

# Transcriptome *Penaeus monodon*

9/22/2021

The samples used for this analysis were:

Group	Sample ID
knockdown_PmSTAT	1_S1
knockdown_PmSTAT	2_S2
knockdown_PmSTAT	3_S3
WSSV_infection	4_S4
WSSV_infection	5_S5
WSSV_infection	6_S6
WSSV_infection	7_S7
WSSV_infection	8_S8
knockdown_PmSTAT_WSSV_infection	9_S9
knockdown_PmSTAT_WSSV_infection	10_S10
knockdown_PmSTAT_WSSV_infection	11_S11
knockdown_PmSTAT_WSSV_infection	12_S12

## Quality pre-treatment

The raw data were filtered using *Trimmomatic* with this parameters:

1. Remotion of adapters and N's
2. Average quality of reads > Q20

In the next table and graph we can observe the number of reads before (1.Raw\_data) and after (2.Quality\_filtered) the quality pre-treatment.

Sample	Raw data	Quality filtered	% after quality
1_S1	16 202 030	16 097 983	99.35
2_S2	12 890 555	12 788 438	99.20
3_S3	839 588	830 354	98.90
4_S4	14 467 820	14 371 003	99.33
5_S5	13 316 992	13 245 638	99.46
6_S6	43 615 302	43 404 612	99.51
7_S7	13 142 443	13 070 733	99.45
8_S8	13 871 940	13 791 480	99.41
9_S9	10 403 660	10 335 646	99.34
10_S10	14 615 414	14 500 097	99.21
11_S11	14 428 705	14 355 271	99.49
12_S12	11 557 788	11 479 499	99.32

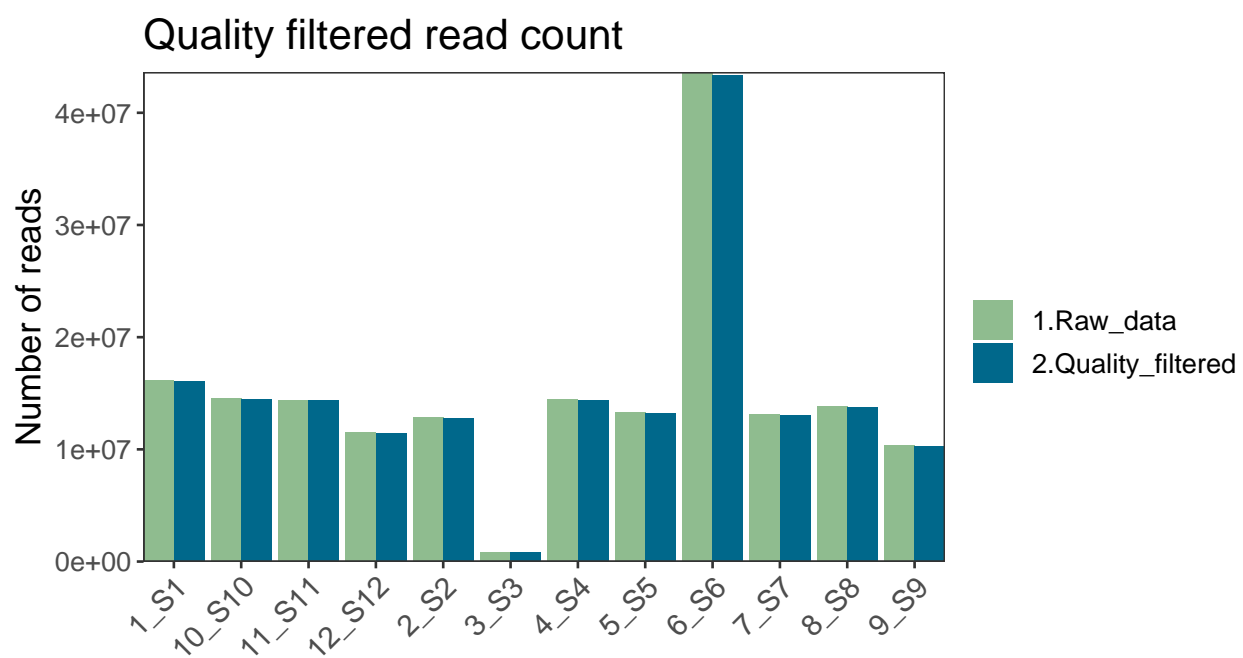


Figure 1: Read count before and after the quality filtering

## *De novo* assembly with Trinity

With the quality filtered data, we made a *de novo* assembly with Trinity. These are the *Trinity.fasta* stats:

---

<b>Counts of transcripts, etc.</b>	
Total trinity ‘genes’	586 928
Total trinity transcripts	778 680
Percent GC	37,7
-	-
<b>Stats based on ALL transcript contigs</b>	
Contig N10	3 271
Contig N20	1 700
Contig N30	799
Contig N40	574
Contig N50	472
Median contig length	331
Average contig	467,55
Total assembled bases	364 068 637
-	-
<b>Stats based on ONLY LONGEST ISOFORM per ‘GENE’</b>	
Contig N10	1 672
Contig N20	727
Contig N30	565
Contig N40	479
Contig N50	417
Median contig length	325
Average contig	409,1
Total assembled bases	240110795

---

To determine the representativeness of our samples in the assembly, each of them was realigned using bowtie2.

---

Sample	% realigned
1_S1	76.34
2_S2	90.59
3_S3	87.12
4_S4	89.03
5_S5	88.54
6_S6	91.74
7_S7	89.92
8_S8	85.38
9_S9	92.13
10_S10	84.76
11_S11	91.25
12_S12	78.82

---

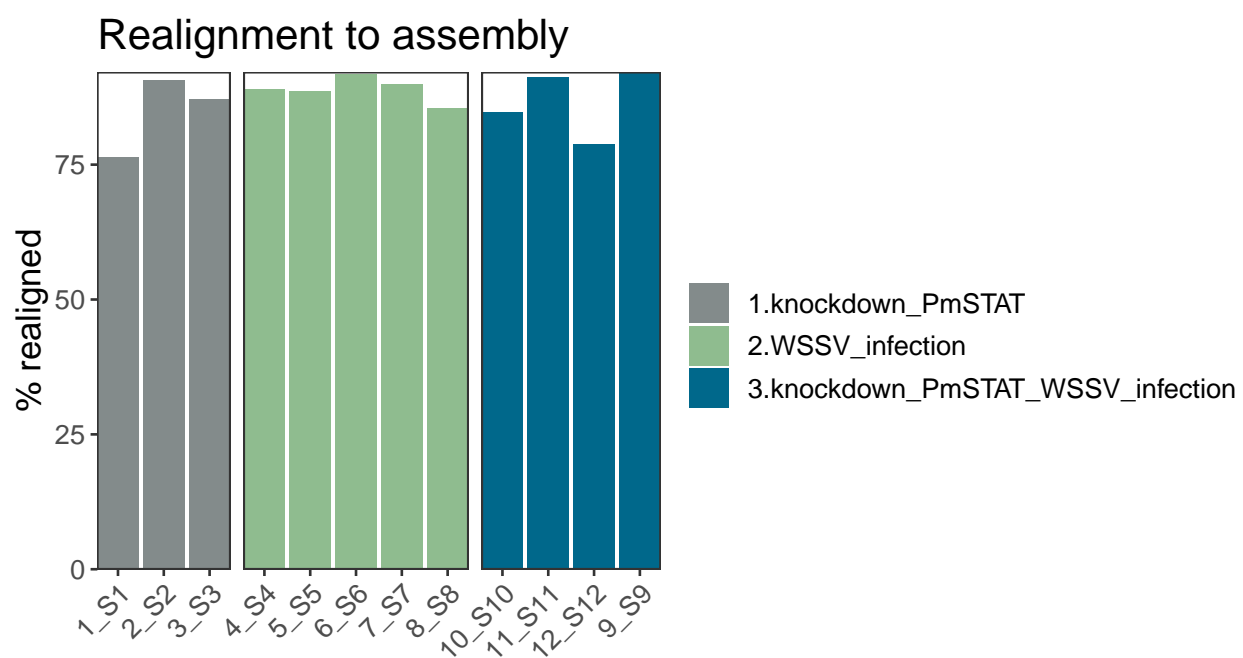


Figure 2: Percentage of reads realigned to the assembly

## Differential expression with DESeq2

We used the DESeq2 with *Log Fold Change*>2 and *p-value*<0.05.

### Considering genes

WSSV infection vs knockdown PmSTAT	DE genes
knockdown PmSTAT-UP	56
WSSV infection-UP	813

WSSV infection vs knockdown PmSTAT WSSV infection	DE genes
knockdown PmSTAT WSSV infection-UP	138
WSSV infection-UP	675

knockdown PmSTAT vs knockdown PmSTAT WSSV infection	DE genes
knockdown PmSTAT-UP	0
knockdown PmSTAT WSSV infection-UP	0

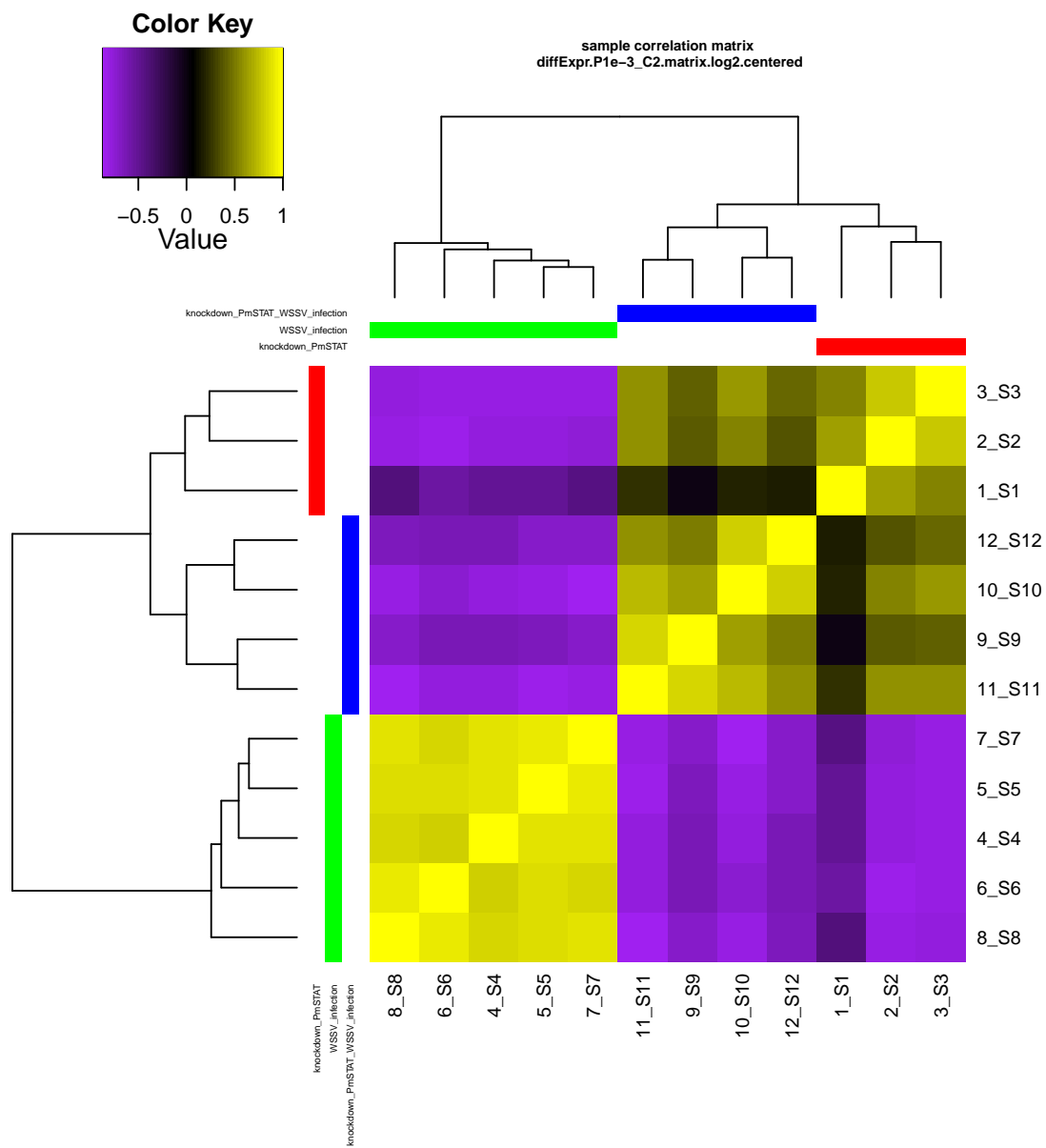


Figure 3: Correlation between samples considering genes

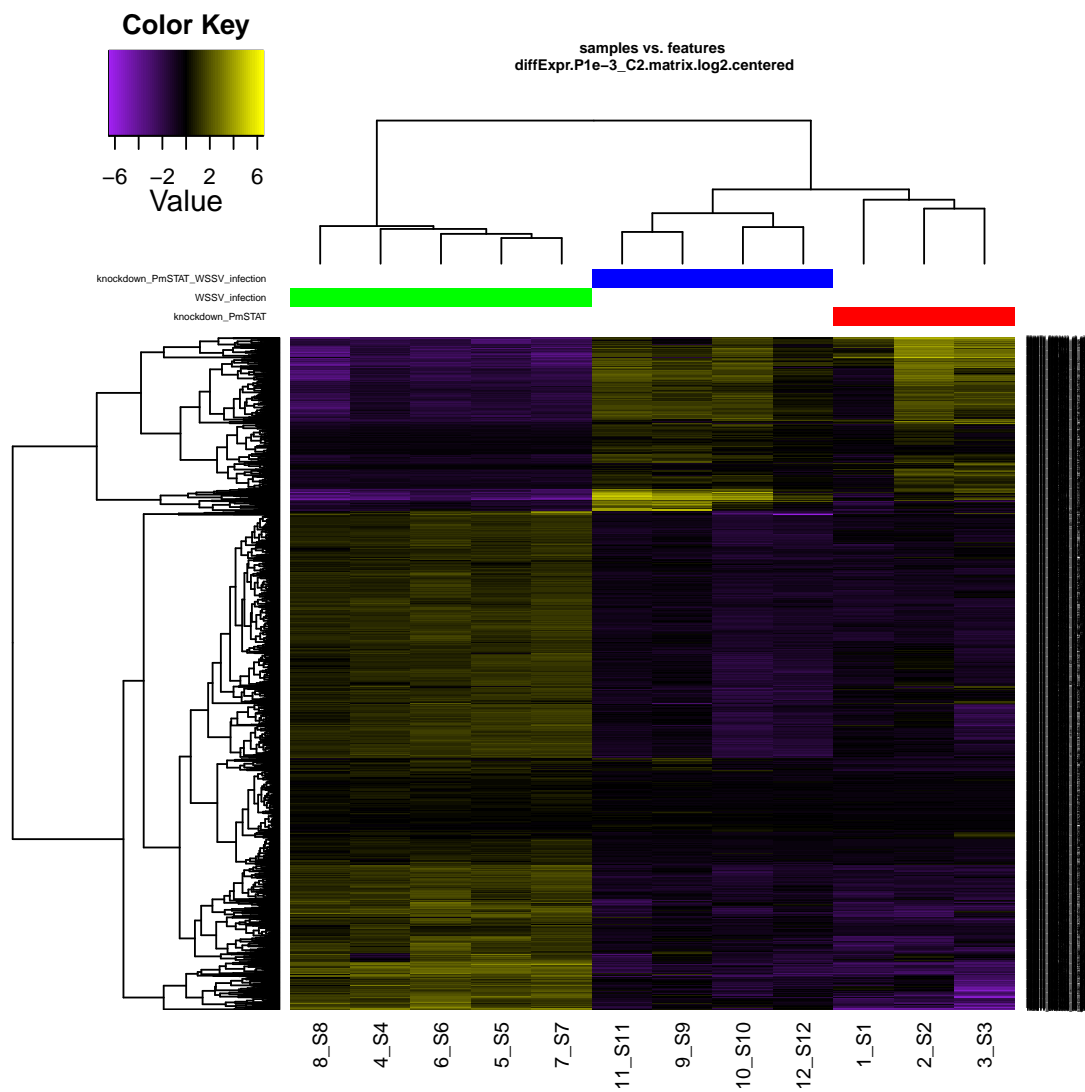


Figure 4: Hierarchical clustering considering genes

## Considering isoforms

WSSV infection vs knockdown PmSTAT	DE isoformsd
knockdown PmSTAT-UP	92
WSSV infection-UP	496

WSSV infection vs knockdown PmSTAT WSSV infection	DE isoforms
knockdown PmSTAT WSSV infection-UP	474
WSSV infection-UP	1242

knockdown PmSTAT vs knockdown PmSTAT WSSV infection	DE isoforms
knockdown PmSTAT-UP	0
knockdown PmSTAT WSSV infection-UP	0



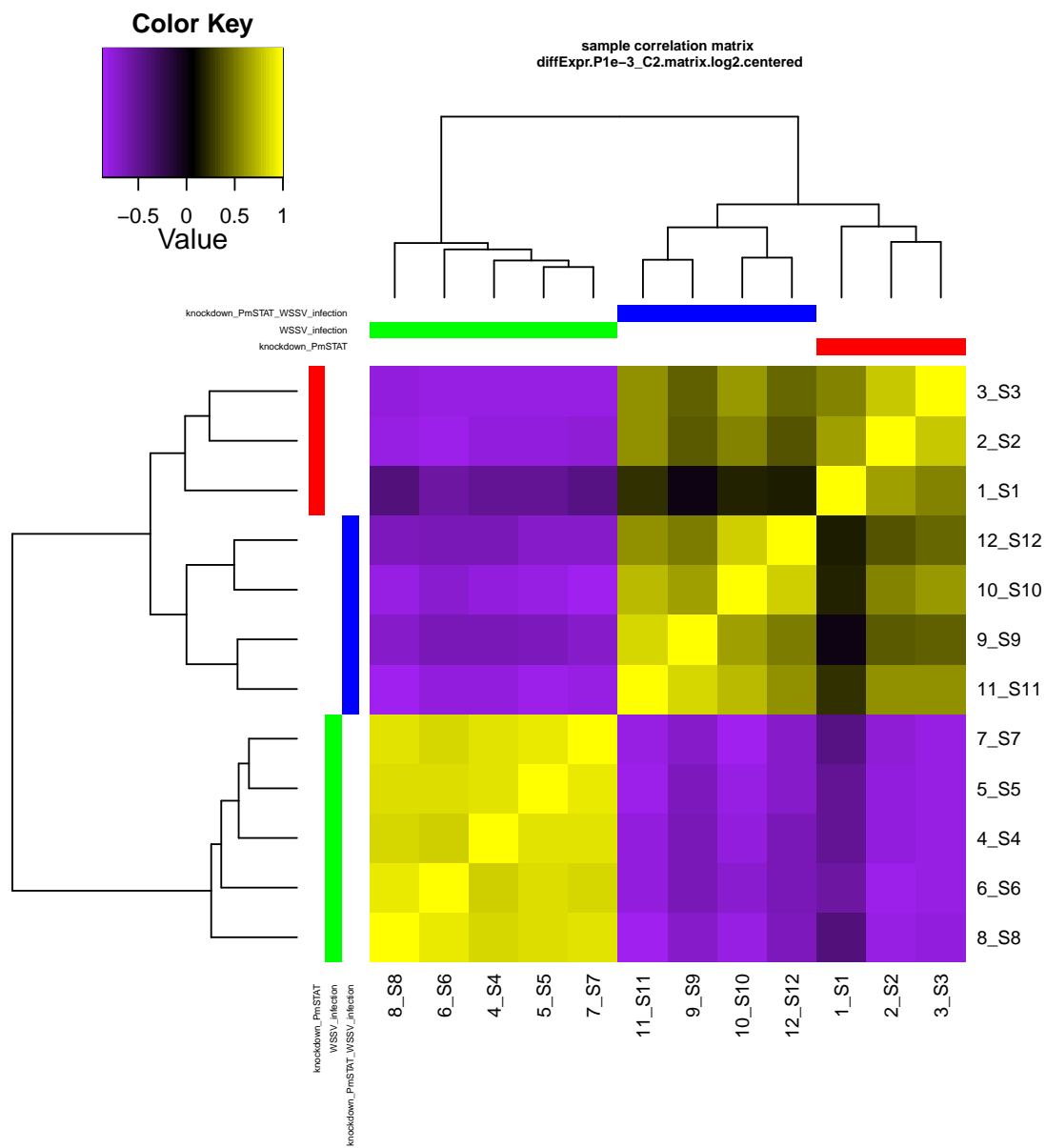


Figure 5: Correlation between samples considering all isoforms

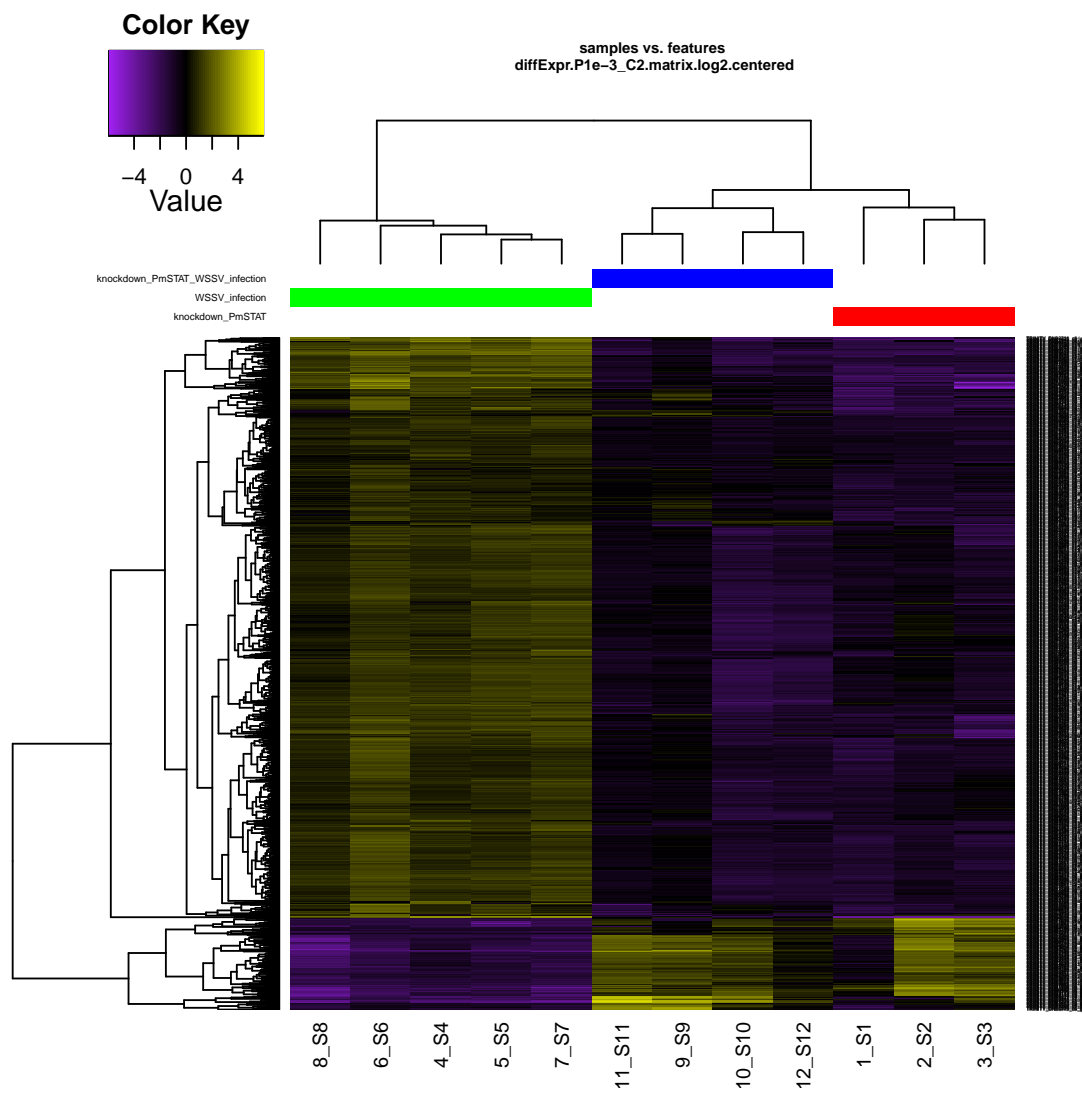


Figure 6: Hierarchical clustering considering all isoforms