512 Project

Nolan Walker

2/17/2021

***Introduction***

**Data Description**

We are looking at the mouse cortex protein expression dataset from Higuera C et. al(2015), where 77 cortex proteins were observed in a mouse model of down syndrome. Some mice were treated with metamine, a drug used to treat alzheimers, to show that learning can be recovered in down syndrome mice. Mice were either assigned to learn, where they were given context and then a shock, or not to learn, where they were only given a shock. This gives a total of 8 classes(2X2X2).There were 15 measurements for each protein, 38 control mice, and 34 down syndrome mice. The goal of this study is to see if there is a difference in protein expression levels across the different classes for the various proteins.

**Data exploration**

dim(data)

## [1] 1080 82

str(data[-c(2:78)])

## 'data.frame': 1080 obs. of 5 variables:  
## $ MouseID : chr "309\_1" "309\_2" "309\_3" "309\_4" ...  
## $ Genotype : chr "Control" "Control" "Control" "Control" ...  
## $ Treatment: chr "Memantine" "Memantine" "Memantine" "Memantine" ...  
## $ Behavior : chr "C/S" "C/S" "C/S" "C/S" ...  
## $ class : chr "c-CS-m" "c-CS-m" "c-CS-m" "c-CS-m" ...

unique(data$class)

## [1] "c-CS-m" "c-SC-m" "c-CS-s" "c-SC-s" "t-CS-m" "t-SC-m" "t-CS-s" "t-SC-s"

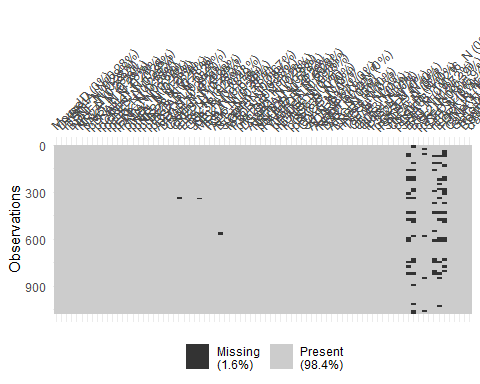
names(data[c(2:78)])

## [1] "DYRK1A\_N" "ITSN1\_N" "BDNF\_N" "NR1\_N"   
## [5] "NR2A\_N" "pAKT\_N" "pBRAF\_N" "pCAMKII\_N"   
## [9] "pCREB\_N" "pELK\_N" "pERK\_N" "pJNK\_N"   
## [13] "PKCA\_N" "pMEK\_N" "pNR1\_N" "pNR2A\_N"   
## [17] "pNR2B\_N" "pPKCAB\_N" "pRSK\_N" "AKT\_N"   
## [21] "BRAF\_N" "CAMKII\_N" "CREB\_N" "ELK\_N"   
## [25] "ERK\_N" "GSK3B\_N" "JNK\_N" "MEK\_N"   
## [29] "TRKA\_N" "RSK\_N" "APP\_N" "Bcatenin\_N"   
## [33] "SOD1\_N" "MTOR\_N" "P38\_N" "pMTOR\_N"   
## [37] "DSCR1\_N" "AMPKA\_N" "NR2B\_N" "pNUMB\_N"   
## [41] "RAPTOR\_N" "TIAM1\_N" "pP70S6\_N" "NUMB\_N"   
## [45] "P70S6\_N" "pGSK3B\_N" "pPKCG\_N" "CDK5\_N"   
## [49] "S6\_N" "ADARB1\_N" "AcetylH3K9\_N" "RRP1\_N"   
## [53] "BAX\_N" "ARC\_N" "ERBB4\_N" "nNOS\_N"   
## [57] "Tau\_N" "GFAP\_N" "GluR3\_N" "GluR4\_N"   
## [61] "IL1B\_N" "P3525\_N" "pCASP9\_N" "PSD95\_N"   
## [65] "SNCA\_N" "Ubiquitin\_N" "pGSK3B\_Tyr216\_N" "SHH\_N"   
## [69] "BAD\_N" "BCL2\_N" "pS6\_N" "pCFOS\_N"   
## [73] "SYP\_N" "H3AcK18\_N" "EGR1\_N" "H3MeK4\_N"   
## [77] "CaNA\_N"

As we can see, there are a lot(1080) of measurements for each of the 77 proteins. There are eight classes; combinations of Genotype, Treatment, and Behavior.

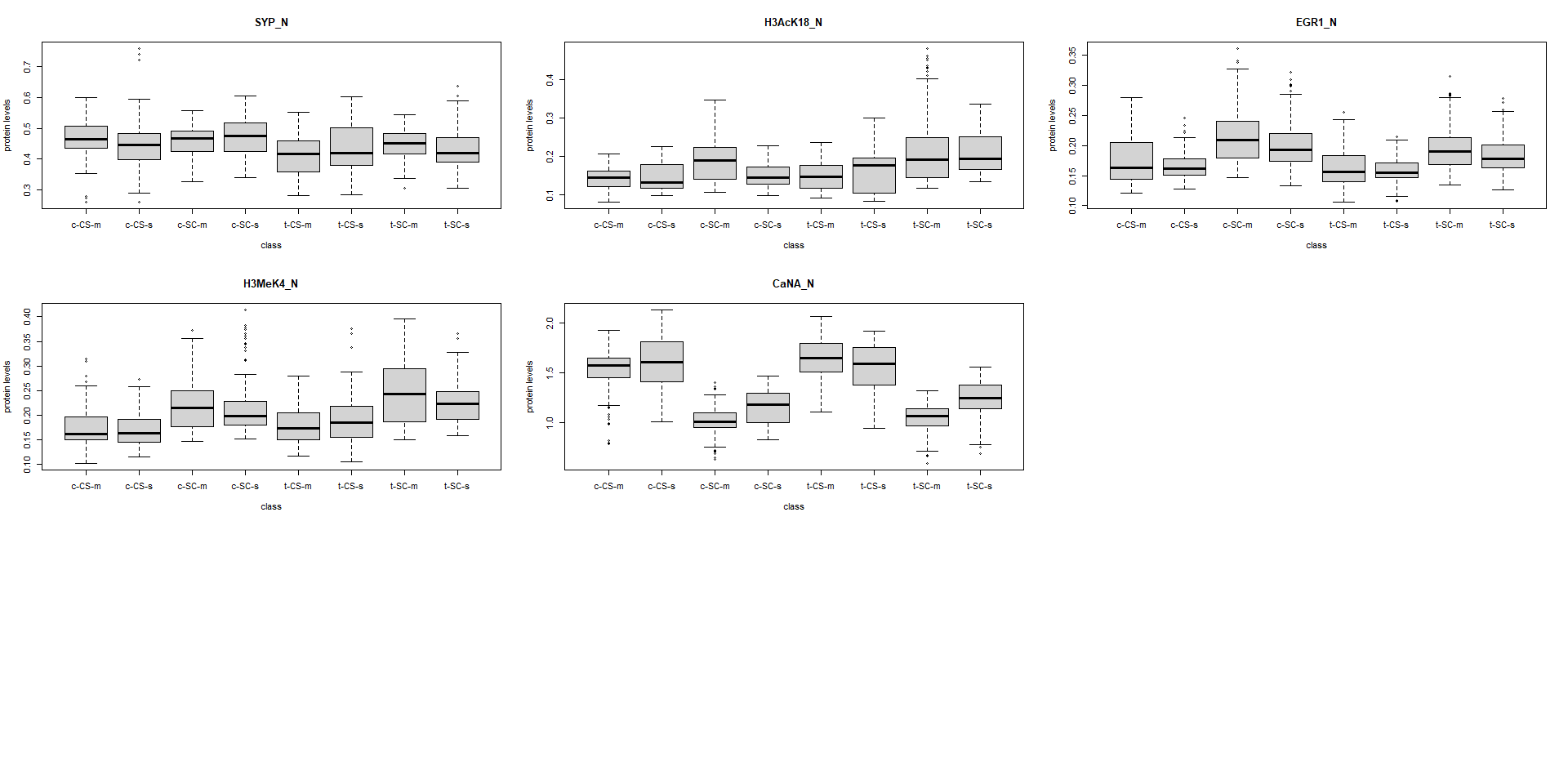
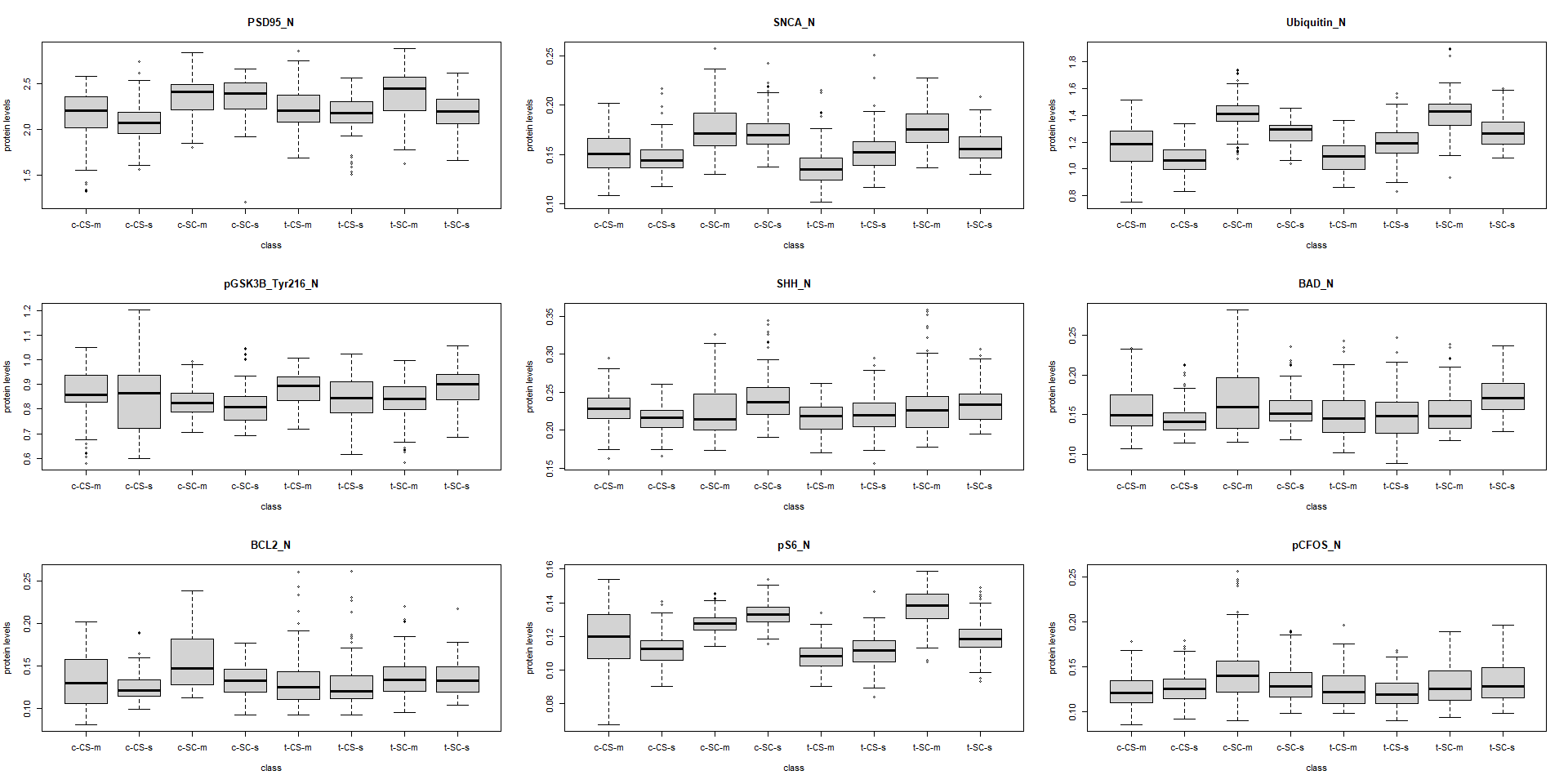
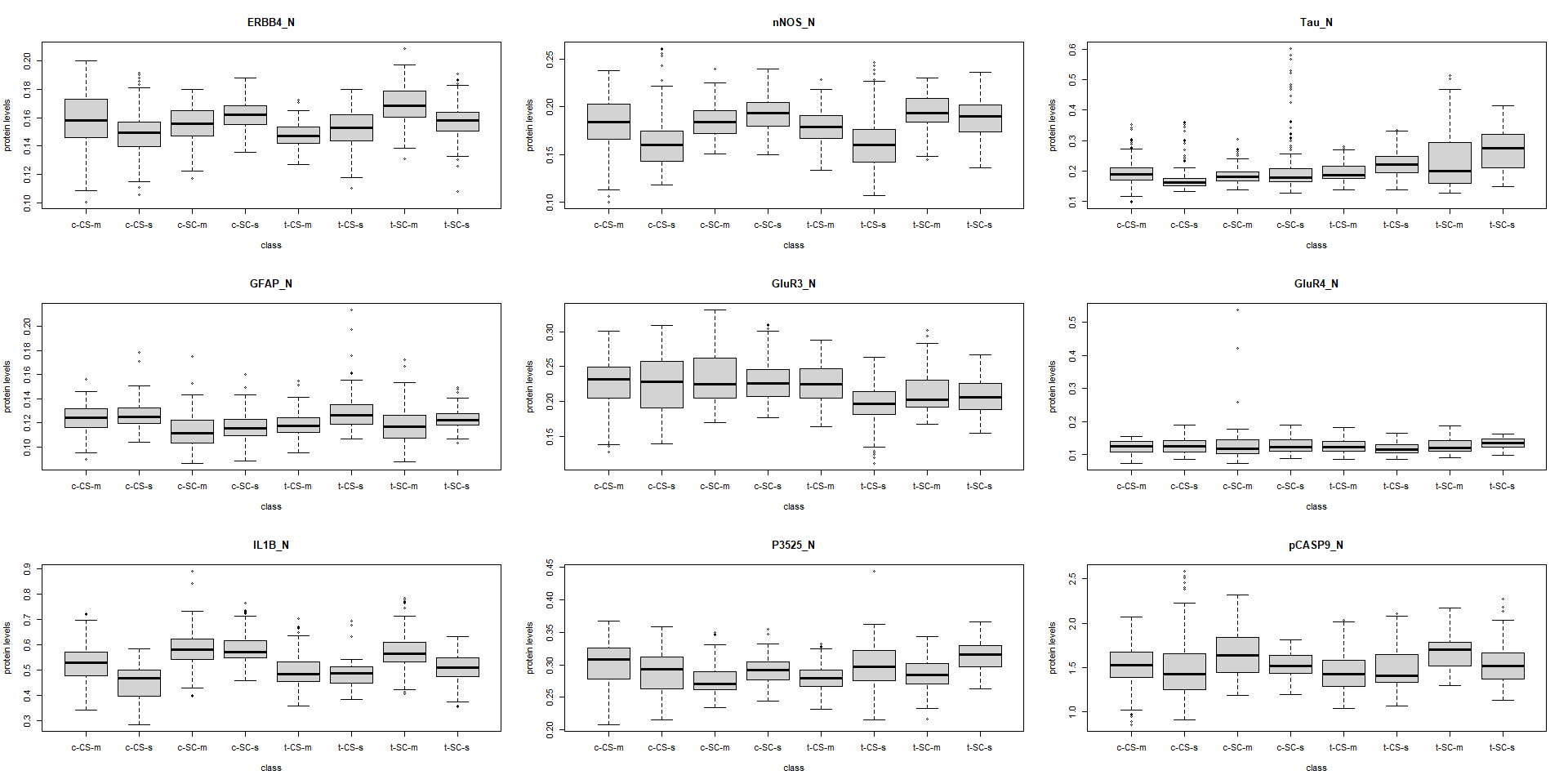
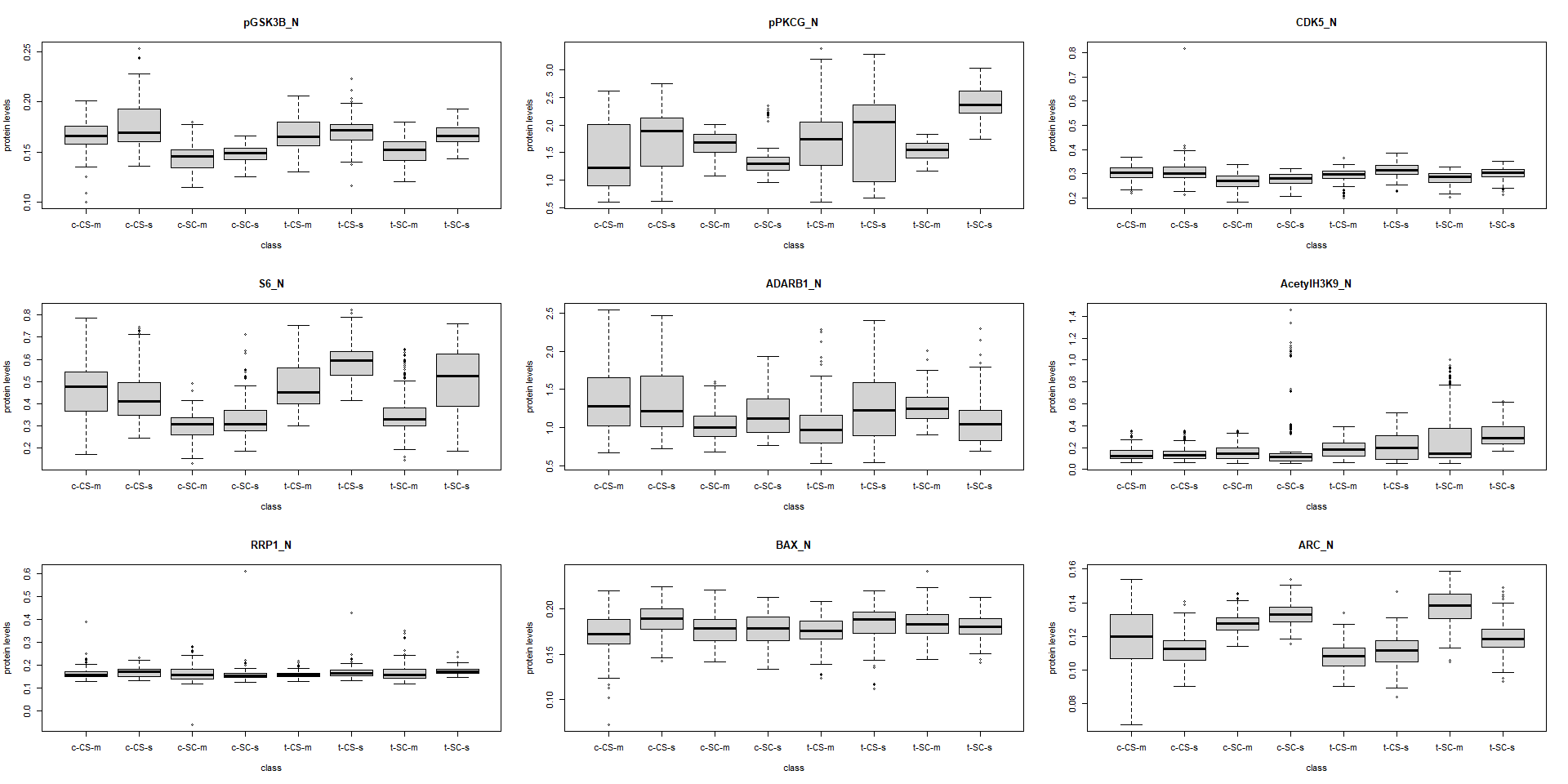
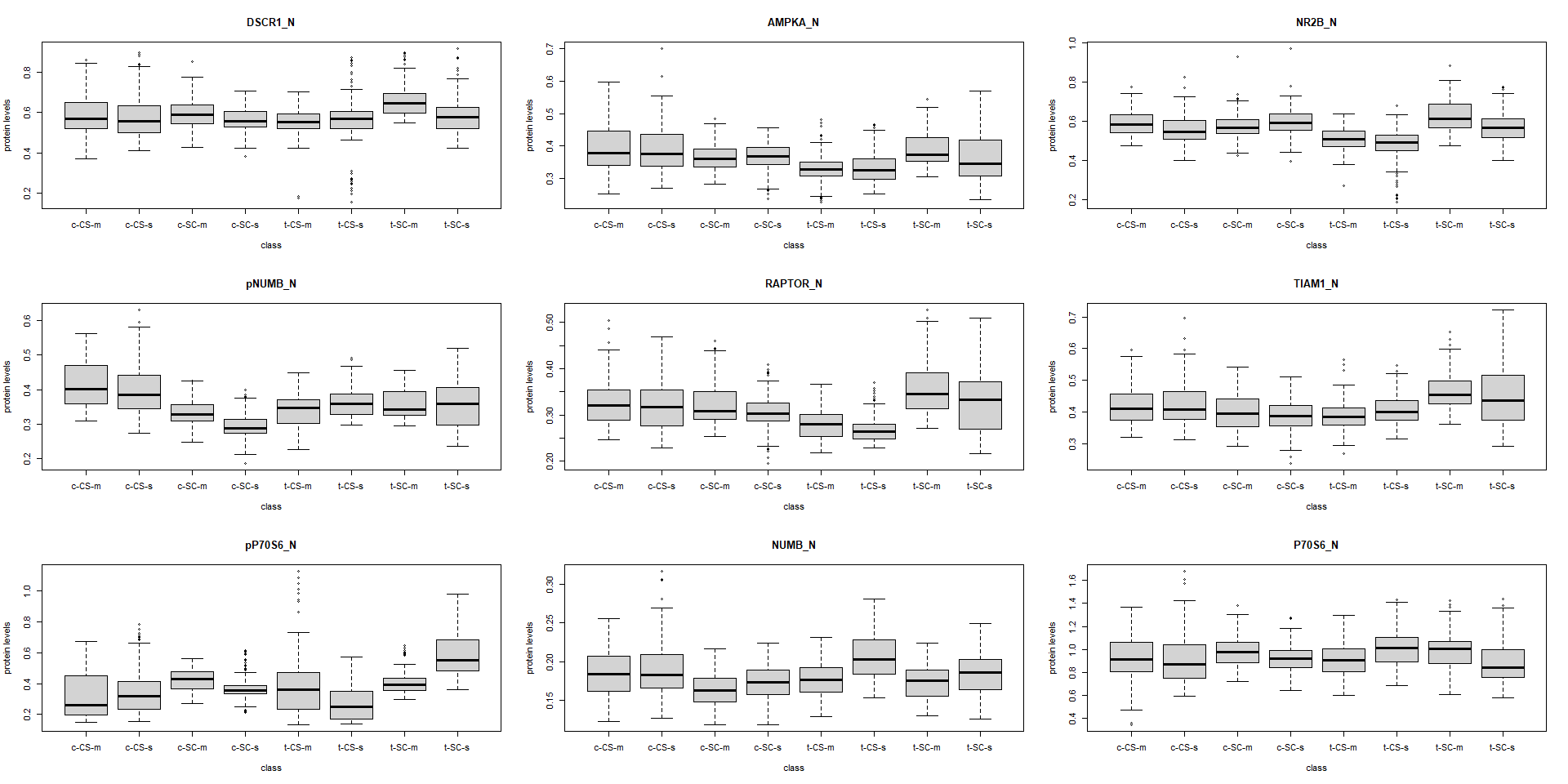
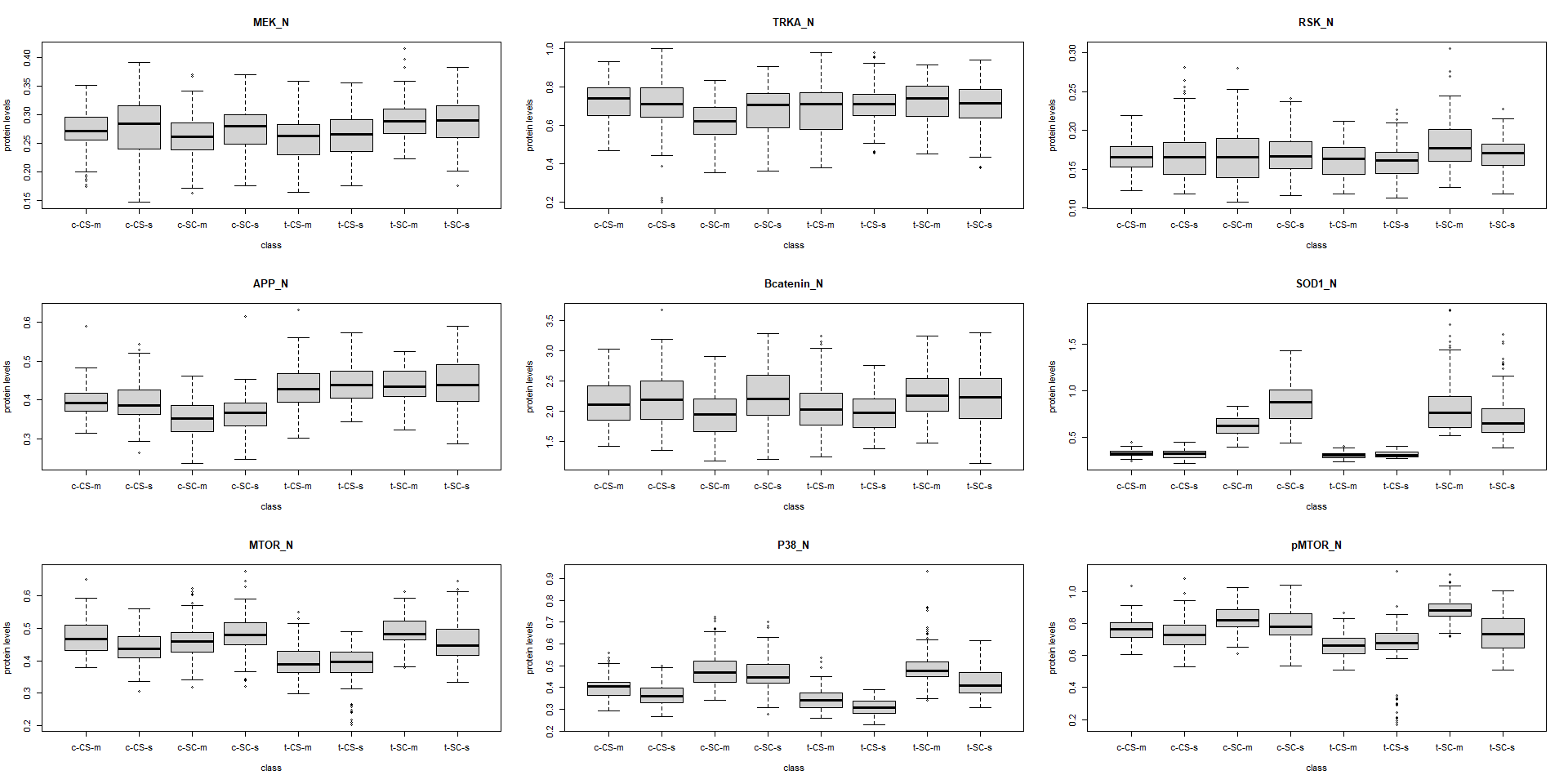
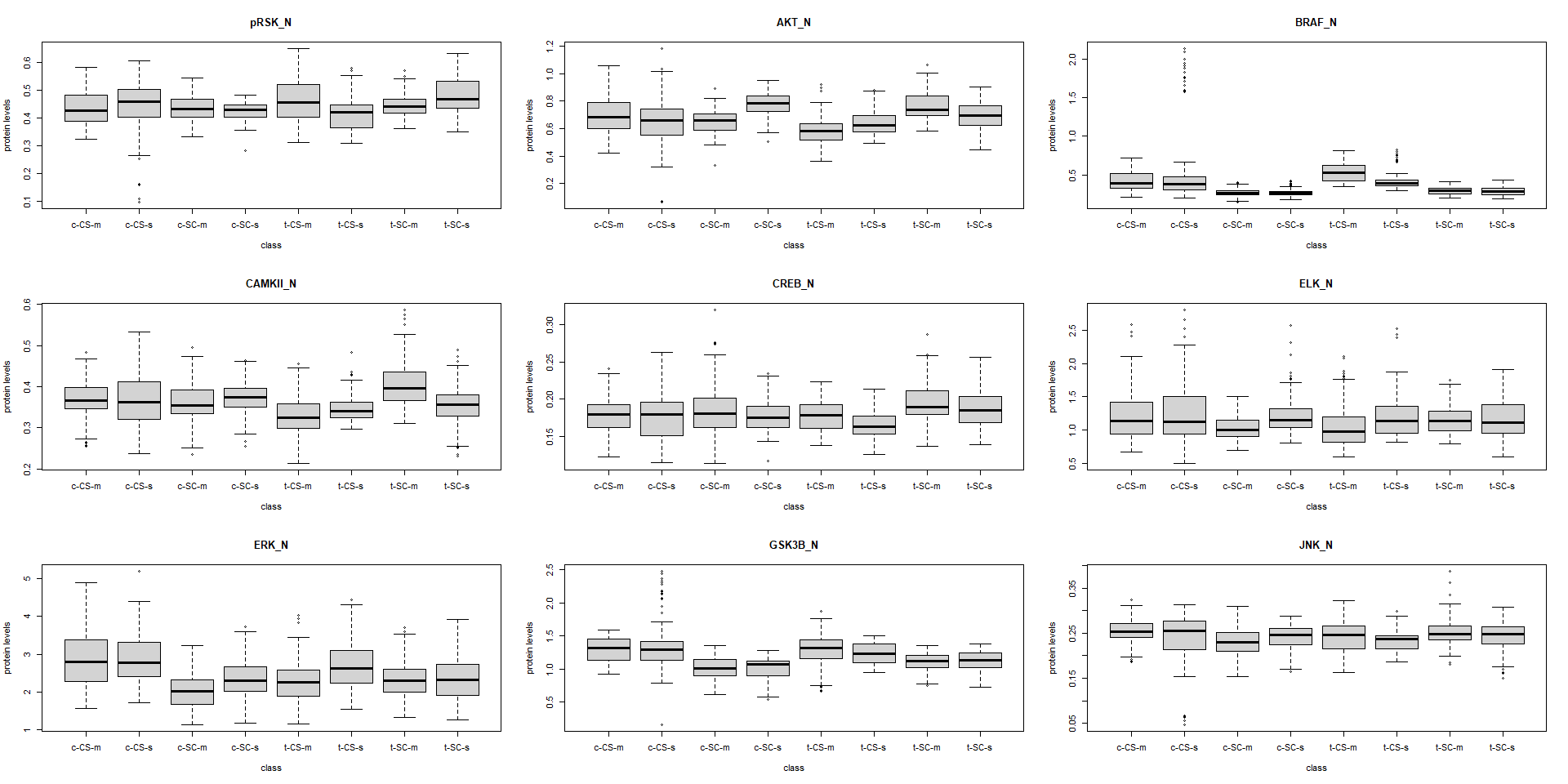
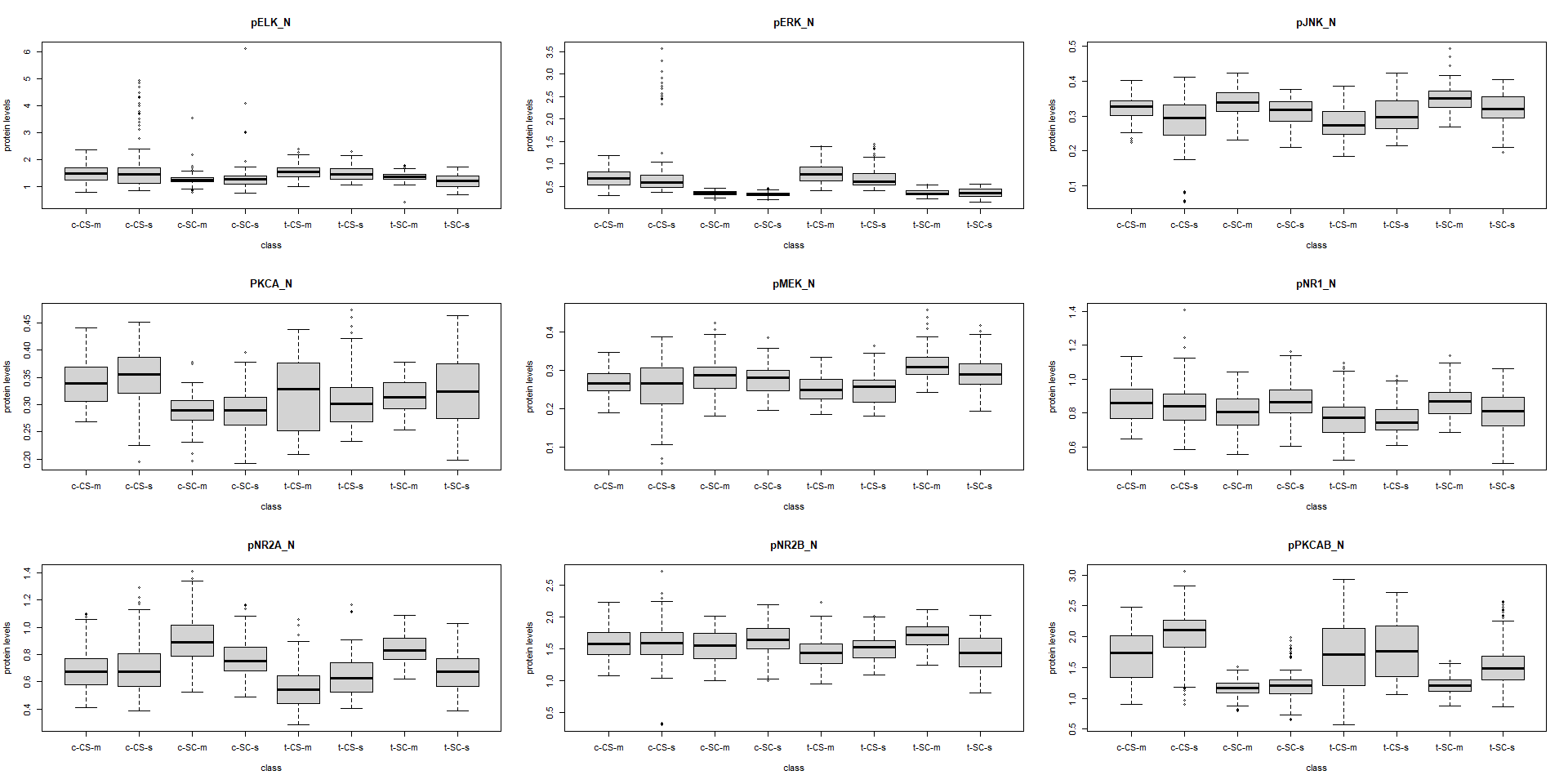
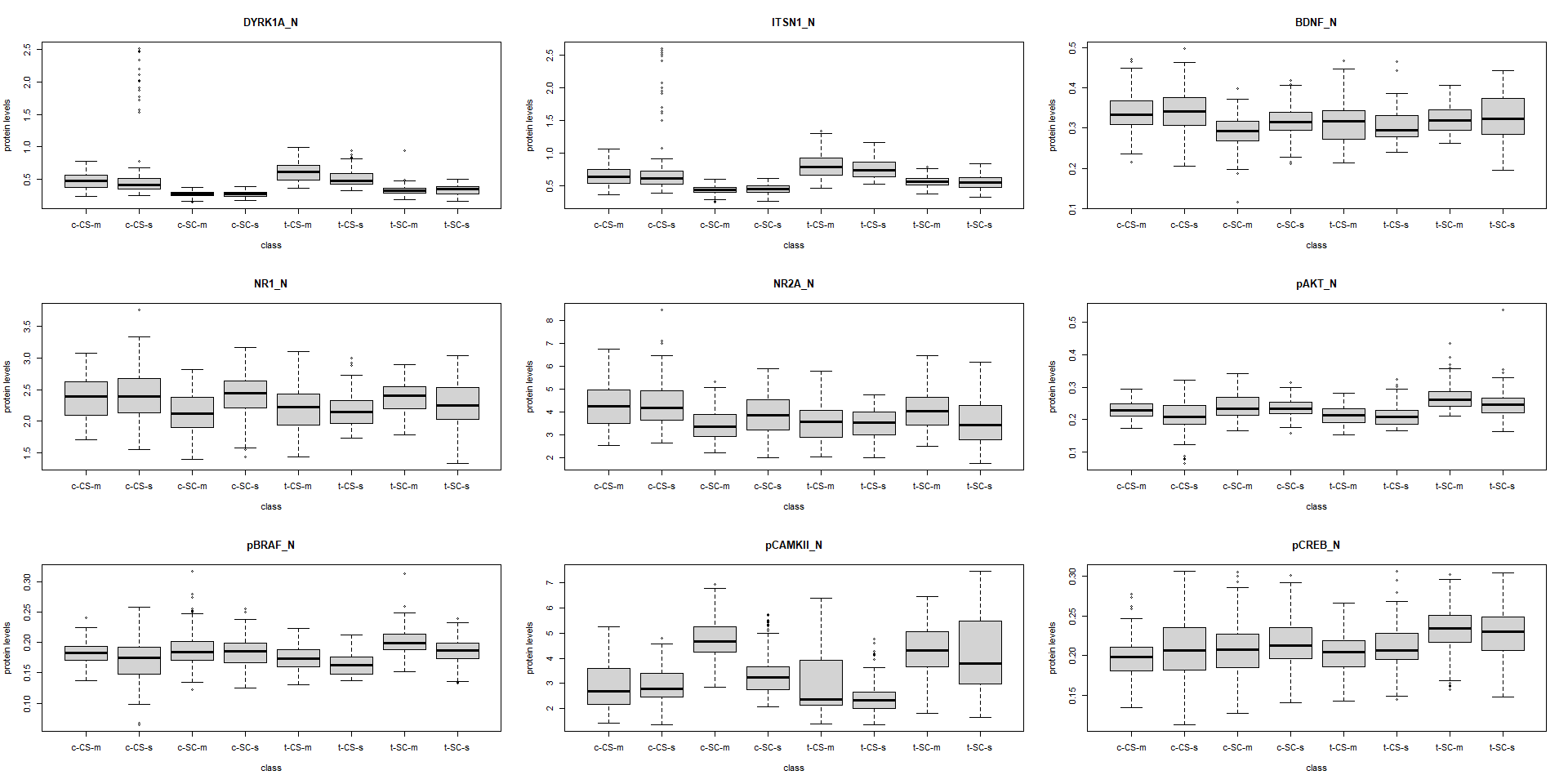
Next to observe missing values.

vis\_miss(data)



As is observed in the plot of missing values, a few of the proteins towards the right of the plot are missing quite a few observations, so some data imputing must be done.

It would be useful to see the distribution of observations for the different proteins.



It looks like some proteins have quite a large difference in measurments across classes. It would make sense to fit multiple mean models to the different proteins to see if there is a difference.

***H\_null****: there are no diferrences in expression level mean between classes.*

***H\_alt****: there is at least one significant difference among the groups.*

**Model fitting**

We fit a simple model to each protein :

{Protein|c-CS-m,c-SC-m,c-CS-s,c-SC-s,t-CS-m,t-SC-m,t-CS-s,t-SC-s}

lm\_herb1:

## Protein Var\_expl p.value  
## 1 DYRK1A\_N 0.29 3.980441e-75  
## 2 ITSN1\_N 0.29 1.230006e-74  
## 3 BDNF\_N 0.11 8.725596e-24  
## 4 NR1\_N 0.09 1.357447e-19  
## 5 NR2A\_N 0.13 1.868668e-28  
## 6 pAKT\_N 0.19 1.779639e-45  
## 7 pBRAF\_N 0.16 1.320758e-36  
## 8 pCAMKII\_N 0.33 2.818994e-88  
## 9 pCREB\_N 0.11 7.944844e-24  
## 10 pELK\_N 0.11 9.005484e-23  
## 11 pERK\_N 0.36 4.569332e-99  
## 12 pJNK\_N 0.21 6.603188e-51  
## 13 PKCA\_N 0.16 5.708838e-38  
## 14 pMEK\_N 0.19 1.157002e-44  
## 15 pNR1\_N 0.11 5.342219e-25  
## 16 pNR2A\_N 0.31 5.188787e-82  
## 17 pNR2B\_N 0.11 1.000318e-22  
## 18 pPKCAB\_N 0.39 6.452687e-109  
## 19 pRSK\_N 0.08 3.286824e-16  
## 20 AKT\_N 0.23 3.370599e-57  
## 21 BRAF\_N 0.26 4.992471e-65  
## 22 CAMKII\_N 0.15 1.045206e-33  
## 23 CREB\_N 0.09 6.970974e-19  
## 24 ELK\_N 0.06 1.983440e-11  
## 25 ERK\_N 0.19 3.024075e-44  
## 26 GSK3B\_N 0.26 2.865826e-65  
## 27 JNK\_N 0.06 9.683309e-11  
## 28 MEK\_N 0.07 5.170426e-14  
## 29 TRKA\_N 0.08 2.068313e-15  
## 30 RSK\_N 0.05 5.885899e-10  
## 31 APP\_N 0.33 4.171656e-88  
## 32 Bcatenin\_N 0.08 1.451842e-15  
## 33 SOD1\_N 0.66 1.209388e-243  
## 34 MTOR\_N 0.27 5.161274e-70  
## 35 P38\_N 0.47 3.545236e-142  
## 36 pMTOR\_N 0.36 1.039129e-98  
## 37 DSCR1\_N 0.12 2.887391e-25  
## 38 AMPKA\_N 0.15 5.249088e-33  
## 39 NR2B\_N 0.27 2.816364e-68  
## 40 pNUMB\_N 0.30 4.061030e-80  
## 41 RAPTOR\_N 0.21 2.472756e-50  
## 42 TIAM1\_N 0.15 3.951100e-35  
## 43 pP70S6\_N 0.27 1.437178e-70  
## 44 NUMB\_N 0.17 6.765151e-40  
## 45 P70S6\_N 0.06 7.397068e-11  
## 46 pGSK3B\_N 0.36 7.114063e-100  
## 47 pPKCG\_N 0.25 1.082395e-62  
## 48 CDK5\_N 0.17 8.453641e-39  
## 49 S6\_N 0.39 1.250355e-110  
## 50 ADARB1\_N 0.13 2.549053e-30  
## 51 AcetylH3K9\_N 0.13 7.566911e-29  
## 52 RRP1\_N 0.03 1.475502e-05  
## 53 BAX\_N 0.07 6.381704e-14  
## 54 ARC\_N 0.48 1.728614e-145  
## 55 ERBB4\_N 0.19 7.883676e-46  
## 56 nNOS\_N 0.20 8.639413e-47  
## 57 Tau\_N 0.18 2.985743e-42  
## 58 GFAP\_N 0.14 3.946997e-32  
## 59 GluR3\_N 0.11 7.395391e-24  
## 60 GluR4\_N 0.02 2.040815e-03  
## 61 IL1B\_N 0.32 2.060474e-84  
## 62 P3525\_N 0.14 1.033929e-32  
## 63 pCASP9\_N 0.08 5.008883e-16  
## 64 PSD95\_N 0.18 1.390435e-41  
## 65 SNCA\_N 0.33 3.281180e-90  
## 66 Ubiquitin\_N 0.51 1.628986e-162  
## 67 pGSK3B\_Tyr216\_N 0.07 9.536395e-14  
## 68 SHH\_N 0.08 1.039664e-17  
## 69 BAD\_N 0.10 2.102935e-20  
## 70 BCL2\_N 0.09 1.736128e-19  
## 71 pS6\_N 0.48 1.728614e-145  
## 72 pCFOS\_N 0.07 2.001019e-14  
## 73 SYP\_N 0.08 2.974103e-16  
## 74 H3AcK18\_N 0.17 1.058385e-38  
## 75 EGR1\_N 0.21 1.819861e-50  
## 76 H3MeK4\_N 0.19 7.181599e-44  
## 77 CaNA\_N 0.61 5.372624e-215

AOV\_df$reject\_null <- AOV\_df$p.value < alpha  
alpha\_adj <- alpha / length(unique(Data$class))\*(length(unique(Data$class))-1)/2