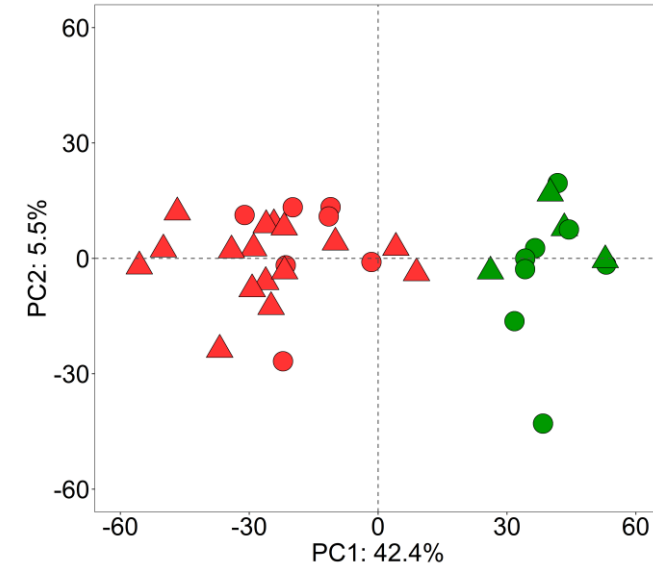


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

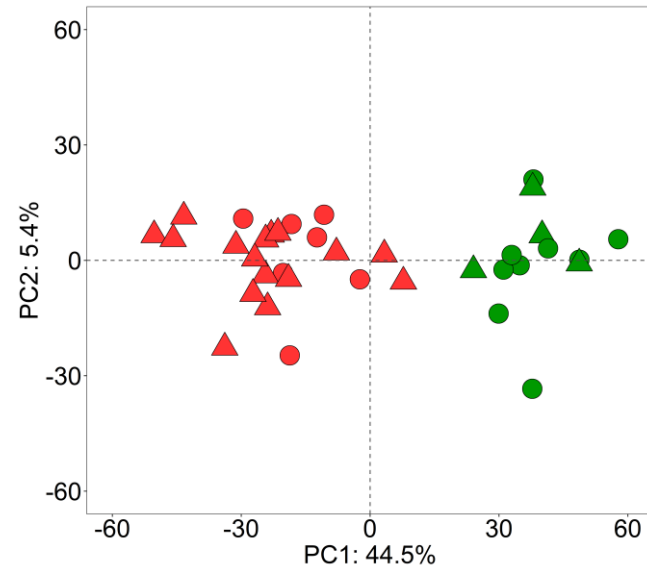
DESeq2 - DEG analysis : CD CrF(23) vs control(13) PCA



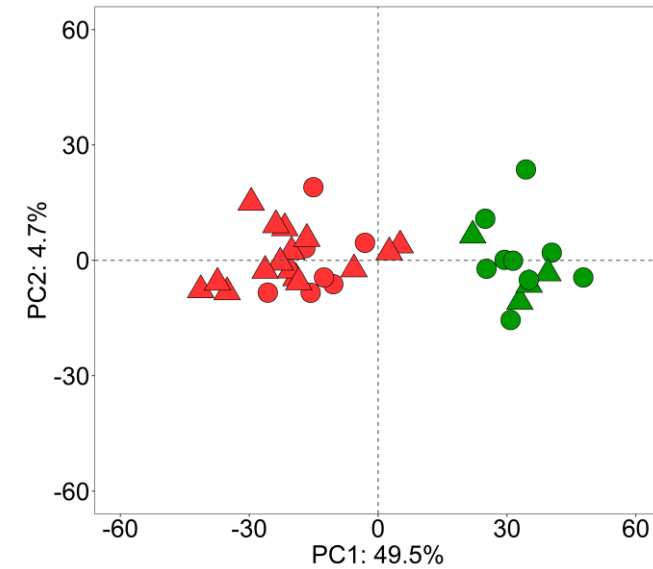
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

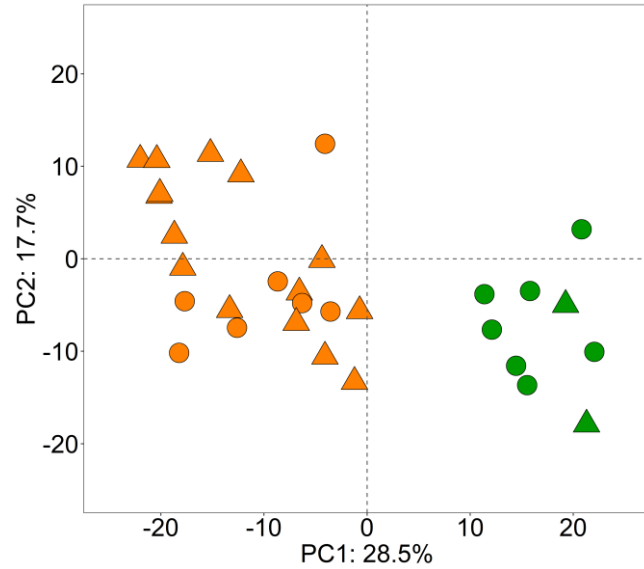


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

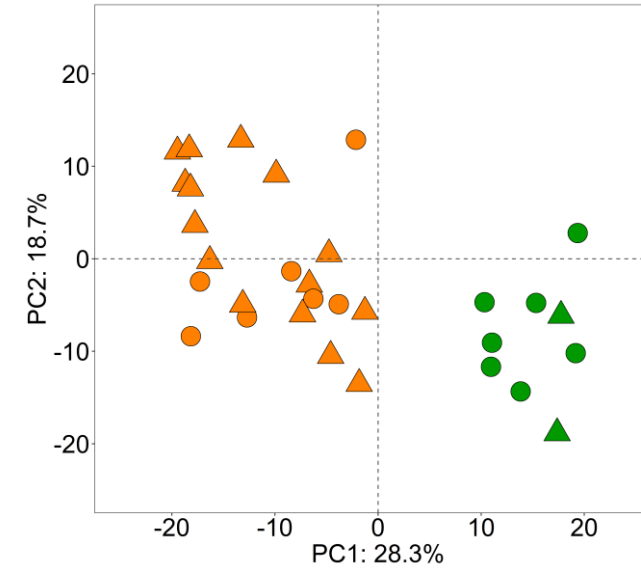


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

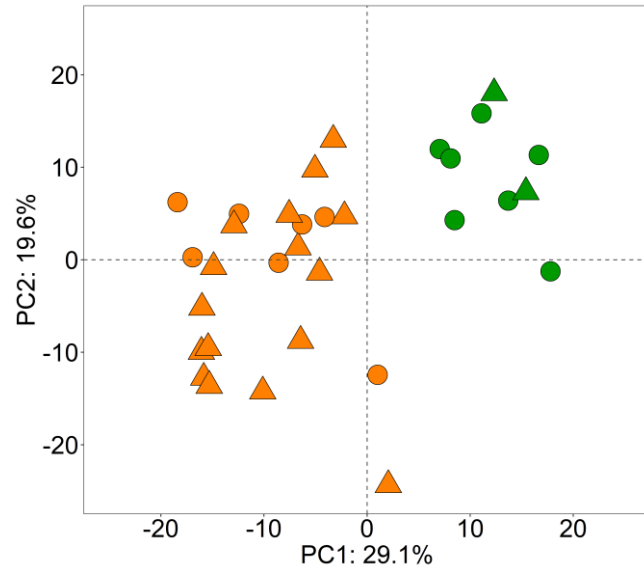
DESeq2 - DEG analysis : CD MAT(23) vs control(13) PCA



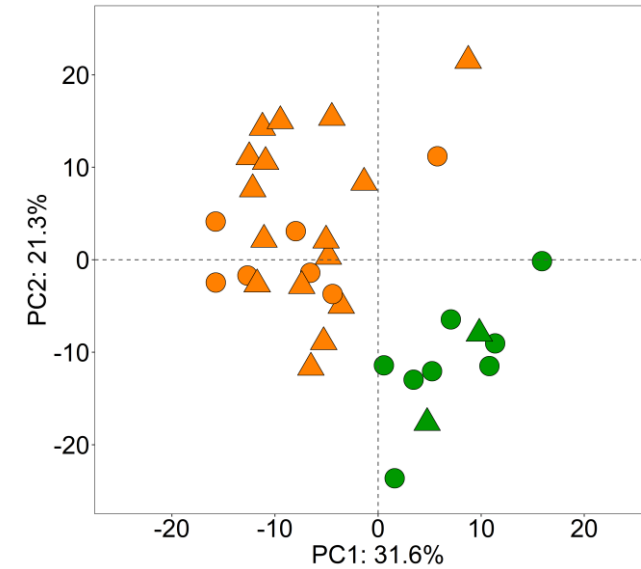
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

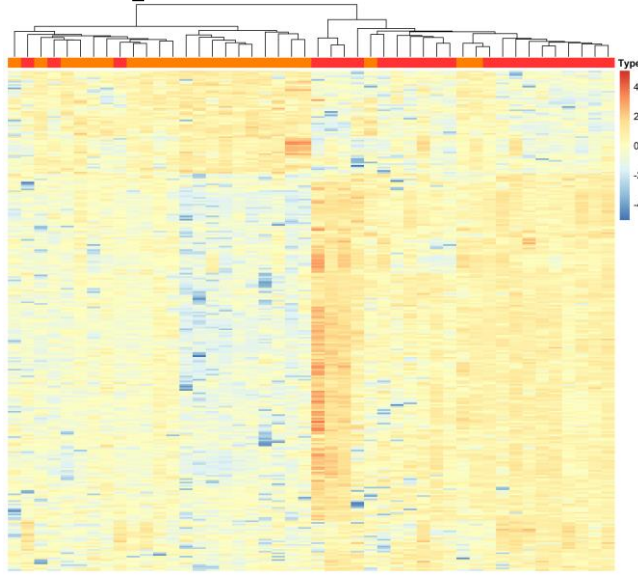


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

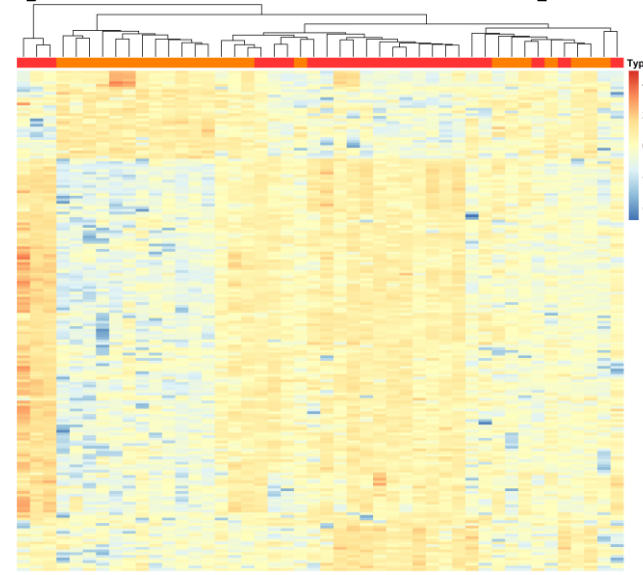


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

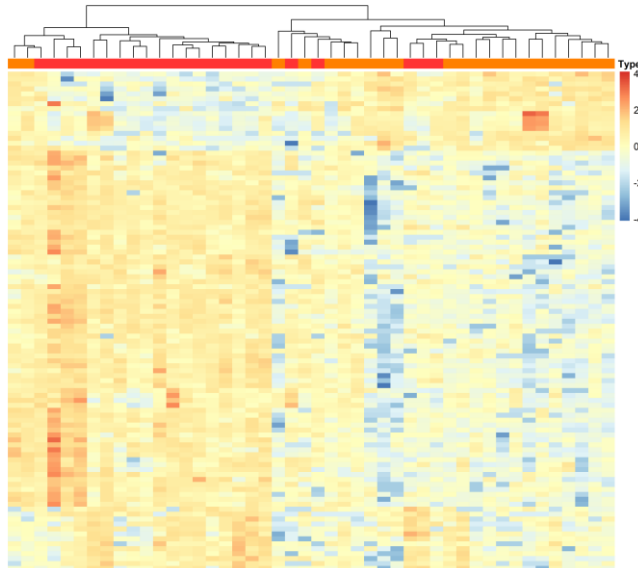
DESeq2 - DEG analysis : CD CrF(23) vs CD MAT(23) heatmap



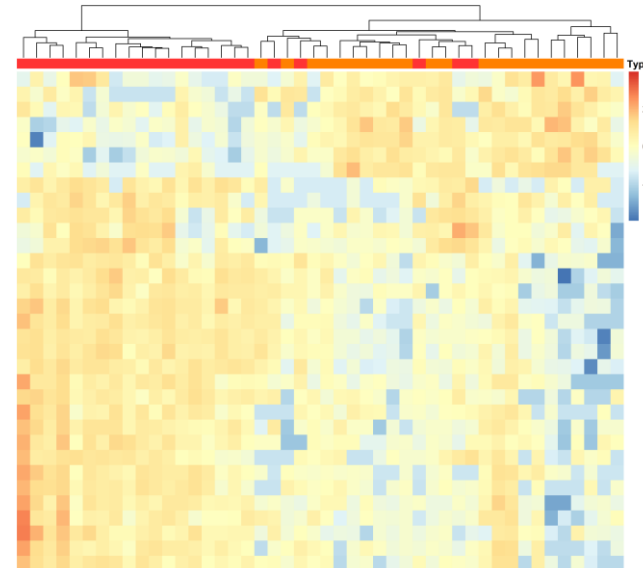
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

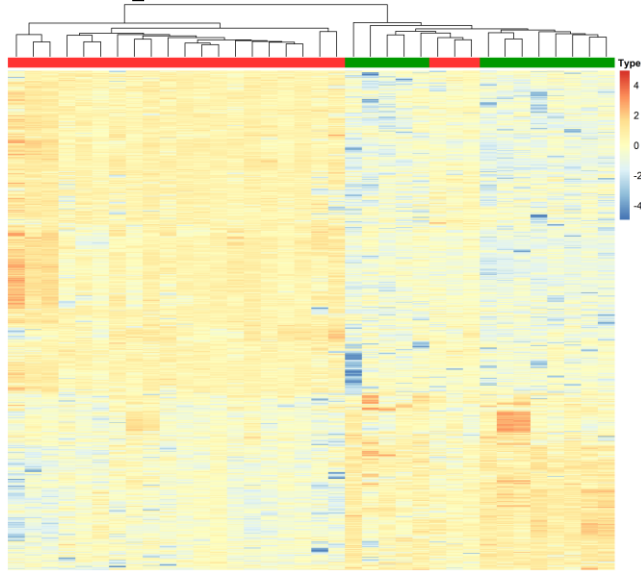


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

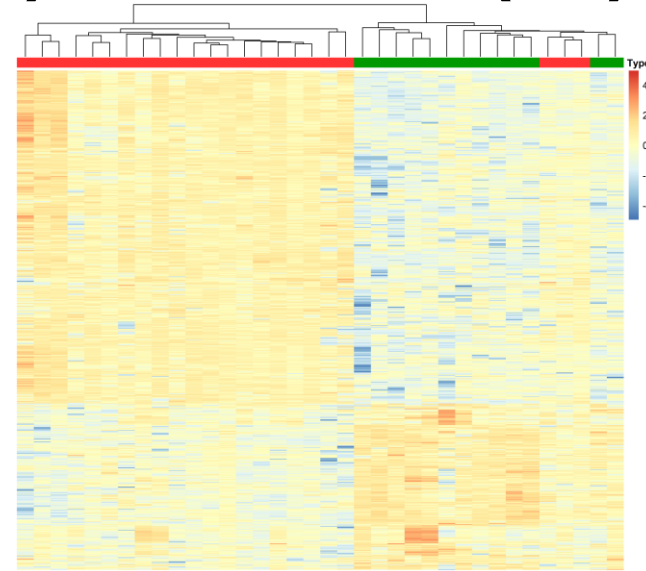


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

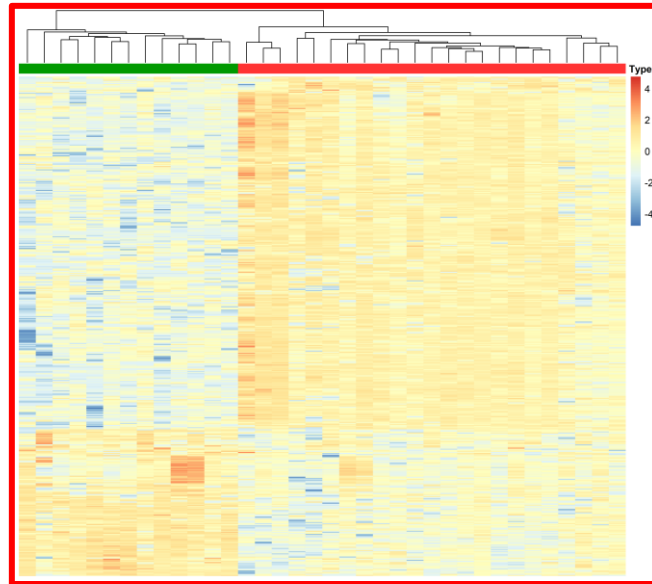
DESeq2 - DEG analysis : CD CrF(23) vs control(13) heatmap



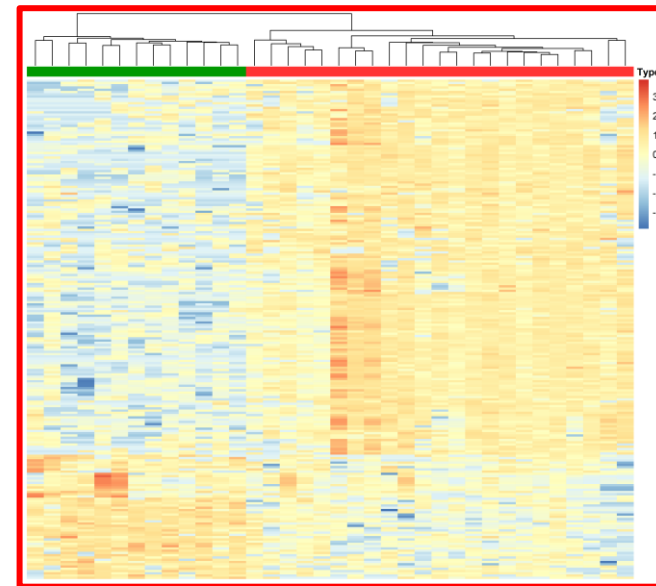
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$

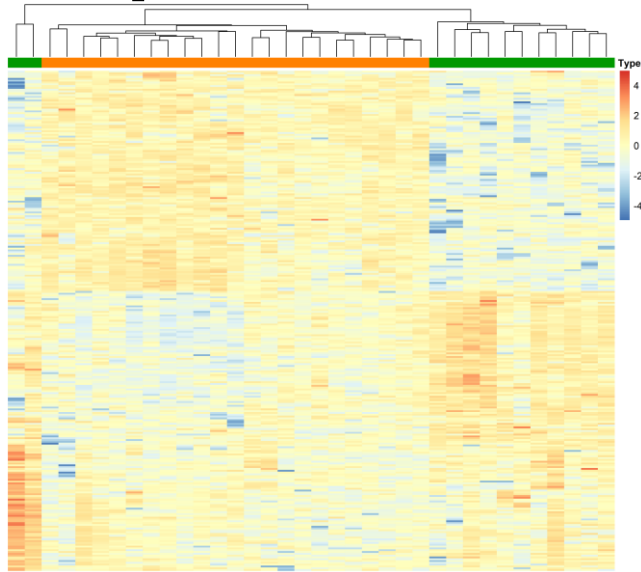


➤ $|\text{Log}_2 \text{ fold change}| > 1.5$

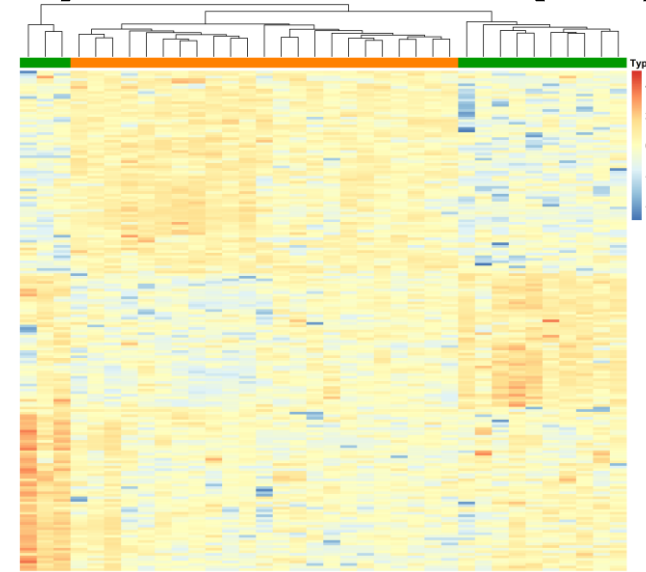


➤ $|\text{Log}_2 \text{ fold change}| > 2.0$

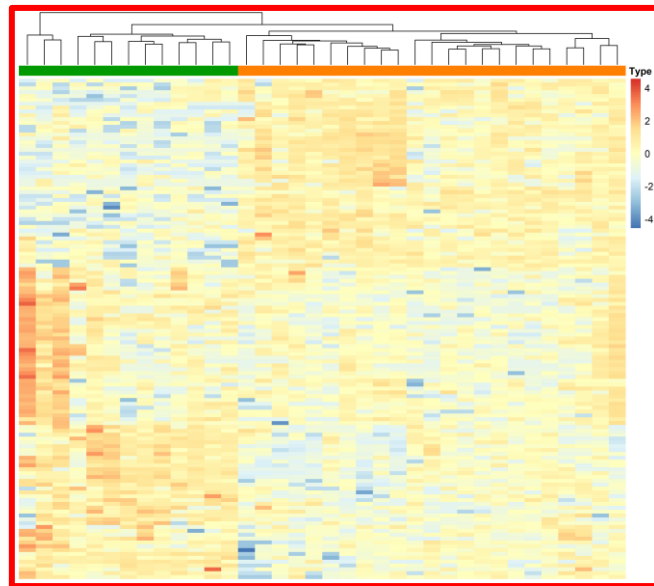
DESeq2 - DEG analysis : CD MAT(23) vs control(13) heatmap



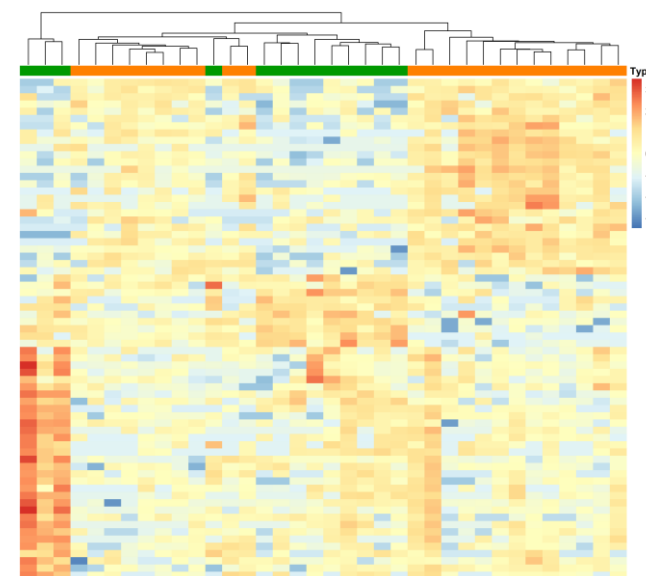
➤ $|\text{Log}_2 \text{ fold change}| > 1.0$



➤ $|\text{Log}_2 \text{ fold change}| > 1.2$



➤ $|\text{Log}_2 \text{ fold change}| > 1.5$



➤ $|\text{Log}_2 \text{ fold change}| > 2.0$