check **Ammarnas brain sex-biased gene expression issue** (exceptional more
number of female-biased genes, formed a distinct second clouds in volcano plot)
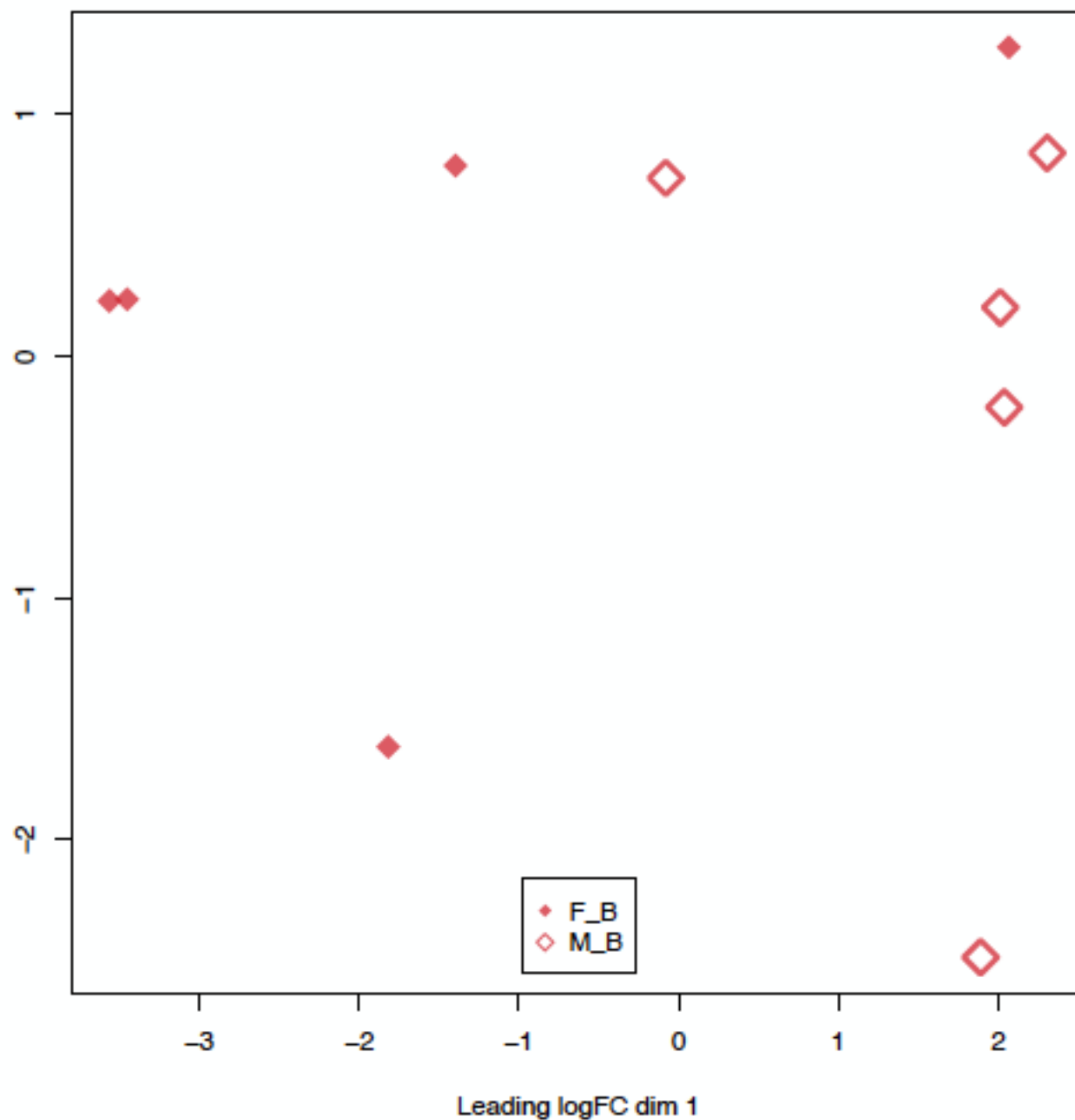
1. re-run analysis

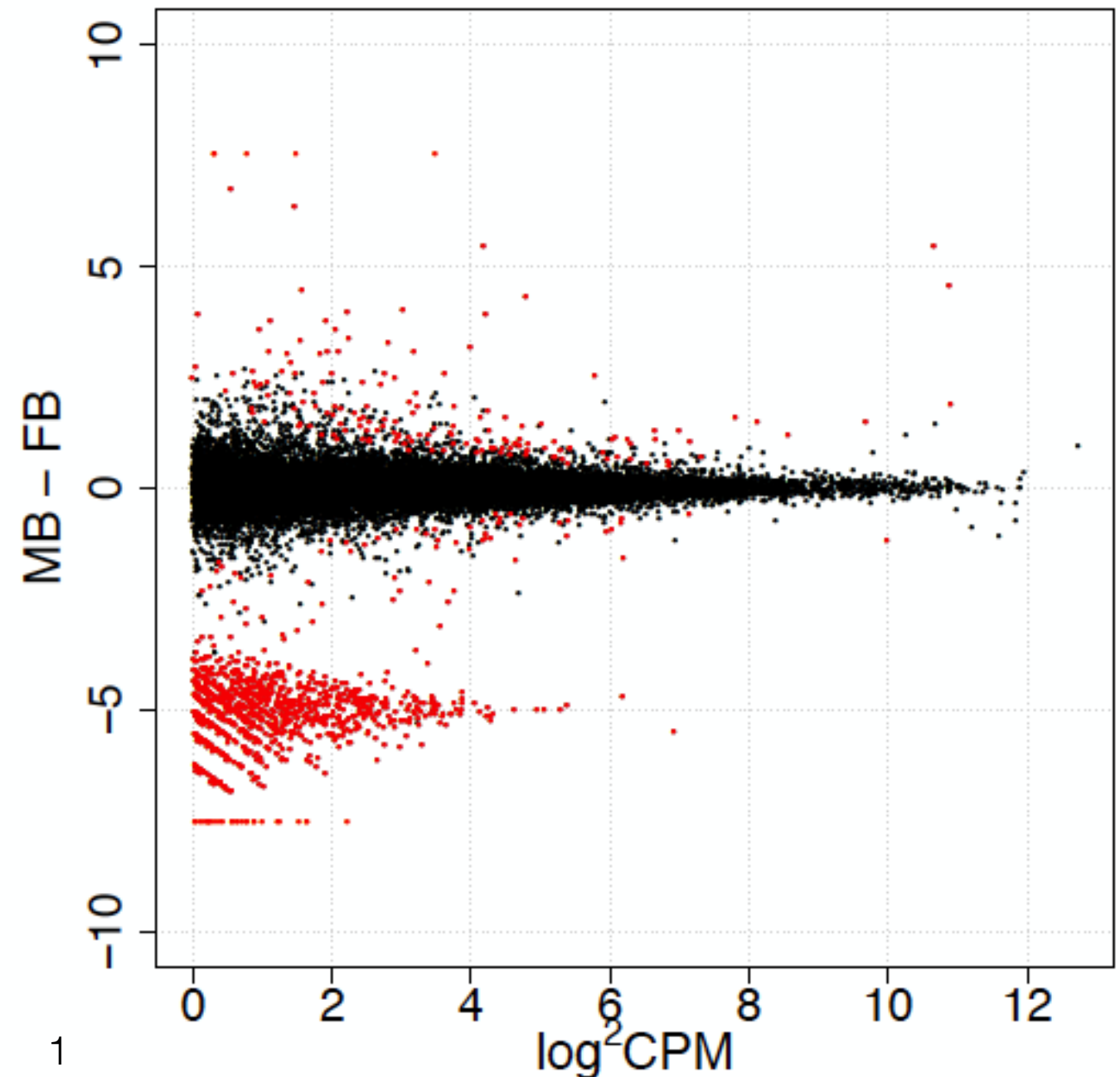the results remain the same as before (for number of SB genes and the second clouds)

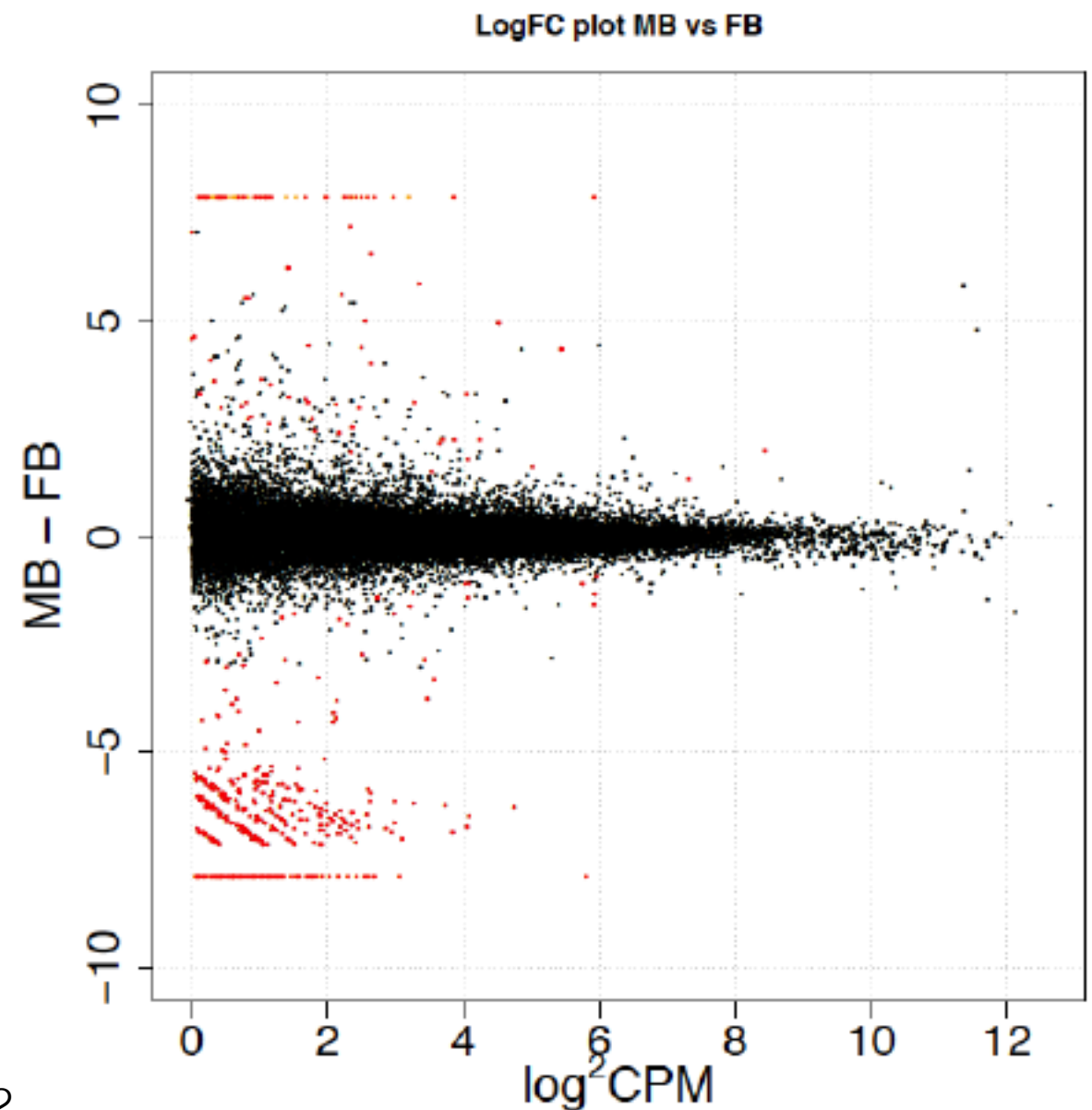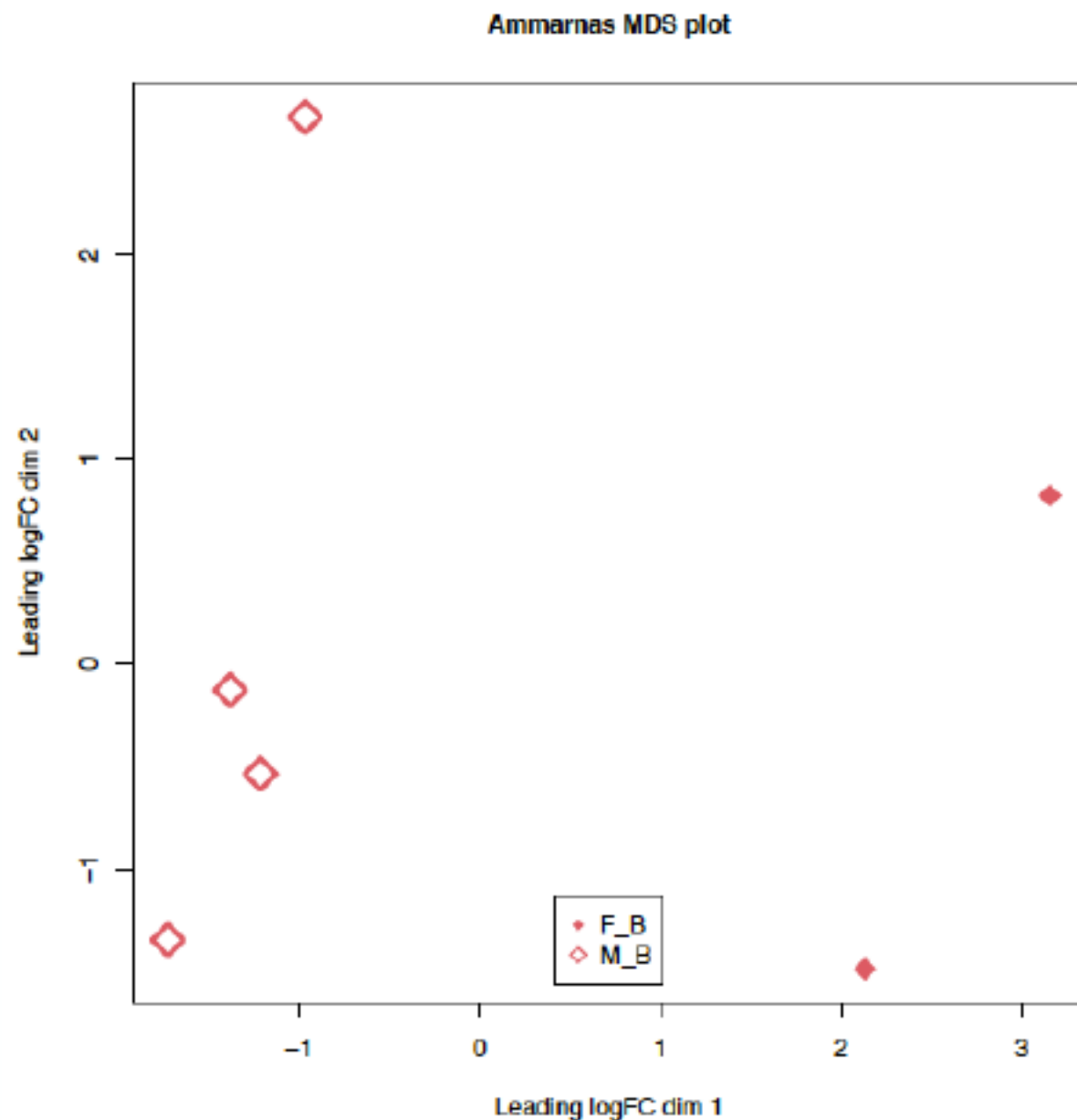FB:1072   (FDR<0.05)
MB:166

Ammarnas MDS plot

LogFC plot MB vs FB

check **Ammarnas brain sex-biased gene expression issue** (exceptional more number of female-biased genes, formed a distinct second clouds in volcano plot)

2. remove seem-to-be problematic samples, (remain 2 females and 4 males) the extra female-biased gene cloud remains.

Ammarnas MDS plot

LogFC plot MB vs FB

2

3. GO term analysis:
suggests no weird/possible contaminated-tissue specific genes.

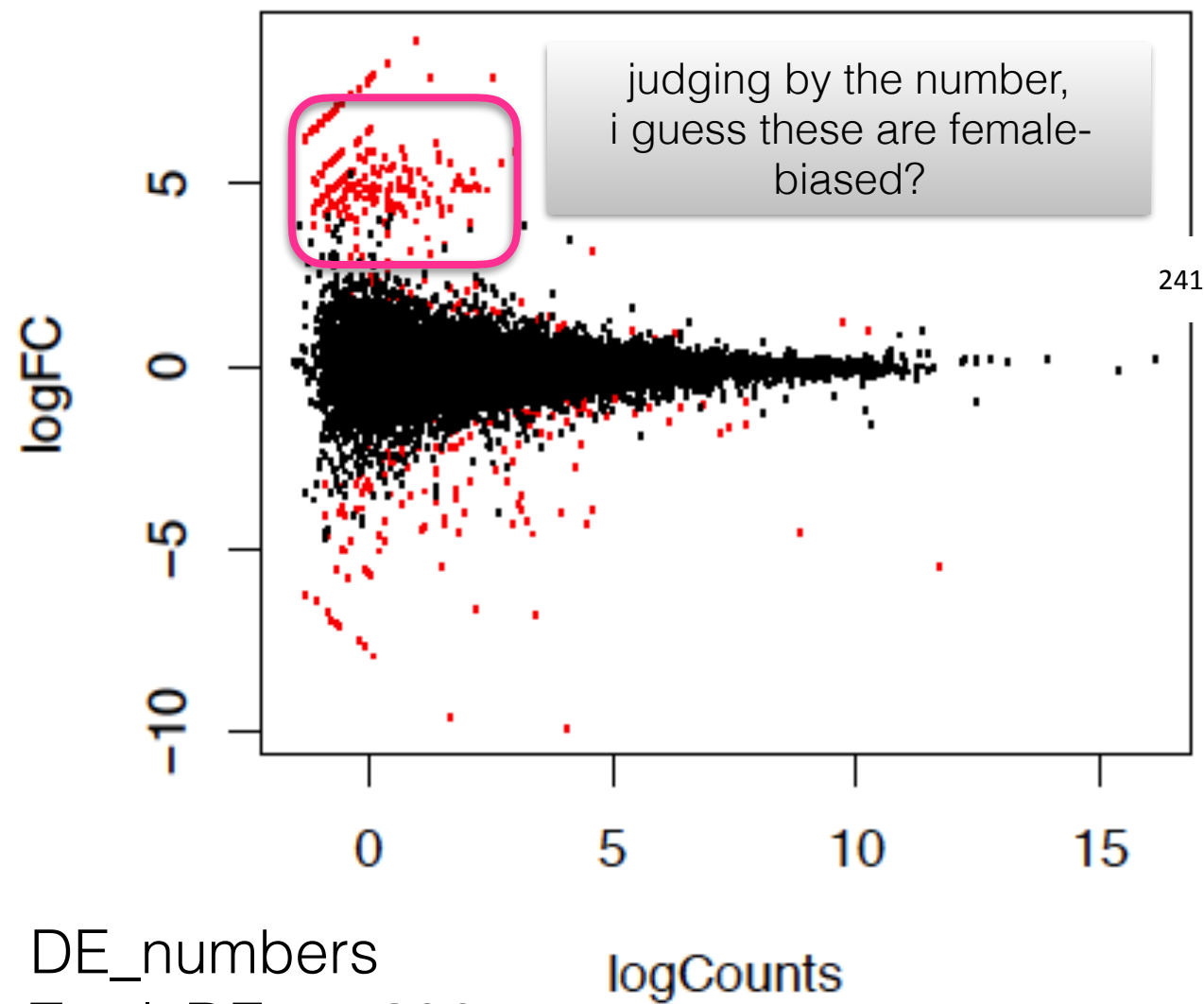| GO category	GO	Term	Annotated	Significant	Expected	topGO (Fisher)	classic (Fisher)
BP	GO:0019752	carboxylic acid metabolic process	13	3	0.49	0.008	0.008
BP	GO:0043043	peptide biosynthetic process	19	3	0.72	0.030	0.025

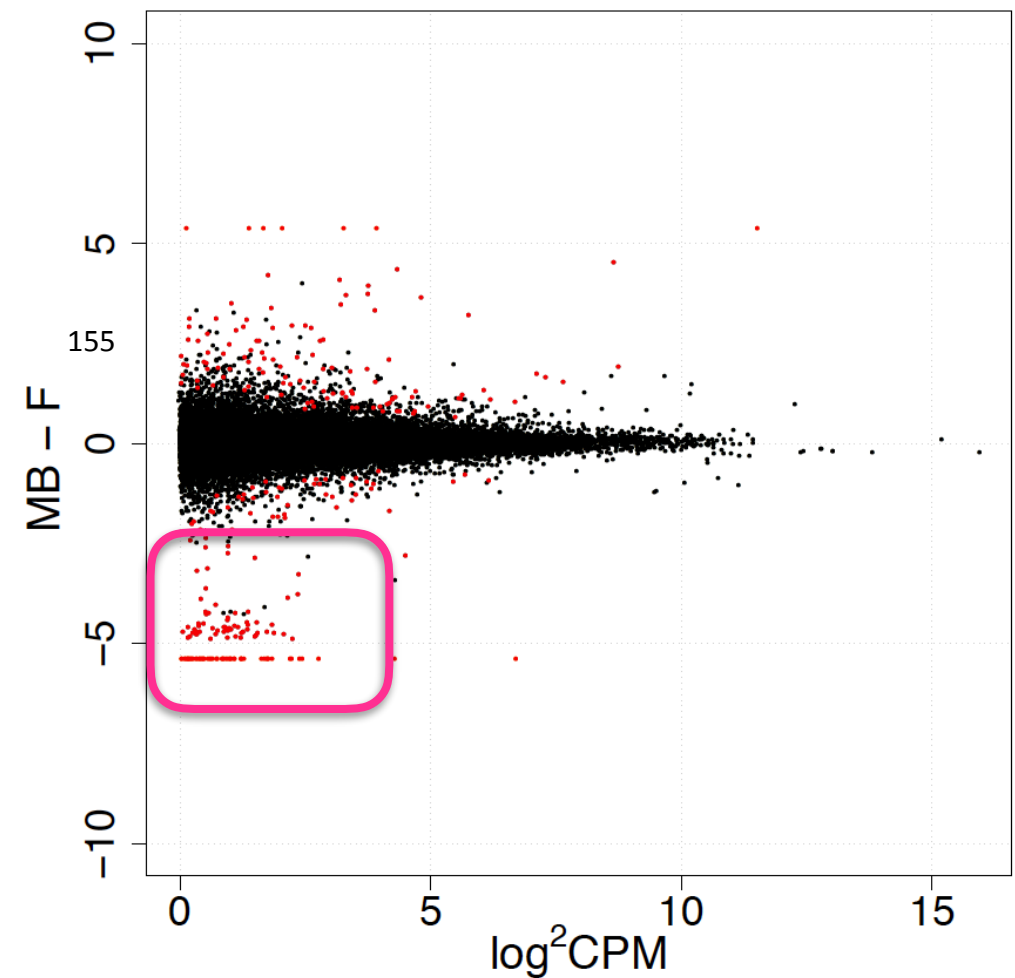# 4. Using transcriptome from Amm, Tve, Kijp 3 pops, mapping the reads

Criteria:

1.
dgl <- dgl[aveLogCPM(dgl) > 0,] # filter by average reads

DE_numbers
Total_DE   396
Male-bias   155
Female-bias 241
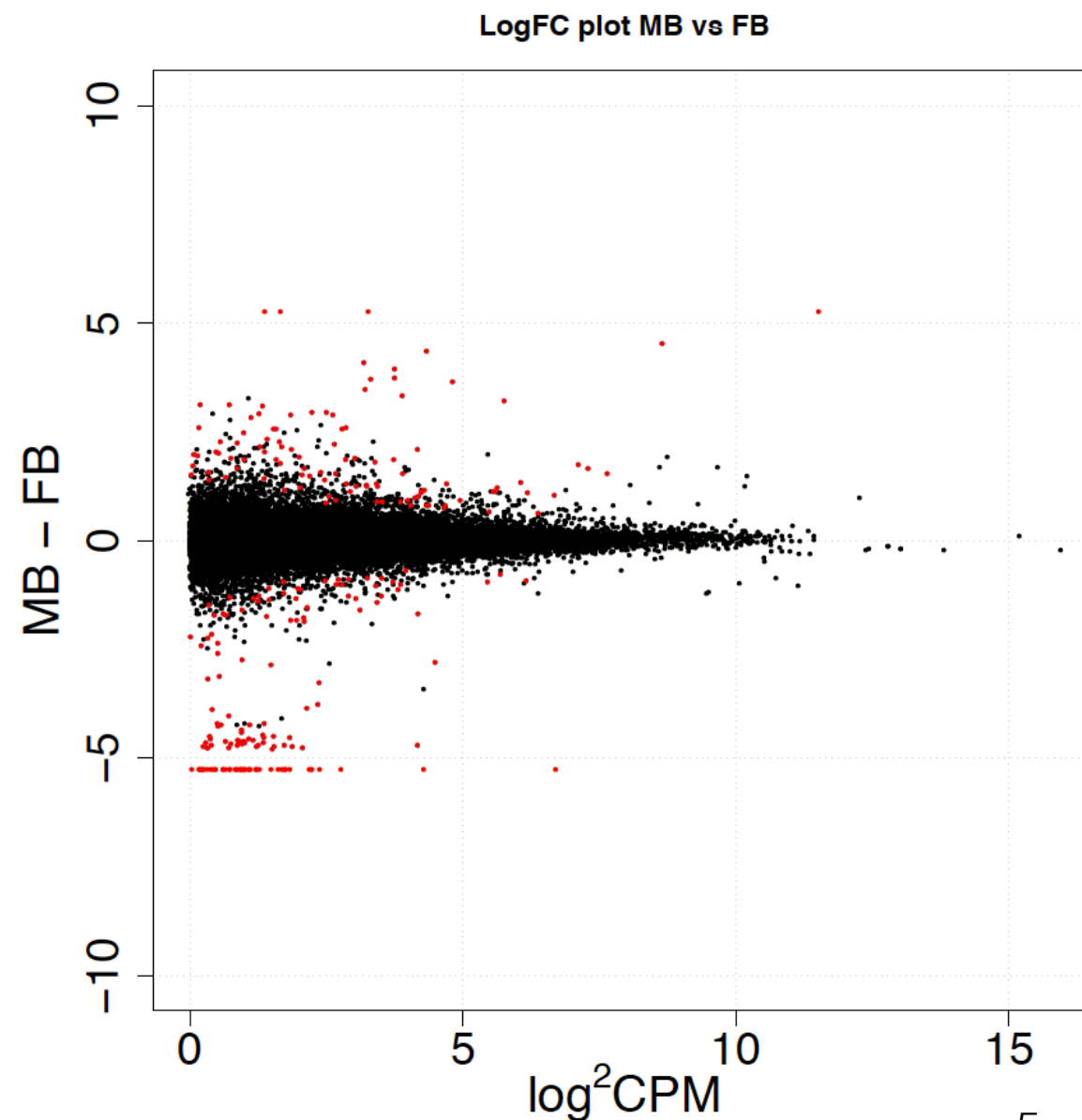
DE_numbers
Total_DE   324
Male-bias   142
Female-bias 182

4. Using transcriptome from Amm, Tve, Kijp 3 pops, mapping the reads

Criteria:

2.

```
dgl <- dgl[aveLogCPM(dgl) > 0,] # filter by average reads
dgl <- dgl[rowSums(cpm(dgl)>1) >= 3,] #filter by minimum expression in at least half of sample size
per sex
```


LogFC plot MB vs FB

DE_numbers
Total_DE   287
Male-bias   121
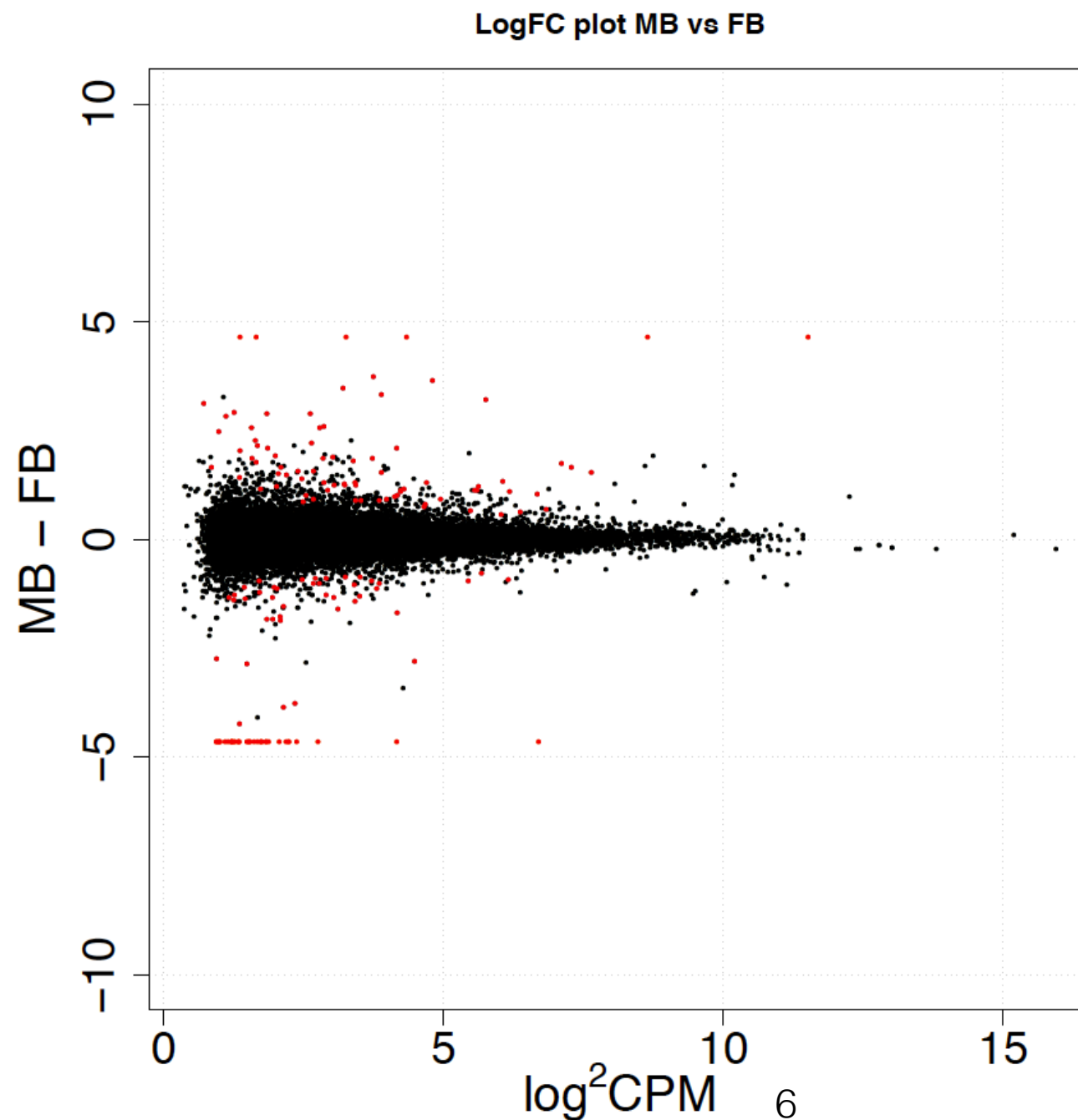Female-bias 166

4. Using transcriptome from Amm, Tve, Kijp 3 pops, mapping the reads

Criteria:
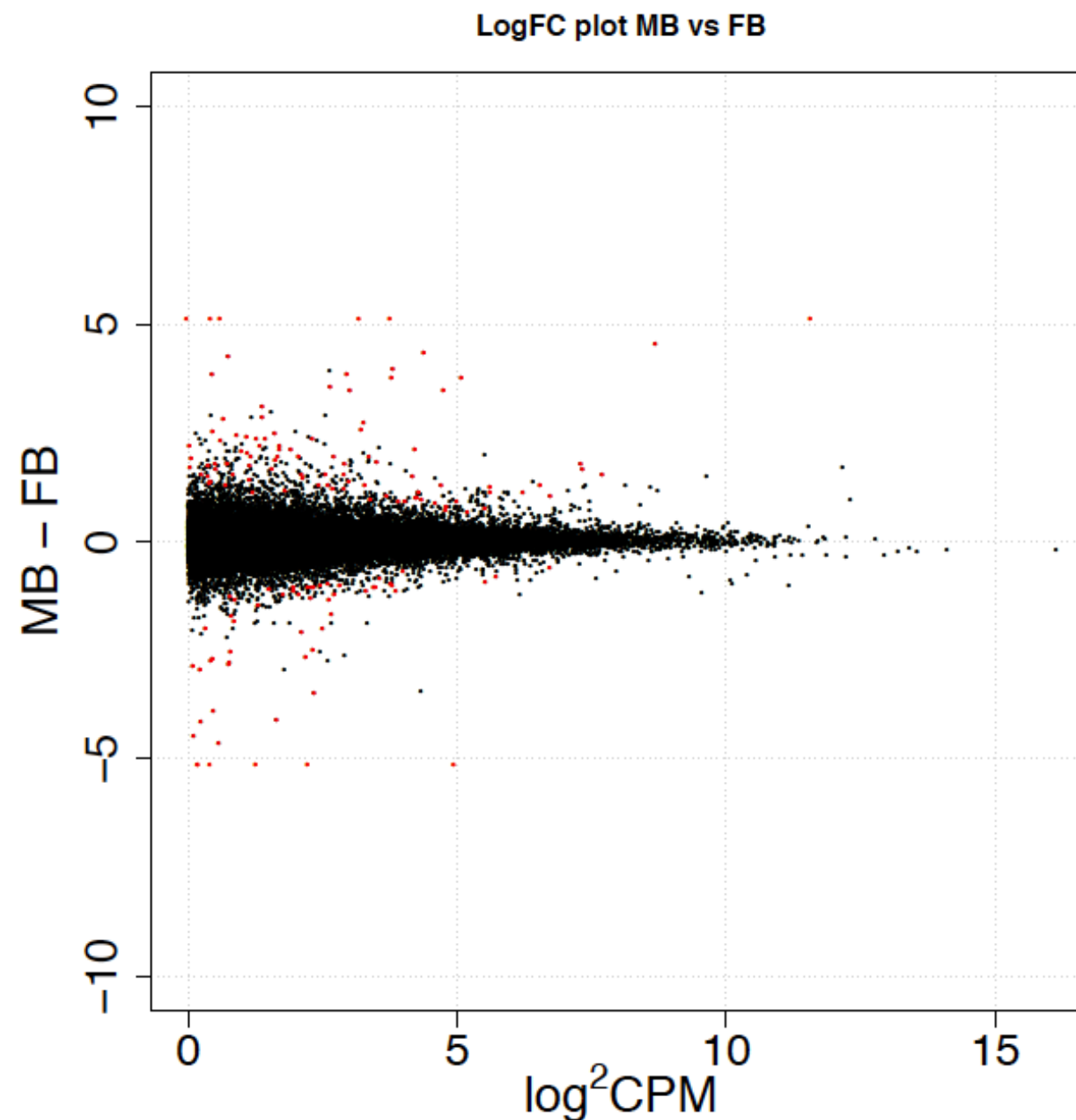
3.

```
dgl <- dgl[aveLogCPM(dgl) > 0,] # filter by average reads
dgl <- dgl[rowSums(cpm(dgl)>2) >= 4,] #filter by minimum expression in at least half of sample size
per sex
```

LogFC plot MB vs FB

DE_numbers
Total_DE   158
Male-bias   80
Female-bias 78

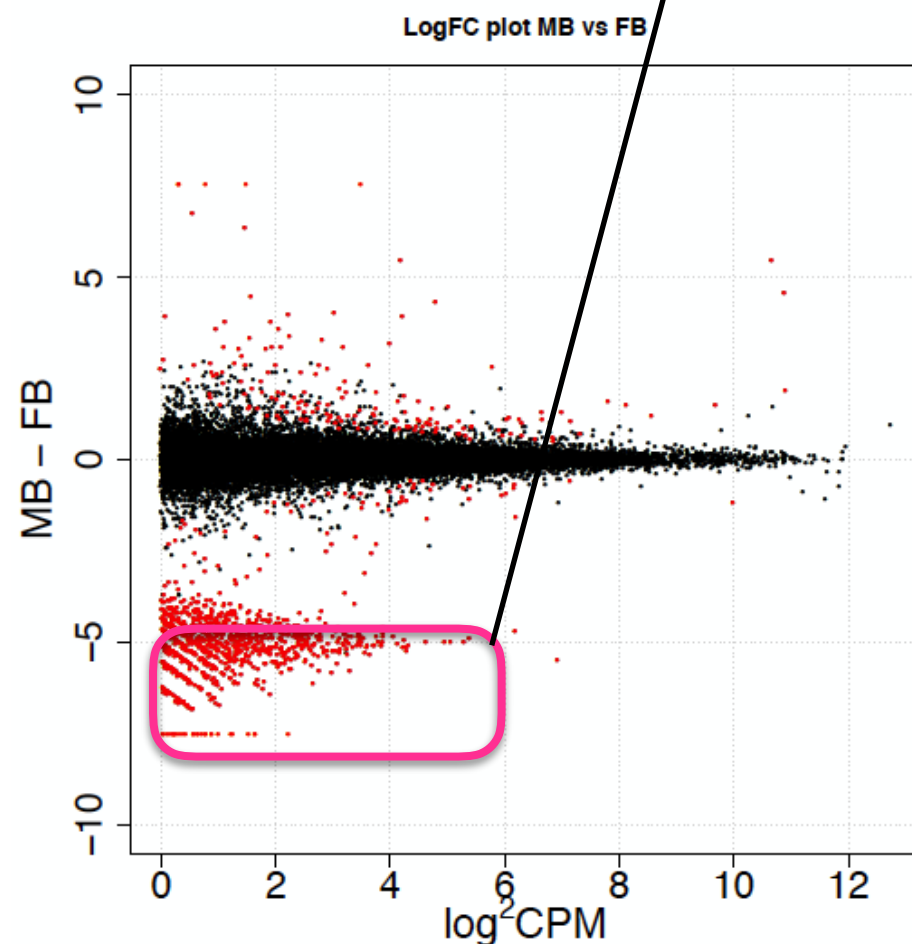# 5. Using transcriptome from Tvedora (with only embryonic and larva tissues), mapping the reads

```
dgl <- dgl[aveLogCPM(dgl) > 0,] # filter by average reads
dgl <- dgl[rowSums(cpm(dgl)>1) >= 3,] #filter by minimum expression in at least half of sample size per sex
```


LogFC plot MB vs FB

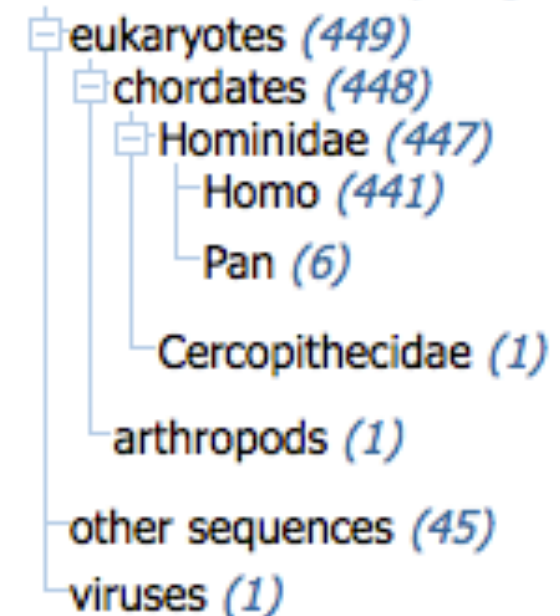DE_numbers
Total_DE   152
Male-bias   97
Female-bias 55

Amm transcriptome: substract genes highly expressed in females (log2 >= 5) reflecting the transcripts from the second clouds —>
Blast 519 highly expressed in Amm female brain, 444 out of 495 (90%) blast results are human sapiens.

```
dgl <- dgl[aveLogCPM(dgl) > 0,] # filter by average reads
dgl <- dgl[rowSums(cpm(dgl)>1) >= 3,] #filter by minimum expression in at least half of sample size per sex
```


LogFC plot MB vs FB

Results by taxon

Taxonomic Groups [List]
⊟eukaryotes *(449)*
　　⊟chordates *(448)*
　　　⊟Hominidae *(447)*
　　　　├Homo *(441)*
　　　　└Pan *(6)*
　　　└Cercopithecidae *(1)*
　　├arthropods *(1)*
　├other sequences *(45)*
　└viruses *(1)*

Would it be somewhat contamination by human sample?