RPPA Analysis: Stuart Gallagher

Lucy Liu

Contents

Introduction

The aim of this analysis is to investigate the protein expression of ~52 melanoma cell lines. The data from this experiment will be used to explore a number of research questions:

1. A portion of cell lines (11 cell lines) have been treated with IBET151. The researcher would like to investigate how protein expression differs between treatment with IBET151 vs DMSO.
   (Note: of the 11 cell lines treated with IBET151, there are only 3 that were also treated with DMSO - C106M, A06M & D14M2. All cell lines treated with DMSO were at time point 48h while the IBET151 treated cell lines were at 24h)
2. Is sensitivity of specific cell lines to a certain drug correlated to protein expression?
3. Is the mutation status/transcriptional signature of the melanoma cell lines correlated to its protein expression profile?

Experiment design

There were two biological replicates in this experiment. The samples are detailed below:

Lysate.ID	Cell.Line/Tissue.type	Treatment	Time.point
SJG-1	C078M	DMSO	48
SJG-2	D17M1	DMSO	48
SJG-3	D11M2	DMSO	48
SJG-4	D04M	DMSO	48
SJG-5	D23M	DMSO	48
SJG-6	A15M2	DMSO	48
SJG-7	D20M	DMSO	48
SJG-8	D35M1	DMSO	48
SJG-9	D38M2	DMSO	48
SJG-10	A11M	DMSO	48
SJG-11	A06M	DMSO	48
SJG-12	C044M	DMSO	48
SJG-13	C054M	DMSO	48
SJG-14	C055M	DMSO	48
SJG-15	C058M	DMSO	48
SJG-16	C062M2	DMSO	48
SJG-17	C074M	DMSO	48
SJG-18	C089M	DMSO	48
SJG-19	C027M	DMSO	48
SJG-20	C037M	DMSO	48
SJG-21	C057M	DMSO	48
SJG-22	C094M	DMSO	48
SJG-23	C013M	DMSO	48
SJG-24	C084M	DMSO	48
SJG-25	C086M	DMSO	48
SJG-26	C091M	DMSO	48
SJG-27	C100M	DMSO	48
SJG-28	C001M	DMSO	48
SJG-29	C002M	DMSO	48
SJG-30	C011M	DMSO	48
SJG-31	C017M	DMSO	48
SJG-32	C106M	DMSO	48
SJG-33	C006M	DMSO	48
SJG-34	C021M	DMSO	48

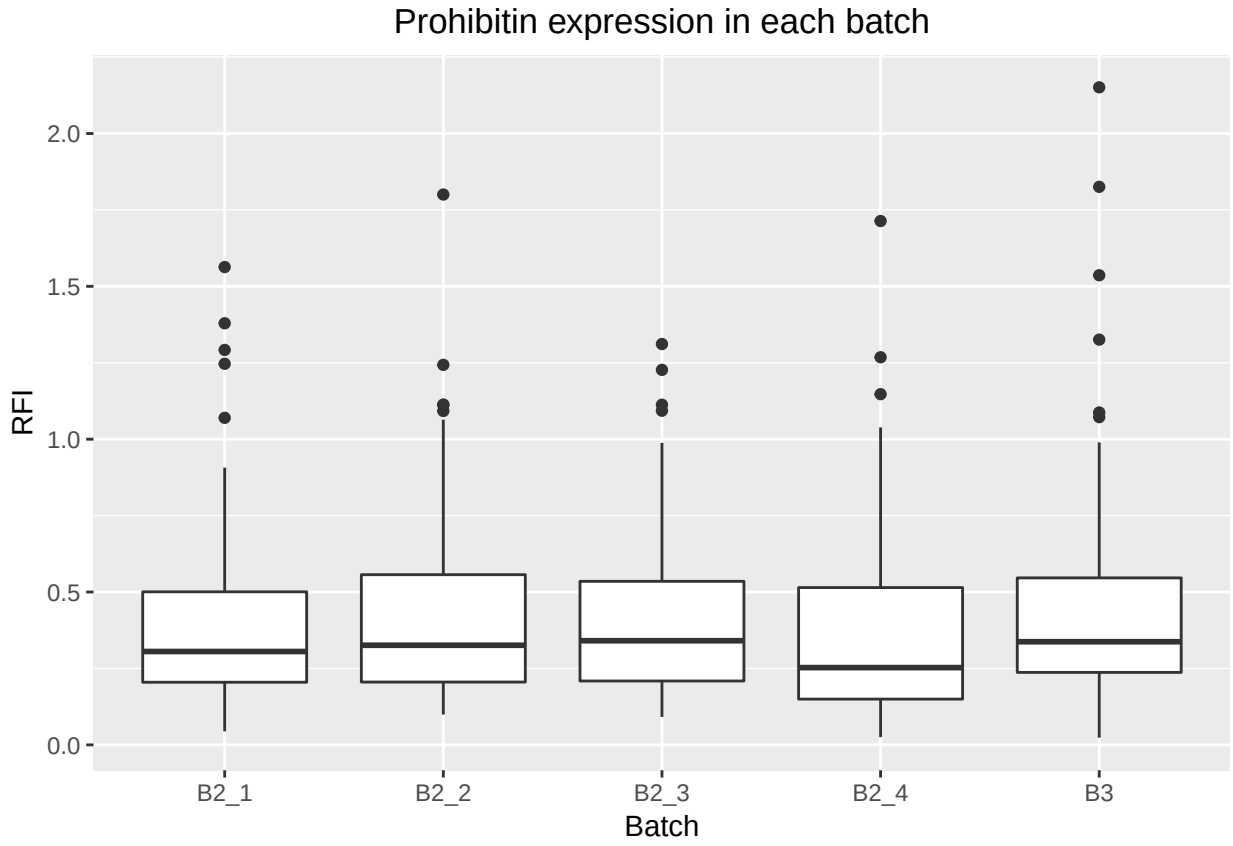
Lysate.ID	Cell.Line/Tissue.type	Treatment	Time.point
SJG-35	C065M	DMSO	48
SJG-36	C076M	DMSO	48
SJG-37	D22M1	DMSO	48
SJG-38	C092M	DMSO	48
SJG-39	A04M	DMSO	48
SJG-40	D36M	DMSO	48
SJG-41	D08M	DMSO	48
SJG-42	A02M1	DMSO	48
SJG-43	D05M1	DMSO	48
SJG-44	D14M2	DMSO	48
SJG-45	D41M	DMSO	48
SJG-46	C060M1	DMSO	48
SJG-47	C088M1	DMSO	48
SJG-48	D24M	DMSO	48
SJG-49	C077M1	DMSO	48
SJG-50	C016M1	DMSO	48
SJG-51	C022M1	DMSO	48
SJG-52	C025M1	DMSO	48
SJG-53	HDF	DMSO	48
SJG-54	HEMn-MP	DMSO	48
SJG-55	HEMn-LP	DMSO	48
SJG-56	CO57M	IBET151	24
SJG-57	CO02M	IBET151	24
SJG-58	C106M	IBET151	24
SJG-59	CO37M	IBET151	24
SJG-60	A06M	IBET151	24
SJG-61	CO92M	IBET151	24
SJG-62	D14M2	IBET151	24
SJG-63	CO65M	IBET151	24
SJG-64	CO25M	IBET151	24
SJG-65	CO44M	IBET151	24
SJG-66	CO76M	IBET151	24
SJG-67	Mel-XY3	DMSO	48
SJG-68	Mel-XY3_Plx-7day	DMSO	168hr
SJG-69	Mel-XY3_Sur	DMSO	48
SJG-70	Mel-XY3_Sur_post	DMSO	48

Batches

There were 2 main batches in the RPPA experiment. Note, in the pilot batch ('Batch 1'), 4 ABs were tested but were all repeated in batch 2. The 2 batches were thus:

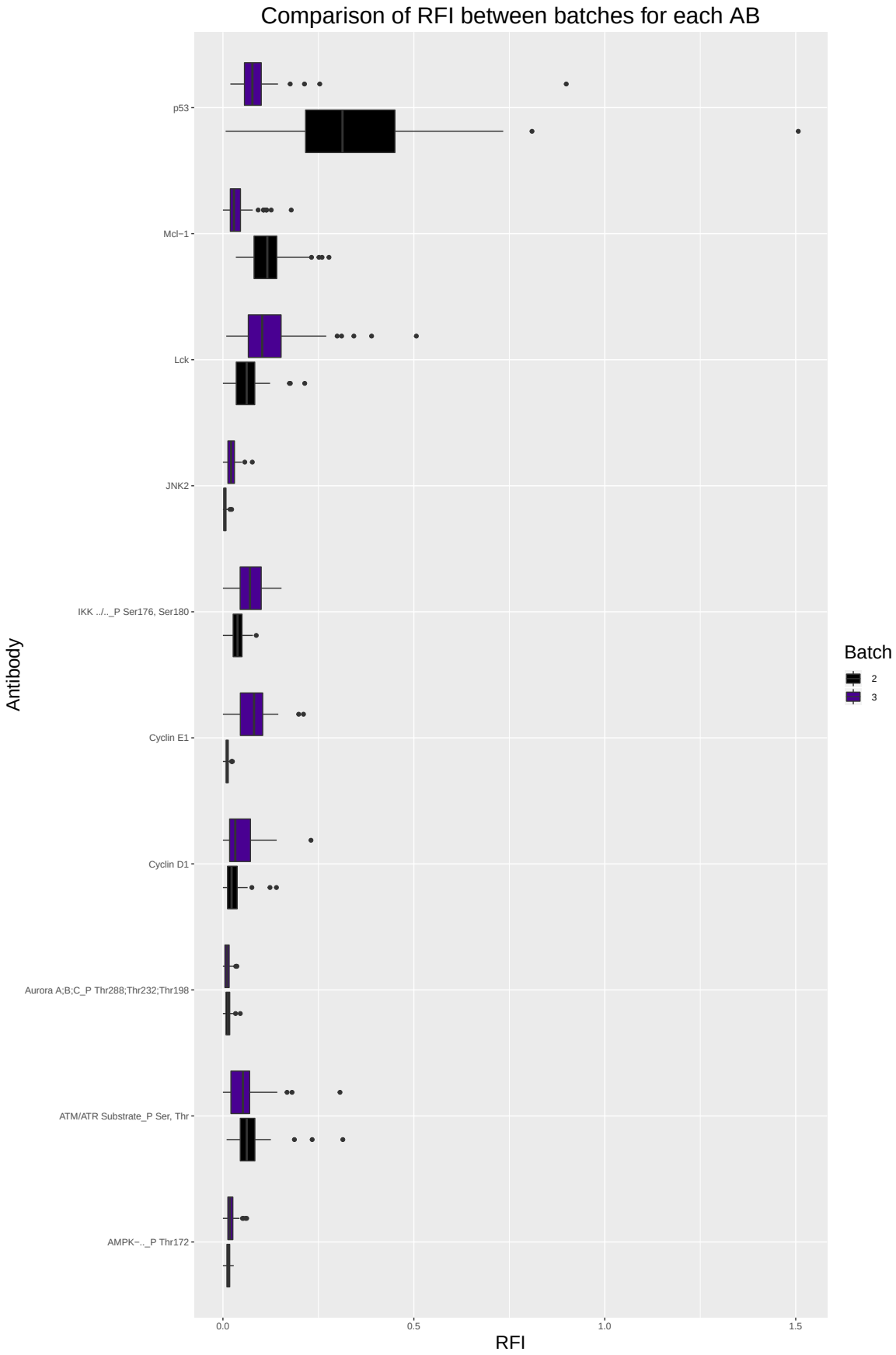
- Batch 2: Full run of all ABs
- Batch 3: Samples and AB's that did not perform well in batch 2 are repeated in batch 3

First, we will compare the RFI values of the positive control AB (Prohibitin) in all batches. Note that Batch 2 was very large and Prohibitin was run 4 times, twice in the first run of ABs and twice in the second run of ABs.

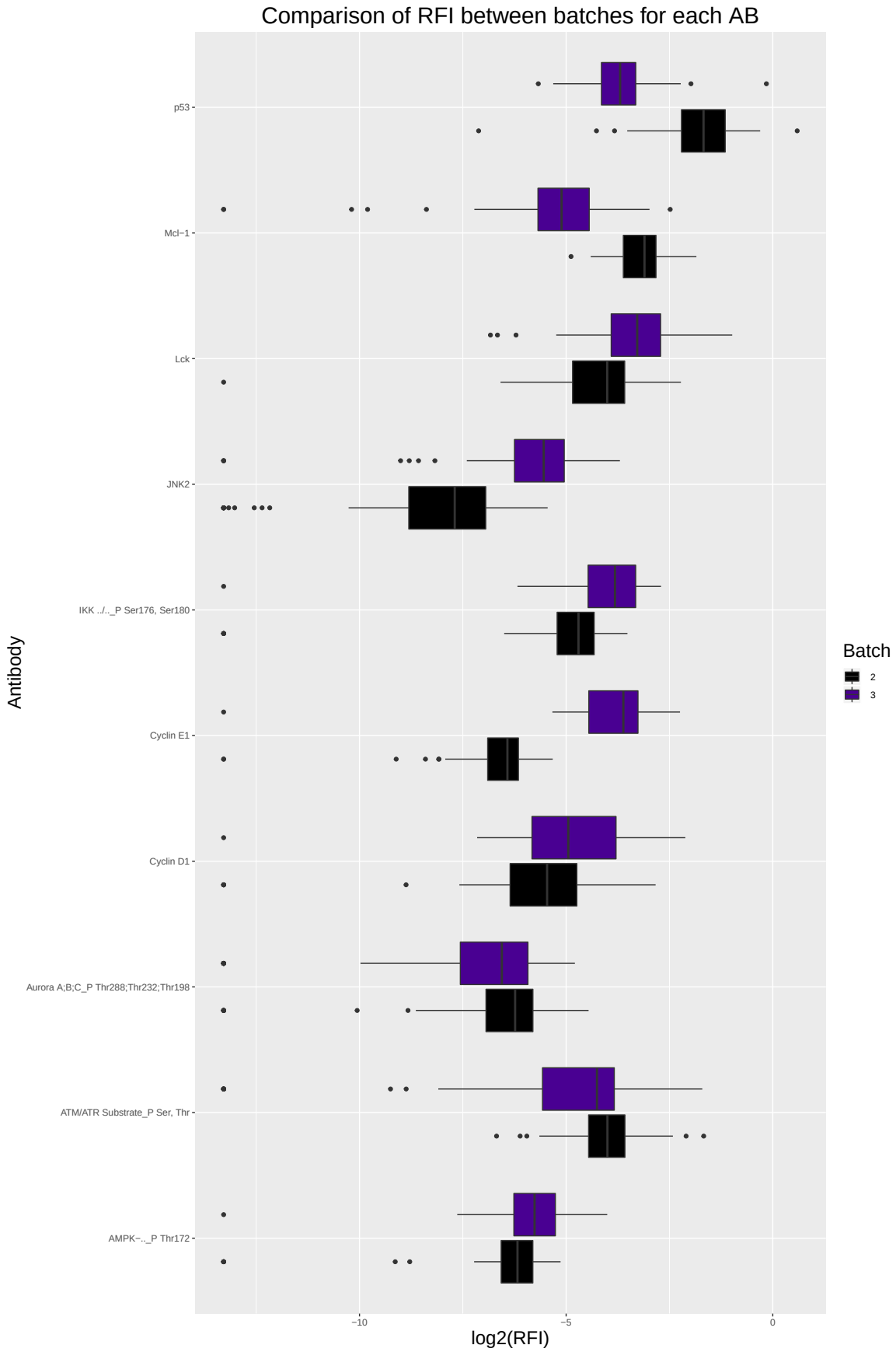


The prohibitin expression across all batches are quite similar, which suggests that there were no significant batch effects.

Most ABs were tested soley within batch 2. This makes it difficult to investigate batch effects as we do not know if the difference in RFI values between batches is due to between array variation or differences in protein expression. There were 10 ABs tested on both batch 2 and 3 (e.g. run in batch 2 for replicate 1 samples (i.e. SJG1-SJG70) and in batch 2 for replicate 2 samples (i.e.SJG71-SJG140)). We compare the RFI values of these ABs between the two batches.


Comparison of RFI between batches for each AB
Antibody
p53
Mcl−1
Lck
JNK2
IKK ../.._P Ser176, Ser180
Cyclin E1
Cyclin D1
Aurora A;B;C_P Thr288;Thr232;Thr198
ATM/ATR Substrate_P Ser, Thr
AMPK−.._P Thr172
0.0
0.5
1.0
1.5
RFI
Batch
2
3

Log RFI values and replot to better visualise ABs with lower RFI values:

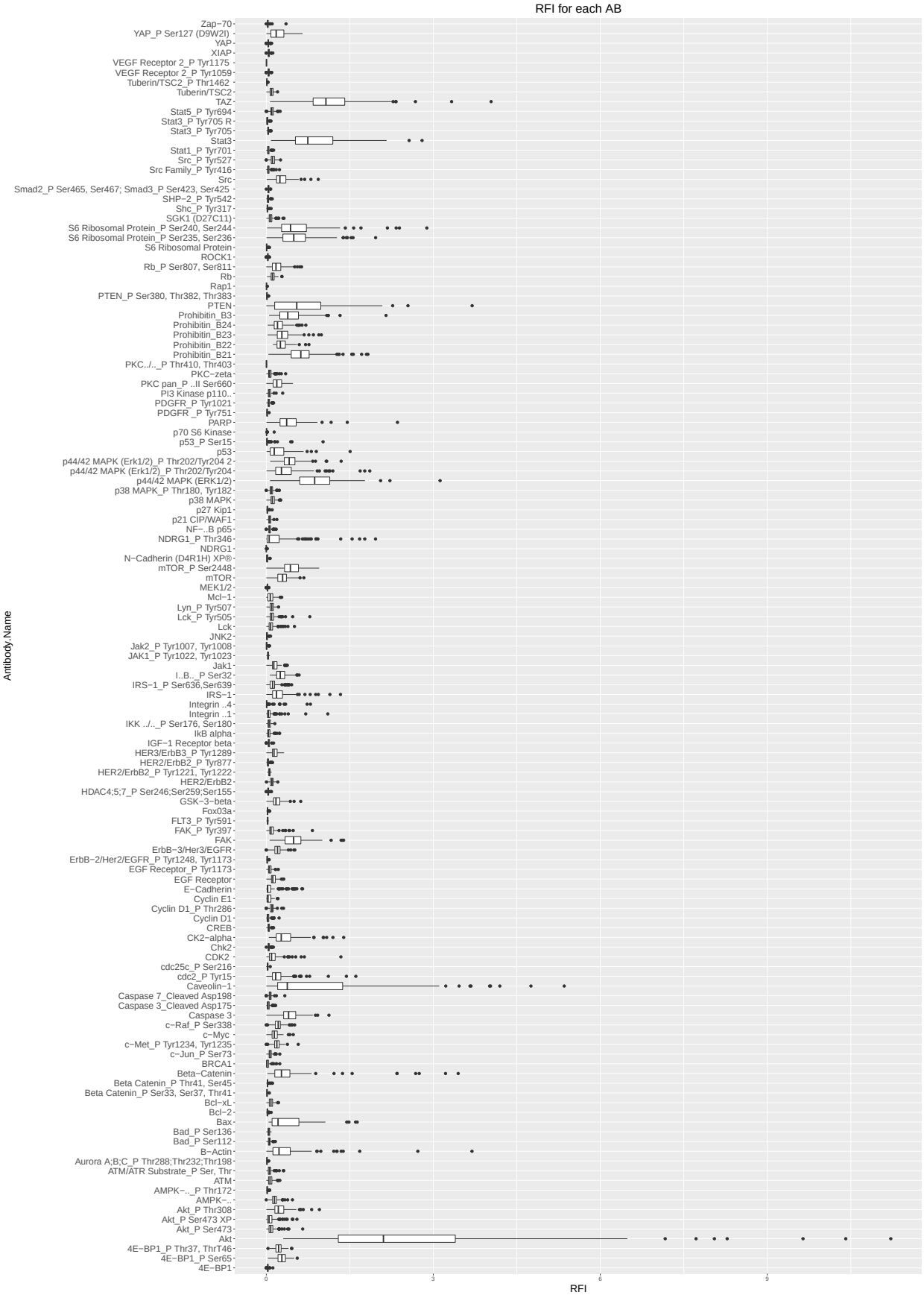

Comparison of RFI between batches for each AB
Antibody
p53
Mcl−1
Lck
JNK2
IKK ../.._P Ser176, Ser180
Cyclin E1
Cyclin D1
Aurora A;B;C_P Thr288;Thr232;Thr198
ATM/ATR Substrate_P Ser, Thr
AMPK−.._P Thr172
−10
−5
0
log2(RFI)
Batch
2
3

Overall, the concordance between batches is not great. However, as mentioned above, it would be difficult to correct for batch effects as the majority of ABs were tested soley in one batch only (thus we do not know if the difference in RFI values between batches is due to between array variation or differences in protein expression). Further, as the vast majority of ABs were tested in batch 2 and the data will be normalised overall for proteins effects, batch correction will not be further considered.

Data exploration

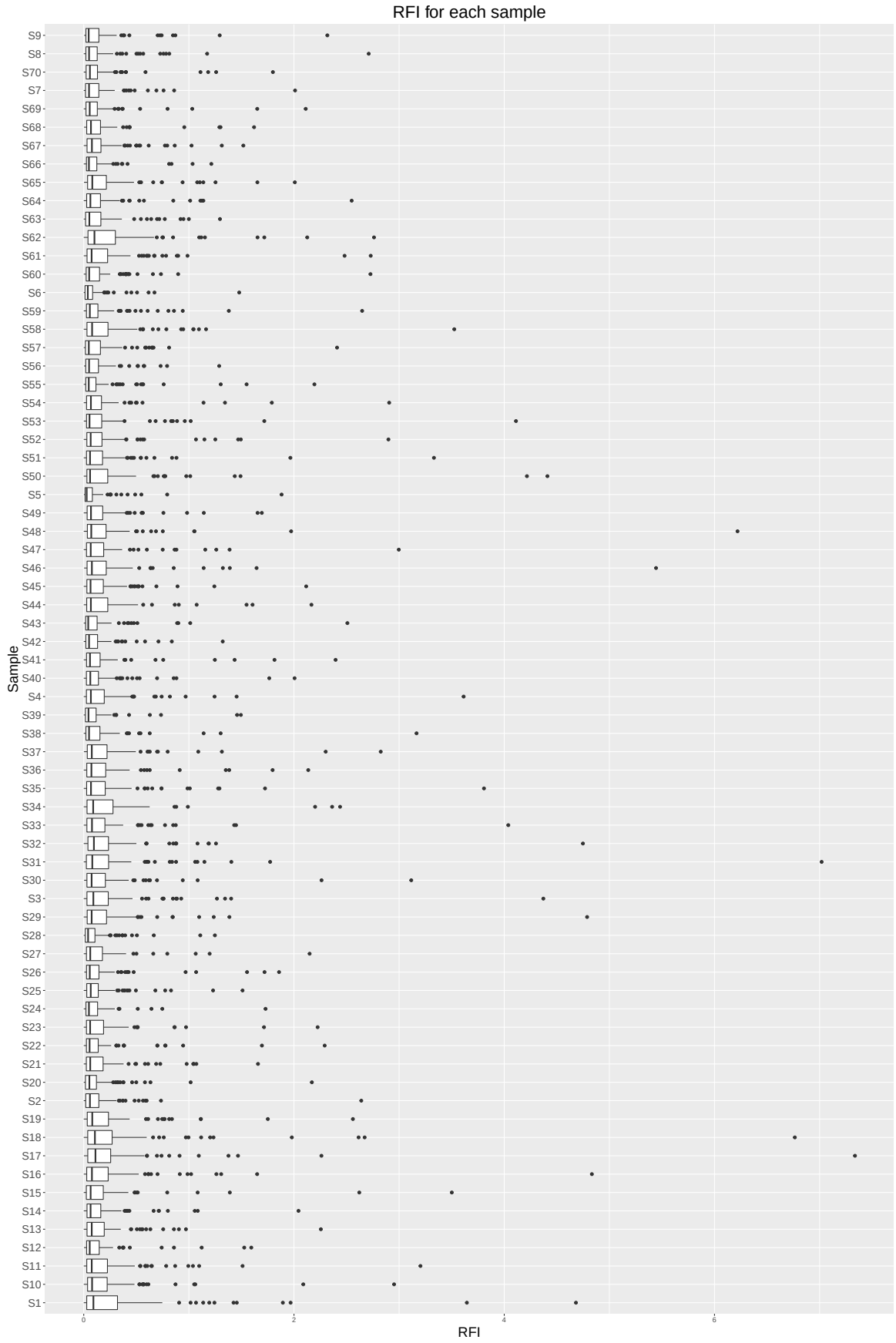
RFI for each protein

A boxplot of the raw RFI values of all samples (both replicates) are plotted for each AB.


RFI for each AB
Antibody Name
Zap−70
YAP_P Ser127 (D9W2I)
YAP
XIAP
VEGF Receptor 2_P Tyr1175
VEGF Receptor 2_P Tyr1059
Tuberin/TSC2_P Thr1462
Tuberin/TSC2
TAZ
Stat5_P Tyr694
Stat3_P Tyr705 R
Stat3_P Tyr705
Stat3
Stat1_P Tyr701
Src_P Tyr527
Src Family_P Tyr416
Src
Smad2_P Ser465, Ser467; Smad3_P Ser423, Ser425
SHP−2_P Tyr542
Shc_P Tyr317
SGK1 (D27C11)
S6 Ribosomal Protein_P Ser240, Ser244
S6 Ribosomal Protein_P Ser235, Ser236
S6 Ribosomal Protein
ROCK1
Rb_P Ser807, Ser811
Rb
Rap1
PTEN_P Ser380, Thr382, Thr383
PTEN
Prohibitin_B3
Prohibitin_B24
Prohibitin_B23
Prohibitin_B22
Prohibitin_B21
PKC../.._P Thr410, Thr403
PKC−zeta
PKC pan_P ..II Ser660
PI3 Kinase p110..
PDGFR_P Tyr1021
PDGFR _P Tyr751
PARP
p70 S6 Kinase
p53_P Ser15
p53
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204 2
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204
p44/42 MAPK (ERK1/2)
p38 MAPK_P Thr180, Tyr182
p38 MAPK
p27 Kip1
p21 CIP/WAF1
NF−..B p65
NDRG1_P Thr346
NDRG1
N−Cadherin (D4R1H) XP®
mTOR_P Ser2448
mTOR
MEK1/2
Mcl−1
Lyn_P Tyr507
Lck_P Tyr505
Lck
JNK2
Jak2_P Tyr1007, Tyr1008
JAK1_P Tyr1022, Tyr1023
Jak1
I..B.._P Ser32
IRS−1_P Ser636,Ser639
IRS−1
Integrin ..4
Integrin ..1
IKK ../.._P Ser176, Ser180
IkB alpha
IGF−1 Receptor beta
HER3/ErbB3_P Tyr1289
HER2/ErbB2_P Tyr877
HER2/ErbB2_P Tyr1221, Tyr1222
HER2/ErbB2
HDAC4;5;7_P Ser246;Ser259;Ser155
GSK−3−beta
Fox03a
FLT3_P Tyr591
FAK_P Tyr397
FAK
ErbB−3/Her3/EGFR
ErbB−2/Her2/EGFR_P Tyr1248, Tyr1173
EGF Receptor_P Tyr1173
EGF Receptor
E−Cadherin
Cyclin E1
Cyclin D1_P Thr286
Cyclin D1
CREB
CK2−alpha
Chk2
CDK2
cdc25c_P Ser216
cdc2_P Tyr15
Caveolin−1
Caspase 7_Cleaved Asp198
Caspase 3_Cleaved Asp175
Caspase 3
c−Raf_P Ser338
c−Myc
c−Met_P Tyr1234, Tyr1235
c−Jun_P Ser73
BRCA1
Beta−Catenin
Beta Catenin_P Thr41, Ser45
Beta Catenin_P Ser33, Ser37, Thr41
Bcl−xL
Bcl−2
Bax
Bad_P Ser136
Bad_P Ser112
B−Actin
Aurora A;B;C_P Thr288;Thr232;Thr198
ATM/ATR Substrate_P Ser, Thr
ATM
AMPK−.._P Thr172
AMPK−..
Akt_P Thr308
Akt_P Ser473 XP
Akt_P Ser473
Akt
4E−BP1_P Thr37, ThrT46
4E−BP1_P Ser65
4E−BP1
0
3
6
9
RFI

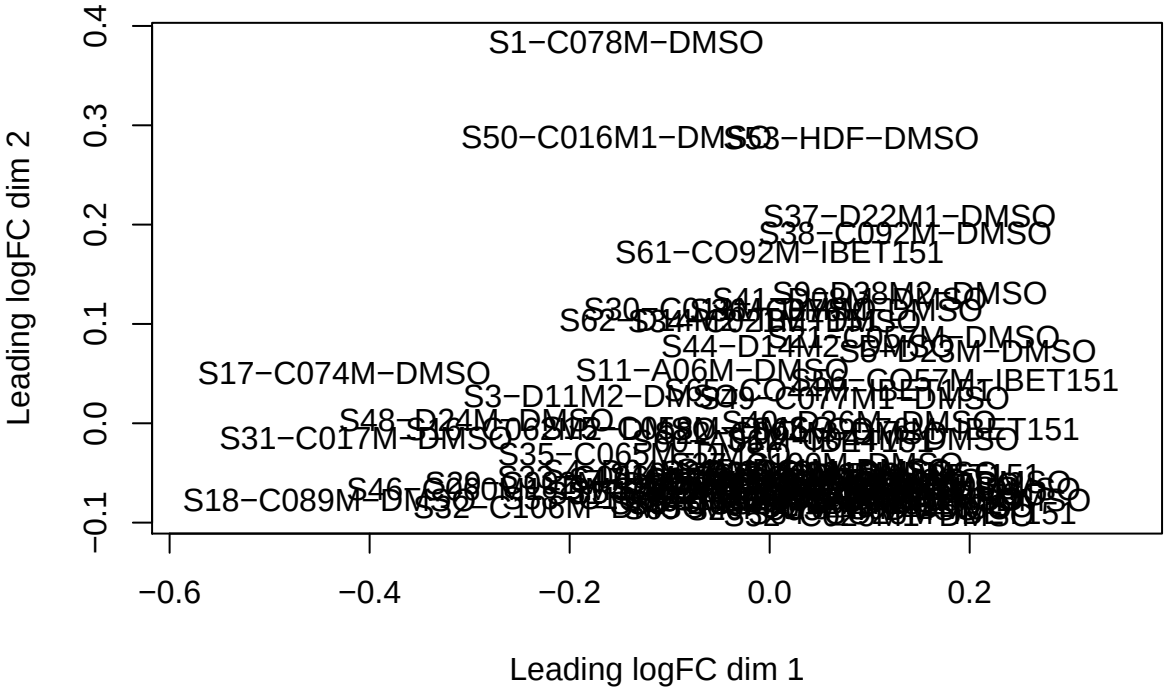
RFI for each sample

Average RFI from the two replicates are plotted for each sample.

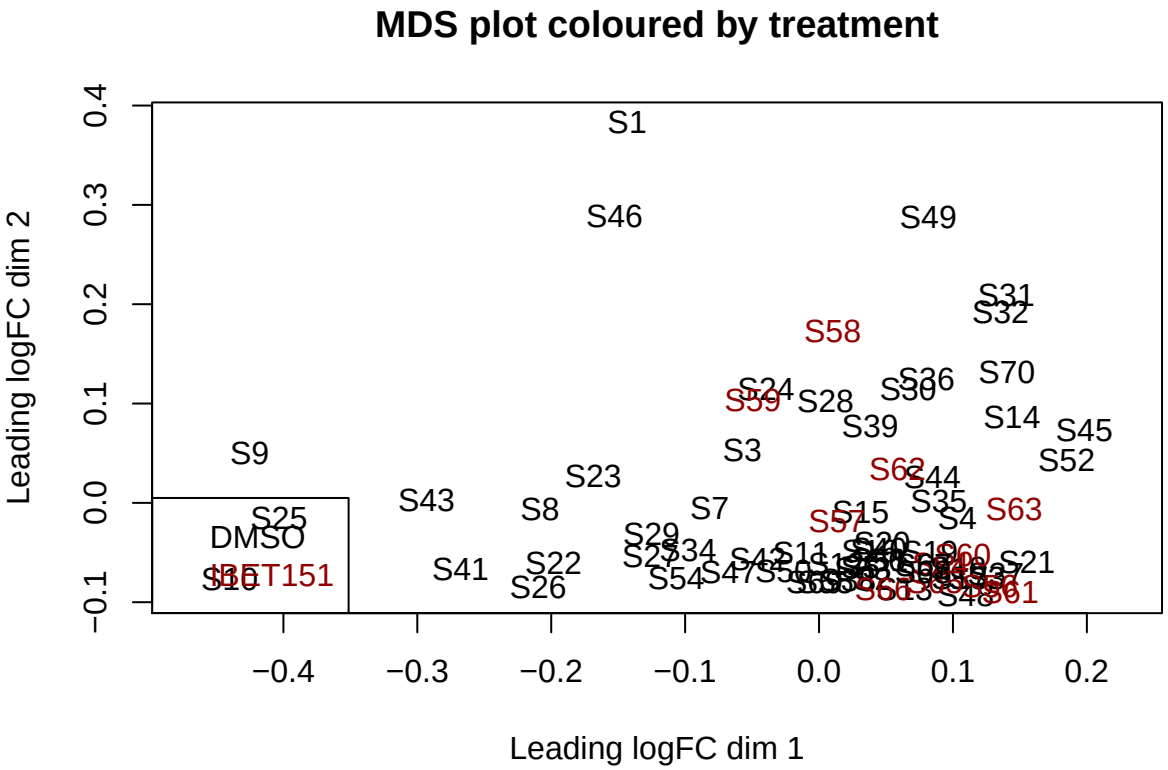

RFI for each sample
Sample
S9
S8
S70
S7
S69
S68
S67
S66
S65
S64
S63
S62
S61
S60
S6
S59
S58
S57
S56
S55
S54
S53
S52
S51
S50
S5
S49
S48
S47
S46
S45
S44
S43
S42
S41
S40
S4
S39
S38
S37
S36
S35
S34
S33
S32
S31
S30
S3
S29
S28
S27
S26
S25
S24
S23
S22
S21
S20
S2
S19
S18
S17
S16
S15
S14
S13
S12
S11
S10
S1
0
2
4
6
RFI

MDS plot of samples

A multidimensional scaling (MDS) plot graphically represents relationships between objects. The distance between two samples approximates their similarity or dissimilarity. The MDS plot is useful for identifying if there are any distinct groups within our samples.



Re-plot MDS using sample names S1-S70 so labels are more legible, and colour sample names by treatment (DMSO or IBET151).



There does not appear to be distinct groups of cell lines within the experiment. The DMSO and IBET151 treated cell lines also do not appear to separate. This is not unexpected as most of the cell lines in the experiment are different (there are only three cell lines that feature twice in the experiment). This may be the reason that the IBET151 treated cell lines do not cluster separately from the DMSO treated cell lines.

Heatmap

Heatmap and dendogram of raw RFI values.


7
6
5
4
3
2
1
0
Akt
B−Actin
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204
CDK2
E−Cadherin
Akt_P Ser473
Akt_P Ser473 XP
Lyn_P Tyr507
ATM
SGK1 (D27C11)
c−Jun_P Ser73
Mcl−1
PKC−zeta
Bcl−xL
Caspase 7_Cleaved Asp198
Lck
Tuberin/TSC2
HER2/ErbB2
Stat5_P Tyr694
Cyclin D1_P Thr286
p38 MAPK_P Thr180, Tyr182
FAK_P Tyr397
Lck_P Tyr505
EGF Receptor
p38 MAPK
Rb
Src_P Tyr527
AMPK−..
c−Myc
IRS−1_P Ser636,Ser639
HER3/ErbB3_P Tyr1289
Jak1
p53_P Ser15
Zap−70
Jak2_P Tyr1007, Tyr1008
S6 Ribosomal Protein
p70 S6 Kinase
NDRG1
PKC../.._P Thr410, Thr403
Rap1
VEGF Receptor 2_P Tyr1175
N−Cadherin (D4R1H) XP®
Aurora A;B;C_P Thr288;Thr232;Thr198
Tuberin/TSC2_P Thr1462
JNK2
PTEN_P Ser380, Thr382, Thr383
Beta Catenin_P Ser33, Ser37, Thr41
MEK1/2
ErbB−2/Her2/EGFR_P Tyr1248, Tyr1173
PDGFR _P Tyr751
Stat3_P Tyr705 R
AMPK−._P Thr172
p27 Kip1
Shc_P Tyr317
Cyclin D1
BRCA1
YAP
JAK1_P Tyr1022, Tyr1023
SHP−2_P Tyr542
Beta Catenin_P Thr41, Ser45
4E−BP1
Bcl−2
cdc25c_P Ser216
FLT3_P Tyr591
Fox03a
ROCK1
HER2/ErbB2_P Tyr877
Stat1_P Tyr701
HDAC4;5;7_P Ser246;Ser259;Ser155
Stat3_P Tyr705
ATM/ATR Substrate_P Ser, Thr
IkB alpha
PI3 Kinase p110..
Caspase 3_Cleaved Asp175
Cyclin E1
Src Family_P Tyr416
Smad2_P Ser465, Ser467; Smad3_P Ser423, Ser425
Chk2
XIAP
Bad_P Ser112
EGF Receptor_P Tyr1173
NF−..B p65
p21 CIP/WAF1
IKK ../.._P Ser176, Ser180
HER2/ErbB2_P Tyr1221, Tyr1222
IGF−1 Receptor beta
CREB
VEGF Receptor 2_P Tyr1059
Bad_P Ser136
PDGFR_P Tyr1021
Integrin ..1
Integrin ..4
NDRG1_P Thr346
CK2−alpha
Prohibitin_B23
Prohibitin_B24
mTOR
4E−BP1_P Ser65
I..B.._P Ser32
Akt_P Thr308
Src
cdc2_P Tyr15
IRS−1
p53
GSK−3−beta
YAP_P Ser127 (D9W2I)
4E−BP1_P Thr37, ThrT46
c−Met_P Tyr1234, Tyr1235
c−Raf_P Ser338
ErbB−3/Her3/EGFR
PKC pan_P ..II Ser660
Rb_P Ser807, Ser811
Beta−Catenin
PTEN
Prohibitin_B21
S6 Ribosomal Protein_P Ser235, Ser236
S6 Ribosomal Protein_P Ser240, Ser244
Bax
PARP
Prohibitin_B3
FAK
Caspase 3
mTOR_P Ser2448
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204 2
Caveolin−1
TAZ
p44/42 MAPK (ERK1/2)
Stat3
S18−C08M−DMSO
S17−C07AM−DMSO
S31−C017M−DMSO
S16−C062M2−DMSO
S48−D24M−DMSO
S29−C02M−DMSO
S46−C060M1−DMSO
S37−D22M1−DMSO
S53−HDF−DMSO
S38−C092M−DMSO
S1−C078M−DMSO
S50−C016M1−DMSO
S32−C108M−DMSO
S58−C108M−IBET151
S3−D11M2−DMSO
S33−C009M−DMSO
S4−D04M−DMSO
S35−C059M−DMSO
S51−C022M1−DMSO
S47−C088M1−DMSO
S8−D39M1−DMSO
S20−C037M−DMSO
S59−C037M−IBET151
S19−C027M−DMSO
S22−C094M−DMSO
S55−HEMn−MP−DMSO
S10−A11M−DMSO
S52−C025M−DMSO
S5−D22M−DMSO
S9−D38M2−DMSO
S31−C057M−DMSO
S56−C057M−IBET151
S2−C076M−IBET151
S60−A06M−IBET151
S43−D05M1−DMSO
S57−CO2O2M−IBET151
S23−C013M−DMSO
S27−C100M−DMSO
S45−D41M−DMSO
S13−C054M−DMSO
S14−C059M−DMSO
S7−D20M−DMSO
S24−C084M−DMSO
S6−A15M2−DMSO
S28−C001M−DMSO
S42−A02M1−DMSO
S25−C088M−DMSO
S39−A04M−DMSO
S68−Mei−JY3_Pik−Tida−DMSO
S69−Mei−JY3_Sur−DMSO
S63−C096M−IBET151
S67−Mei−JY3−DMSO
S70−Mei−JY3_Sur_post−DMSO
S26−C091M−DMSO
S49−C077M1−DMSO
S12−C044M−DMSO
S40−D36M−DMSO
S34−C021M−DMSO
S61−C020M−IBET151
S36−D08M−DMSO
S62−D14M2−IBET151
S65−C044M−IBET151
S15−C058M−DMSO
S44−D1AM2−DMSO
S11−A09M−DMSO
S30−C011M−DMSO

To aid in better discriminating the range of RFI values, a heatmap of log 2 RFI values is plotted as well.


NDRG1
Jak2_P Tyr1007, Tyr1008
N−Cadherin (D4R1H) XP®
Integrin ..4
p70 S6 Kinase
S6 Ribosomal Protein
PKC../.._P Thr410, Thr403
Rap1
VEGF Receptor 2_P Tyr1175
Akt
TAZ
p44/42 MAPK (ERK1/2)
Stat3
Caveolin−1
PTEN
S6 Ribosomal Protein_P Ser235, Ser236
S6 Ribosomal Protein_P Ser240, Ser244
Prohibitin_B3
FAK
Prohibitin_B21
B−Actin
IRS−1
cdc2_P Tyr15
Rb_P Ser807, Ser811
Beta−Catenin
p53
GSK−3−beta
YAP_P Ser127 (D9W2I)
c−Met_P Tyr1234, Tyr1235
4E−BP1_P Thr37, ThrT46
c−Raf_P Ser338
ErbB−3/Her3/EGFR
Prohibitin_B23
Prohibitin_B24
I..B.._P Ser32
4E−BP1_P Ser65
mTOR
Akt_P Thr308
Src
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204
CK2−alpha
Bax
PARP
Caspase 3
mTOR_P Ser2448
p44/42 MAPK (Erk1/2)_P Thr202/Tyr204 2
Beta Catenin_P Ser33, Ser37, Thr41
MEK1/2
Aurora A;B;C_P Thr288;Thr232;Thr198
Tuberin/TSC2_P Thr1462
PTEN_P Ser380, Thr382, Thr383
JNK2
PDGFR _P Tyr751
p53_P Ser15
BRCA1
Cyclin D1
Stat3_P Tyr705 R
Beta Catenin_P Thr41, Ser45
cdc25c_P Ser216
Fox03a
AMPK−.._P Thr172
ErbB−2/Her2/EGFR_P Tyr1248, Tyr1173
Bcl−2
p27 Kip1
Shc_P Tyr317
Zap−70
Smad2_P Ser465, Ser467; Smad3_P Ser423, Ser425
Caspase 3_Cleaved Asp175
JAK1_P Tyr1022, Tyr1023
FLT3_P Tyr591
SHP−2_P Tyr542
4E−BP1
HER2/ErbB2_P Tyr877
YAP
Stat1_P Tyr701
ROCK1
HDAC4;5;7_P Ser246;Ser259;Ser155
Stat3_P Tyr705
NDRG1_P Thr346
Bad_P Ser112
PKC pan_P ..II Ser660
E−Cadherin
CDK2
Lck_P Tyr505
p38 MAPK_P Thr180, Tyr182
Tuberin/TSC2
p38 MAPK
EGF Receptor
Src_P Tyr527
Rb
Stat5_P Tyr694
Cyclin D1_P Thr286
HER2/ErbB2
IRS−1_P Ser636,Ser639
HER3/ErbB3_P Tyr1289
Jak1
AMPK−..
c−Myc
Lyn_P Tyr507
Akt_P Ser473
FAK_P Tyr397
c−Jun_P Ser73
EGF Receptor_P Tyr1173
Caspase 7_Cleaved Asp198
NF−..B p65
Lck
SGK1 (D27C11)
Mcl−1
Bcl−xL
PKC−zeta
Integrin ..1
Akt_P Ser473 XP
ATM
IkB alpha
PI3 Kinase p110..
ATM/ATR Substrate_P Ser, Thr
IKK../.._P Ser176, Ser180
HER2/ErbB2_P Tyr1221, Tyr1222
p21 CIP/WAF1
CREB
VEGF Receptor 2_P Tyr1059
IGF−1 Receptor beta
Bad_P Ser136
PDGFR_P Tyr1021
Chk2
XIAP
Cyclin E1
Src Family_P Tyr416
0
−5
−10
S7−D20M−DMSO
S6−A13M2−2hMSO
S28−C001M−DMSO
S8−D35M1−DMSO
S37−D22M1−DMSO
S19−C027M−DMSO
S62−D14M2−IBET151
S41−D08M−DMSO
S50−C016M1−DMSO
S1−C078M−DMSO
S31−C017M−DMSO
S61−C092M−IBET151
S38−C092M−DMSO
S49−C077M1−DMSO
S4−D04M−DMSO
S36−C076M−DMSO
S66−C076M−IBET151
S40−D36M−DMSO
S60−A06M−IBET151
S44−D14M2−DMSO
S11−A06M−DMSO
S58−Mel−XY3_Sur−7day−DMSO
S12−C044M−DMSO
S26−C091M−DMSO
S65−C044M−IBET151
S46−C060M1−DMSO
S17−C074M−DMSO
S30−D11M2−DMSO
S3−D11M2−DMSO
S35−C065M−DMSO
S34−C021M−DMSO
S48−D24M−DMSO
S47−C088M1−DMSO
S10−A11M−DMSO
S32−C106M−DMSO
S59−C037M−IBET151
S64−C025M−IBET151
S27−C100M−DMSO
S67−Mel−XY3−DMSO
S13−C054M−DMSO
S14−C055M−DMSO
S52−C025M1−DMSO
S58−C106M−IBET151
S18−C088M−DMSO
S15−C098M−DMSO
S16−C012M−DMSO
S23−C012M−DMSO
S29−C002M−DMSO
S25−C086M−DMSO
S31−C002M−DMSO
S33−C006M−DMSO
S51−C022M1−DMSO
S53−HEP−DMSO
S54−HEMn−MP−DMSO
S55−HEMn−LP−DMSO
S43−D09M1−DMSO
S22−C094M−DMSO
S24−C084M−DMSO
S5−D23M−DMSO
S56−C057M−IBET151
S9−D38M2−DMSO
S20−C037M−DMSO
S69−Mel−XY3_Sur−DMSO
S57−C002M−IBET151
S42−A02M1−DMSO
S70−Mel−XY3_Sur_post−DMSO
S63−C068M−IBET151
S2−D17M1−DMSO
S39−A04M−DMSO

Normalisation

There are three primary factors that need to be normalised for in an RPPA experiment (Liu et al. 2014,Wachter et al. (2015)):

1. Spatial bias: Differences in intensity caused by location of the lysate spot on the slide (e.g. rim effects). The Zepto system already accounts for this using the BSA control spots, thus spatial normalisation will not be further considered.
2. Total amount of protein (loading) of different samples on the slides: Although the total proteins in the lysate are gauged before they are printed, this is confounded by lipids and other biological materials in the samples. Thus, the total protein measurement is only a rough estimate.
3. Non linearity of variances: A MA (differences versus means) plot of the differences (M) in intensity of ABs between two samples varies across the spectrum of A, the mean intensity of the two samples. This is especially the case at the upper range and sometimes the lower range of A. This can be seen best when comparing a pair of technical replicates. As there would not be any differential expression, you would expect their intensities to be equal. Some imbalance often occurs and this imbalance may vary depending on the average intensity (A). Variation in the imbalance When normalised well all the points of a MA plot should align with a horizontal line and be evenly distributed about this line.

Normalisation methods

Housekeeping protein

This normalisation methodology is based on the assumption that the levels of housekeeping (HK) proteins such as β-Actin is uniform across samples and experiment conditions. Thus any differences in level of these housekeeping proteins is due to differing amounts of protein loading. Our previous experience suggests that HK proteins that are effective in western blots are not in RPPA due to the increased sensitivey of RPPA.

Median normalisation

This method assumes that all measured proteins reflect the total protein amount of one sample. Thus the median AB of a sample estimates sample loading. The median value of all AB signals for a sample is used to normalise the raw intensity values for each AB. This is one of the 'simpliest' methods (involving the least steps) however is biased when the number of ABs is <100 (Liu et al. 2014).

Loading control

This approach is utilised by MD Anderson.

Protein effects on intensity are accounted for by dividing all the raw linear intensity by the median for each AB (across all samples). This is the 'median centered ratio'. Sample effects are accounted for by taking the median of the median centered ratios for each sample (across all ABs). This becomes the correction factor for that sample. Raw intensity values are divided by this correction factor to obtain the normalised intensity.

The steps are outlined below:

1. Determine median RFI for each AB (across all samples)
2. Divide each RFI by the median within each AB to get the 'median-centred ratio'.
3. Calculate the median median-cetered ratio for each sample (across all ABs). This is the correction factor for each sample.
4. Divide each median-centred ratio by the correction factor for each sample.

Invariable protein normalisation

This methodology was proposed by Liu et al. (Liu et al. 2014) and was originally suggested to normalise microarray data (Pelz et al. 2008).

The aim of this methodology is to determine the most uniformly expressed proteins and use these as an effective 'housekeeping' protein to normalise to.

The steps are outlined below:

1. Rank the intensity of ABs for each sample so you have ranked ABs for each sample from the highest expressing AB to lowest expressing AB.
2. Calculate the variance of the ranks for each AB. Remove the AB with the highest rank variance.
3. Re-rank the intensity of ABs.
4. Repeat the steps 2 and 3 until the number of remaining markers reaches a predetermined number (100 kept in Liu et al. study).
5. Trim intensities for each AB e.g. the highest and lowest 25% of values are removed from the data set.
6. Average the remaining values of every AB across all samples and use this as a virtual reference sample.
7. Normalise each sample to the virtual reference sample by lowess smoothing using a MA plot. The normalised values are generated using the residuals of the fit.

How to evaluate normalisation methodologies

To be able to evaluate each normalisation methodology, one must be able to compare how effectively each method normalises for total protein loading and non-linearity of variances. Normalisation of non-linearity of variances is effectively assessed using the MA plot. Normalisation of total protein loading is more difficult to assess and must be done indirectly. Two techniques have beens suggested and are discussed below.

MA plot

Normalisation of non-linearity of variances is effectively assessed using a MA plot. A MA plot of replicate samples show the difference in RFI values between the two samples against the average RFI value of the two samples, for each AB.

RLE

RLE (Relative log expression) plots are effective for the assessment for unwanted variation (Gandolfo and Speed 2018). It plots log expression relative to the median of that AB for each sample. Assuming the expression levels for the majority of proteins are stable across cell lines, the boxplots in a RLE plot should be roughly centered on 0 and would be roughly the same size (height).

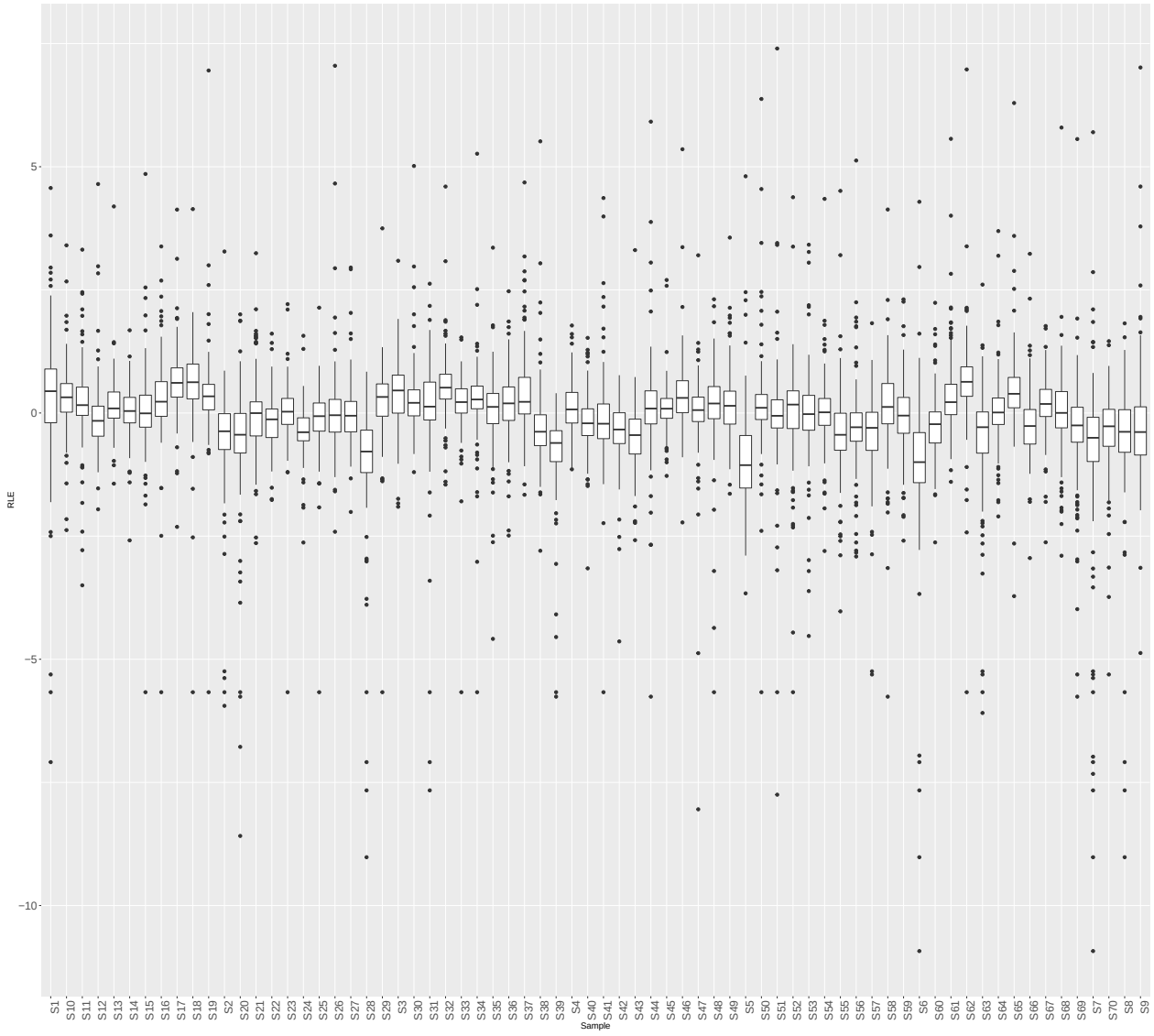
Pairwise correlation coefficients (Spearman ρ)

This plot is used to evaluate the between protein correlations. For each pair of proteins (e.g. for 10 proteins, there would be 45 different possible pairs) the Spearman ρ is calculated across all samples. This is then sorted and plotted. We expect the positive and negative correlation coefficients to be largely equal. If there is an abnormally large number of protein pairs with positive correlation, it may be due to a sample loading effect where some proteins are high in all samples, thus are ranked simiarly highly for those samples. Samples with low loading result in low levels of all proteins, and are ranked similarly low for those samples. There would thus be high correlation between a larger proportion of proteins. We expect protein expression to be inherently different due to expression, phosphorylation levels and/or AB affinity.

RLE plots

Raw data

Plot an RLE graph of the raw data:



There is variation in boxplot position and height and shows taht the data needs to be normalised.

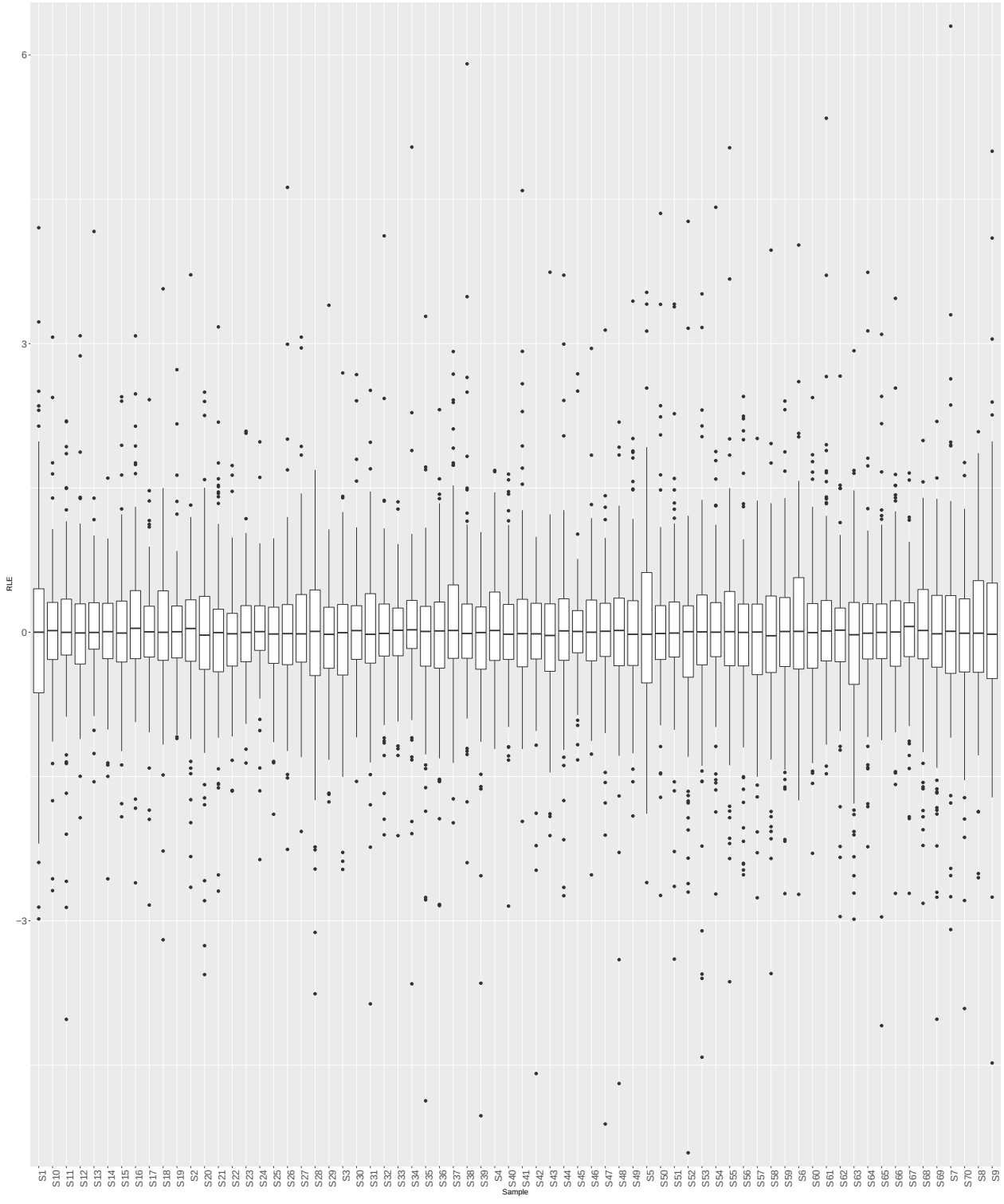
Median normalisation

Median normalisation would not change the above plot as the median value of the median normalised expression values (for each AB) is just 1.

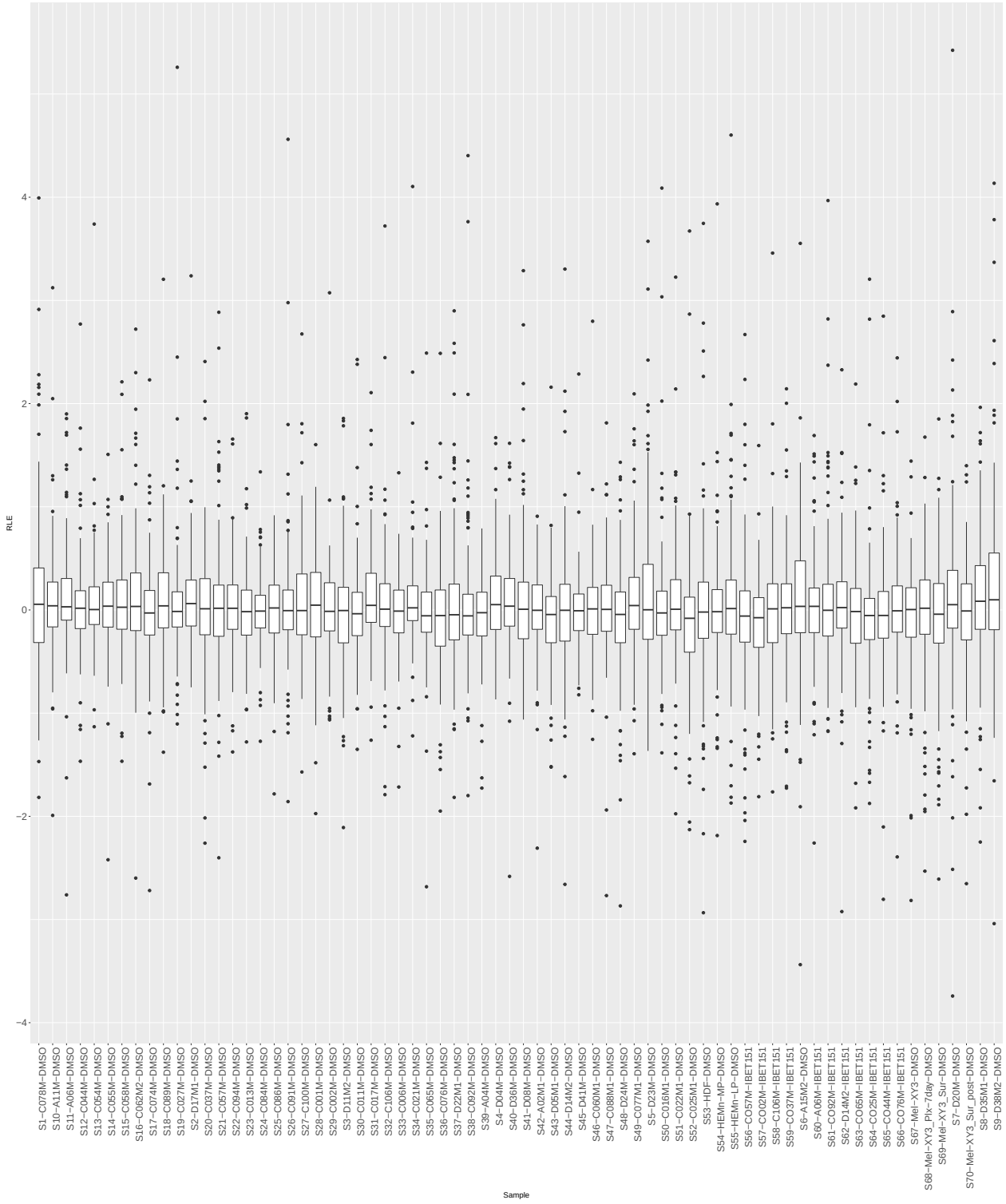
Loading control normalisation

Loading control normalisation causes the median value of each sample to be 0, by virtue of the normalisation approach itself. Nonetheless, the height of each boxplot can be used to assess quality of normalisation. Note

that for many samples there are 1 or 2 very low RLE values which are excluded from the graph to allow easier comparison.



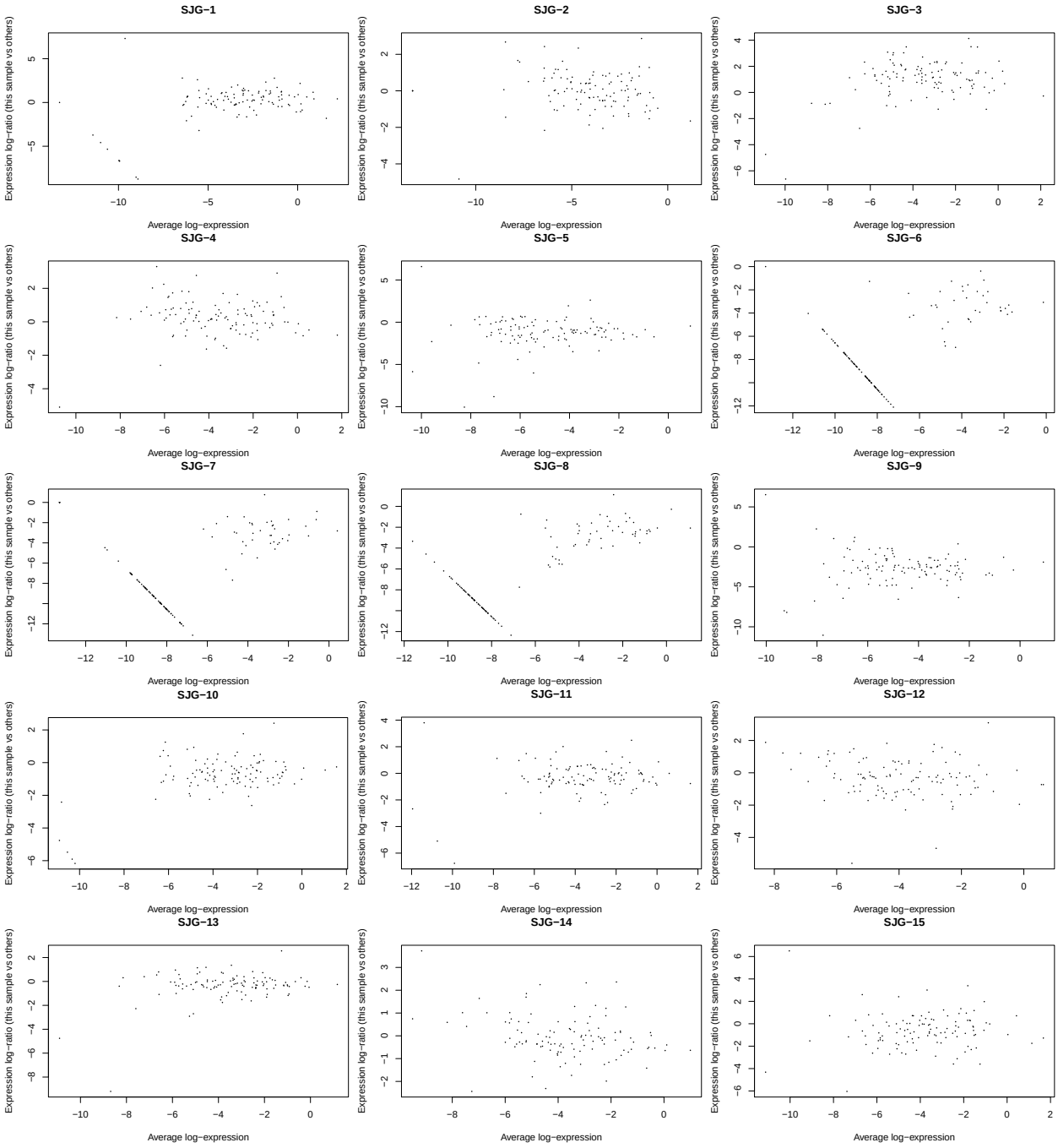
Global rank invariant

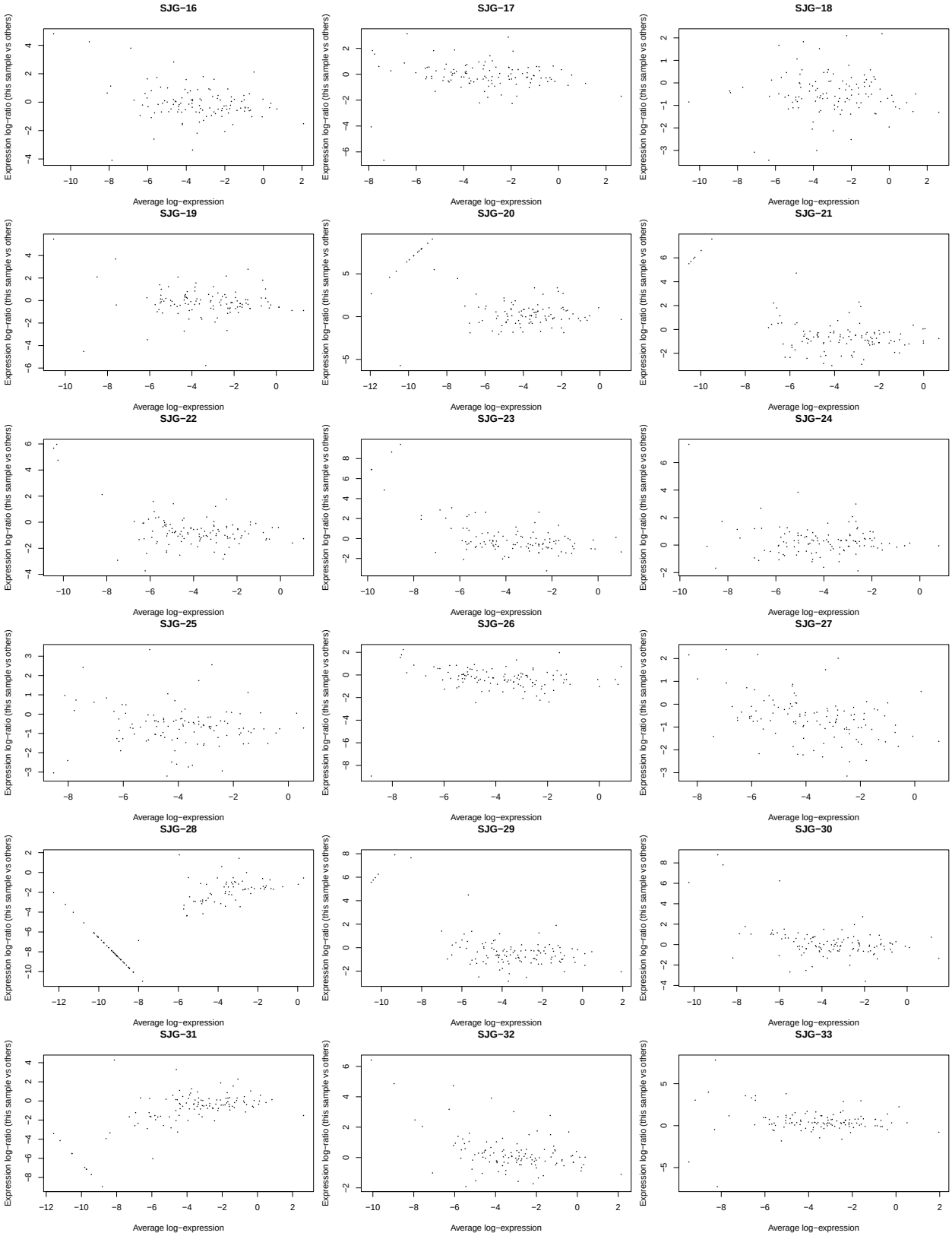


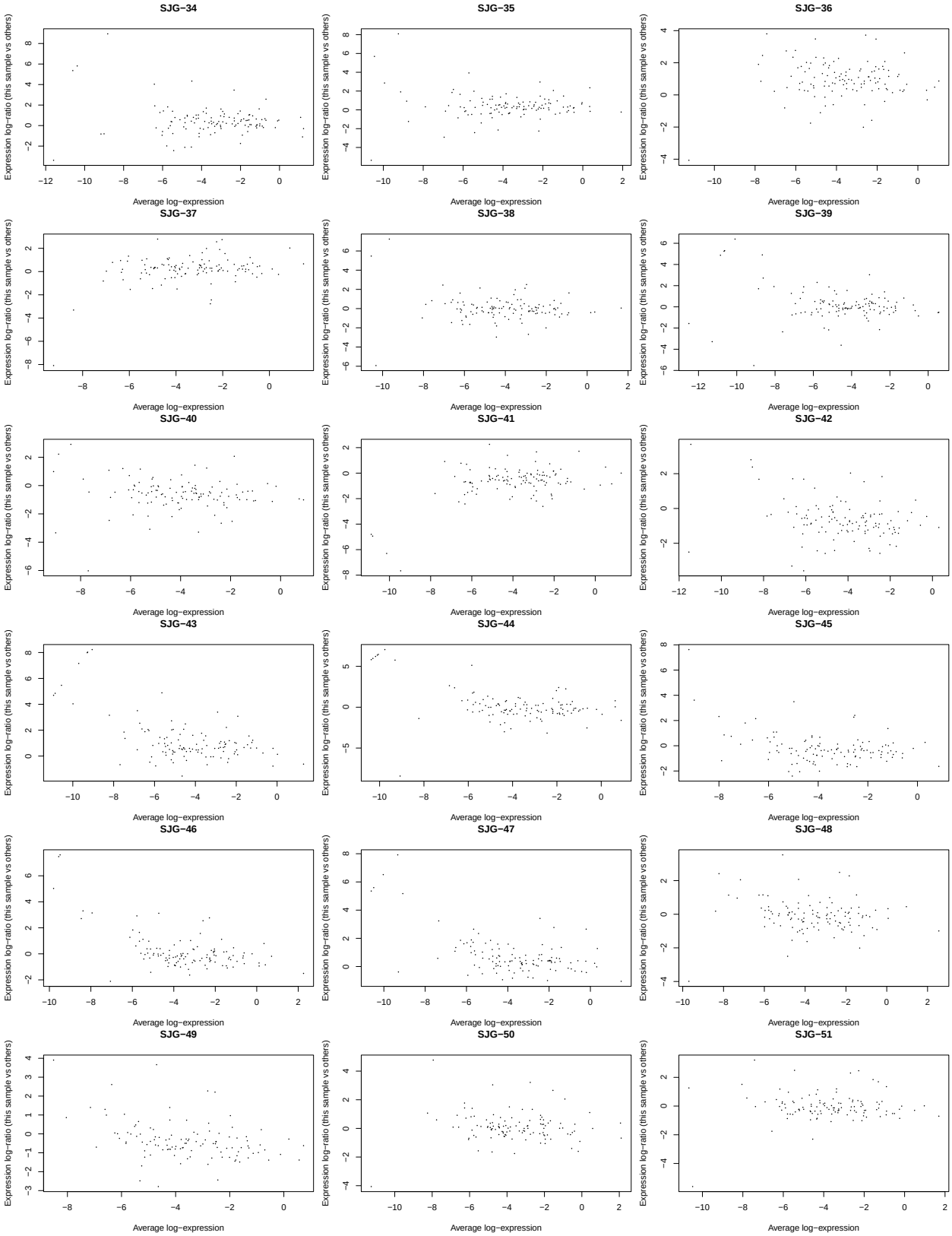
MA plots

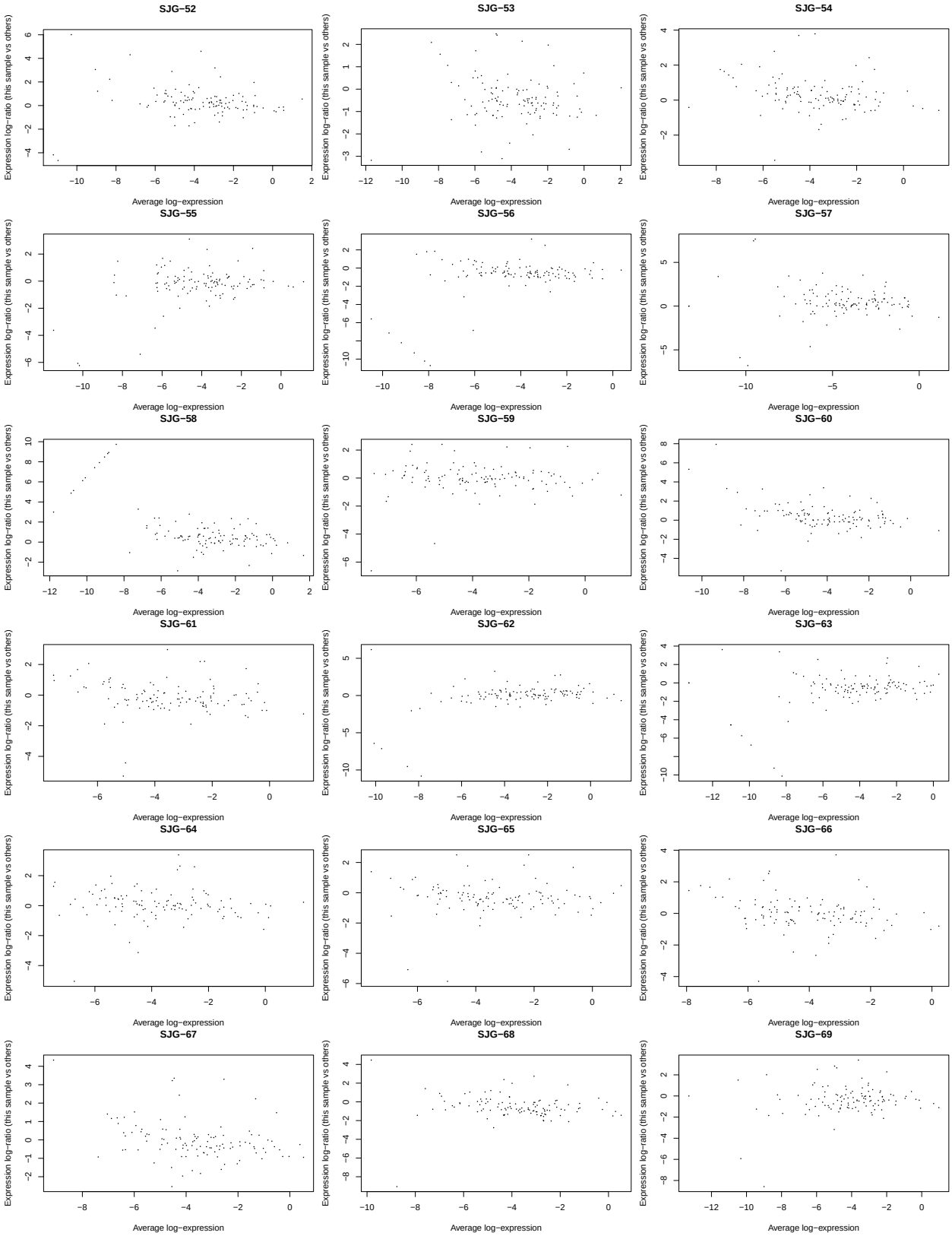
Each point in an MA plot represents an AB. On the y axis is the difference in RFI between two replicates and on the x axis is the mean RFI of the two replicates. For well normalised data, you expect the points to lie along 0 and the variation of points around 0 to be similar along A.

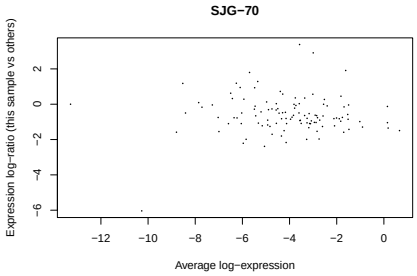
Raw data




SJG−16
SJG−17
SJG−18
SJG−19
SJG−20
SJG−21
SJG−22
SJG−23
SJG−24
SJG−25
SJG−26
SJG−27
SJG−28
SJG−29
SJG−30
SJG−31
SJG−32
SJG−33
Expression log−ratio (this sample vs others)
Average log−expression

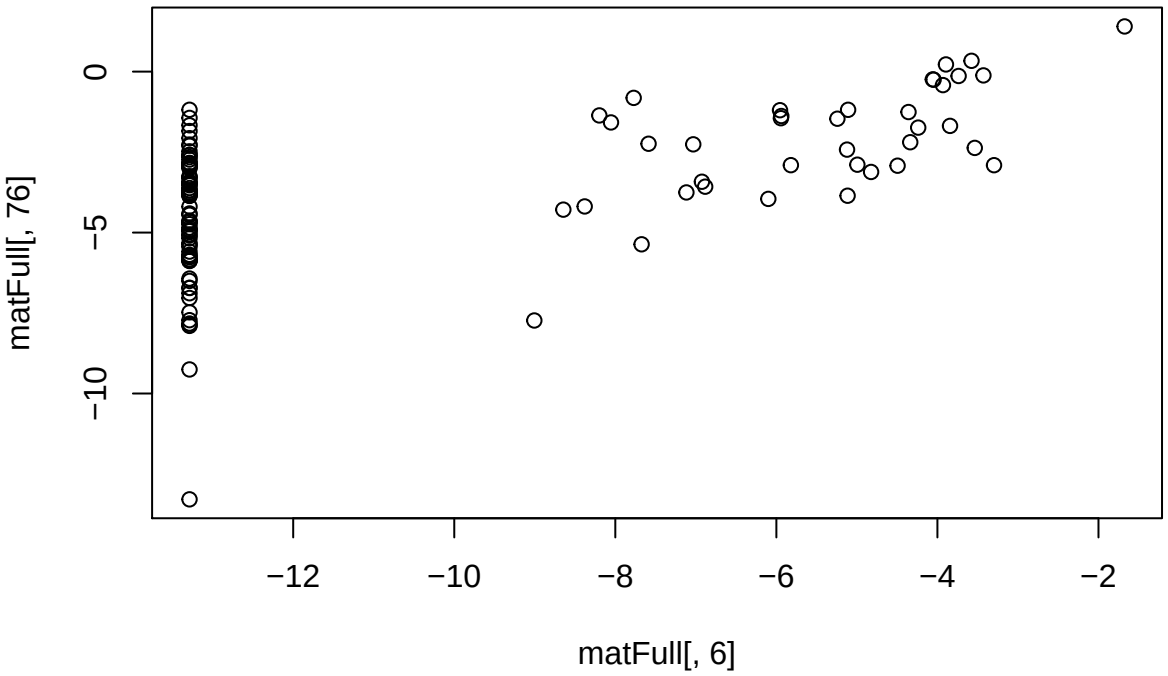

SJG−34
SJG−35
SJG−36
SJG−37
SJG−38
SJG−39
SJG−40
SJG−41
SJG−42
SJG−43
SJG−44
SJG−45
SJG−46
SJG−47
SJG−48
SJG−49
SJG−50
SJG−51
Expression log−ratio (this sample vs others)
Average log−expression


SJG−52
SJG−53
SJG−54
SJG−55
SJG−56
SJG−57
SJG−58
SJG−59
SJG−60
SJG−61
SJG−62
SJG−63
SJG−64
SJG−65
SJG−66
SJG−67
SJG−68
SJG−69
Expression log−ratio (this sample vs others)
Average log−expression



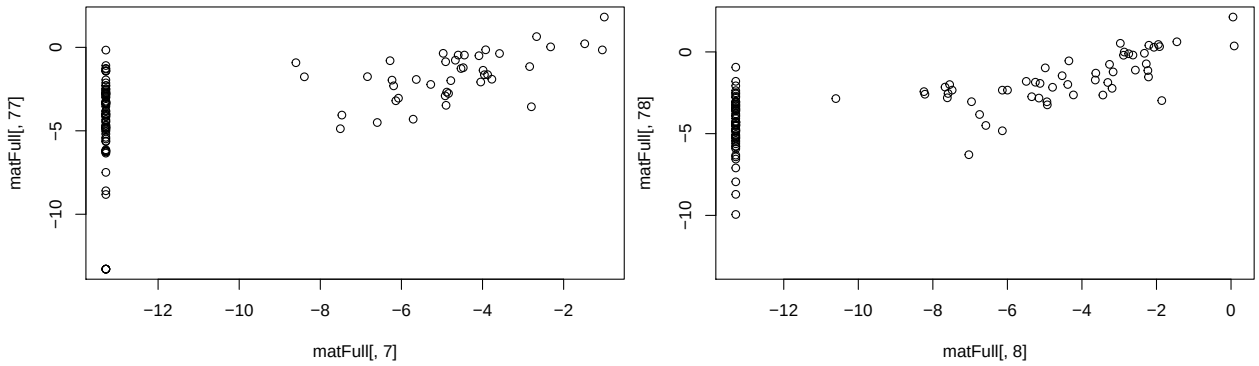
A number of sample pairs have unusual MA plot patterns and are further investigated here:

The RFI values for sample 6 and 76 are plotted:



Can see from the above plot that sample 6 has a large number of uniformly low AB expression values.

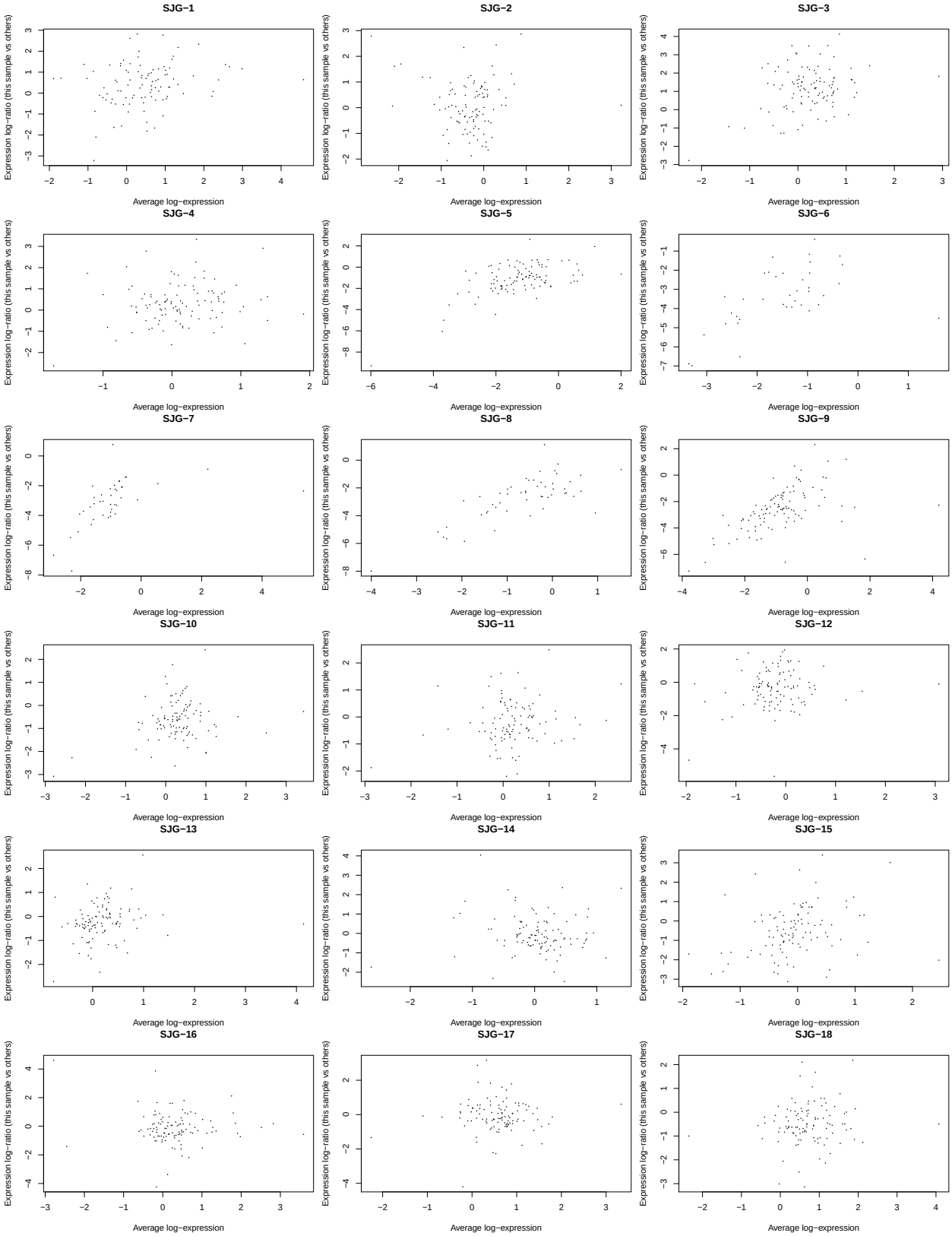
A similar pattern is seem for samples 7 & 77 and 8 & 78:

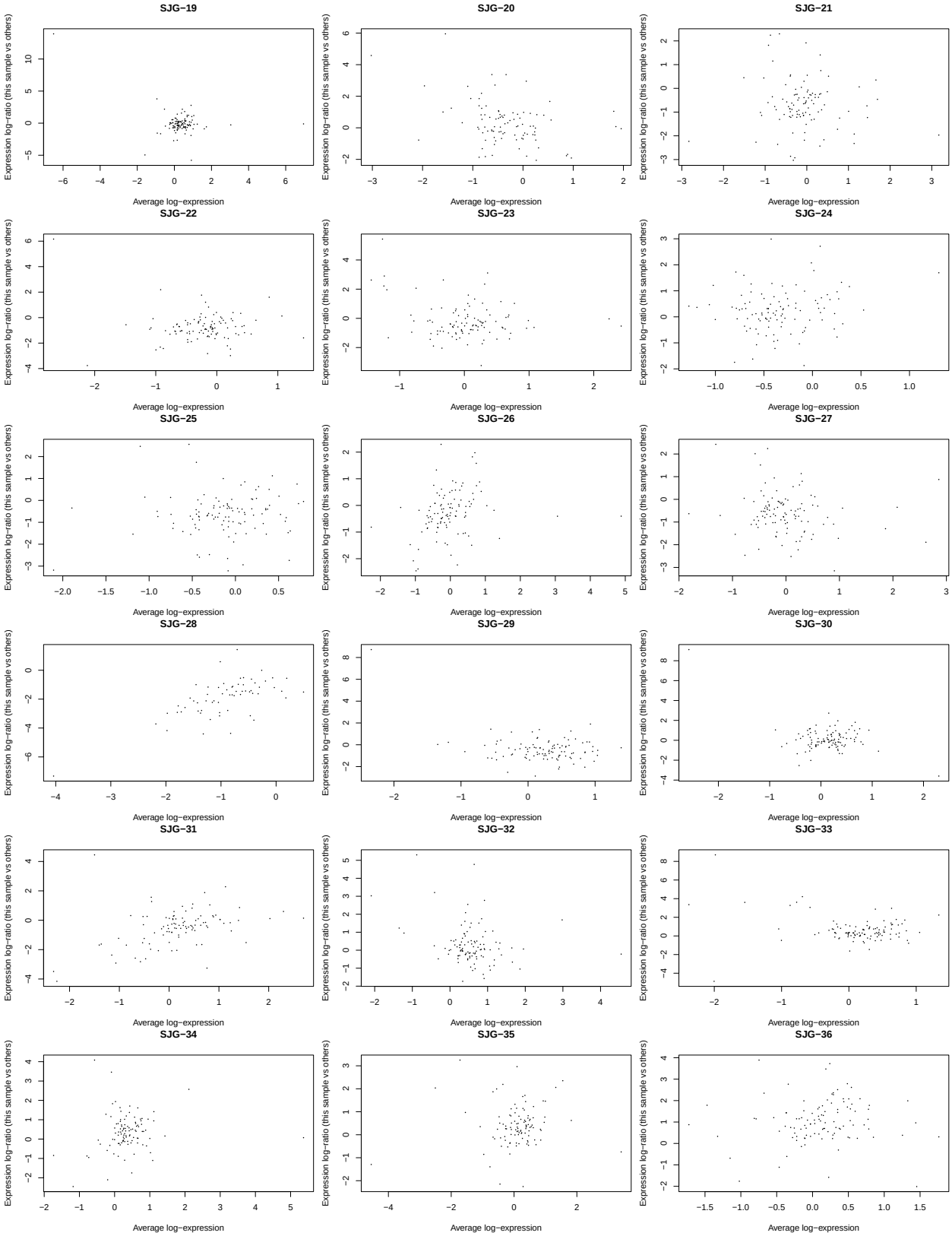


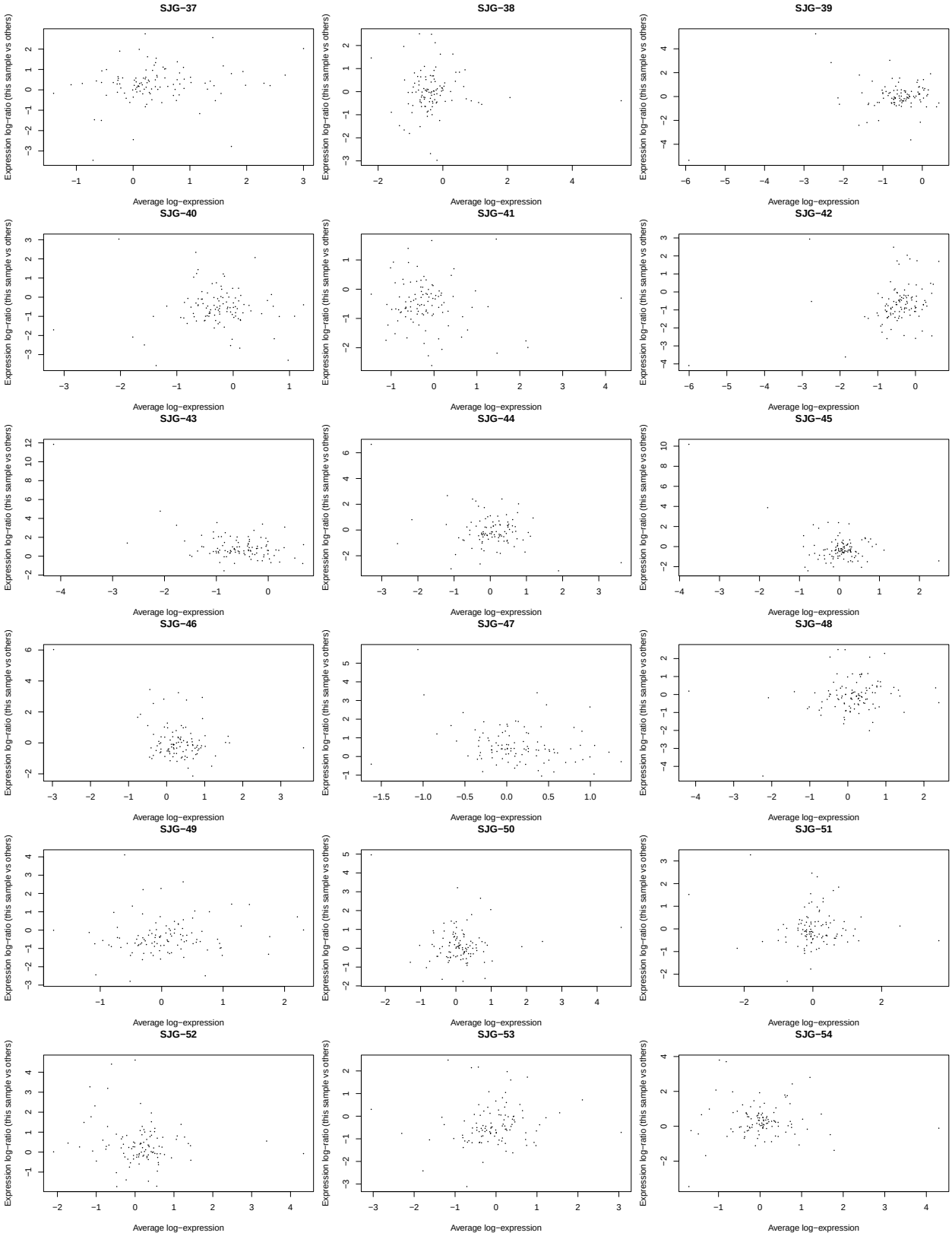
For sample's 20 & 90, 28 & 98 and 58 & 128, there is a similar cause to their unusual MA plots, except it is the second replicate that has the uniformly low RFI values.

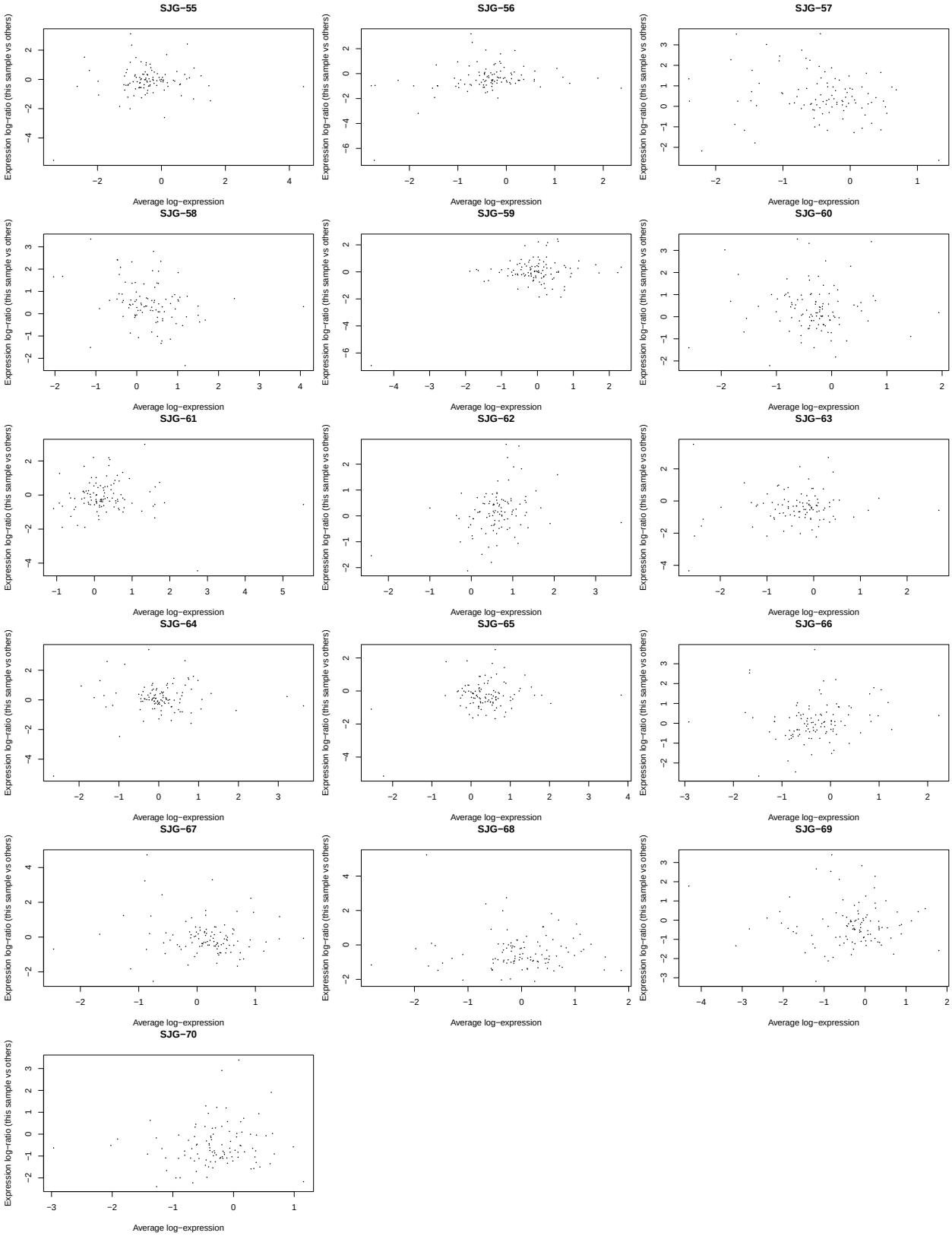
These plots suggest that it may be appropriate to remove an entire replicate and/or entire AB's that have a large proportion of low/NA RFI values. This is explored further in the "Filtering" document.

Median normalisation

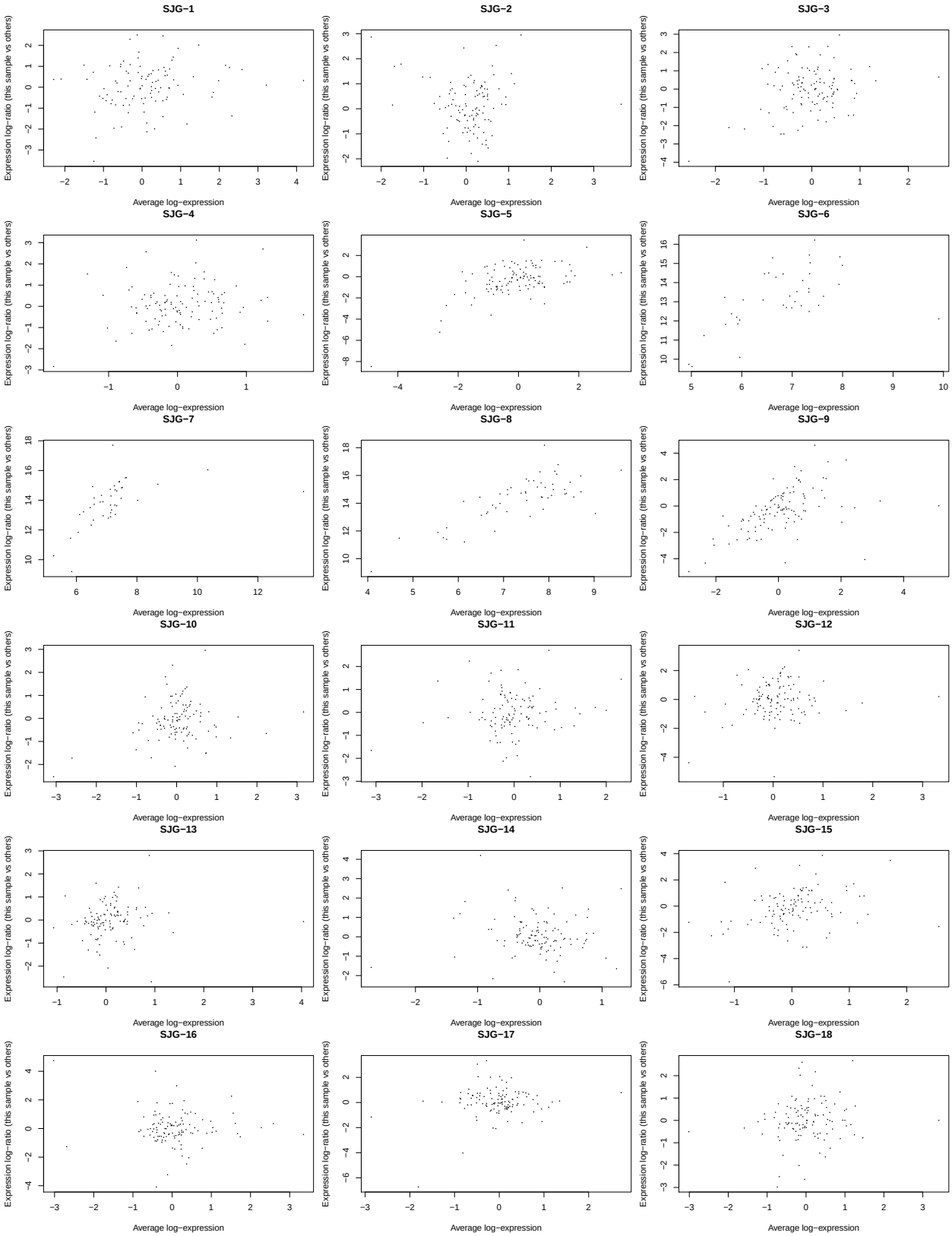


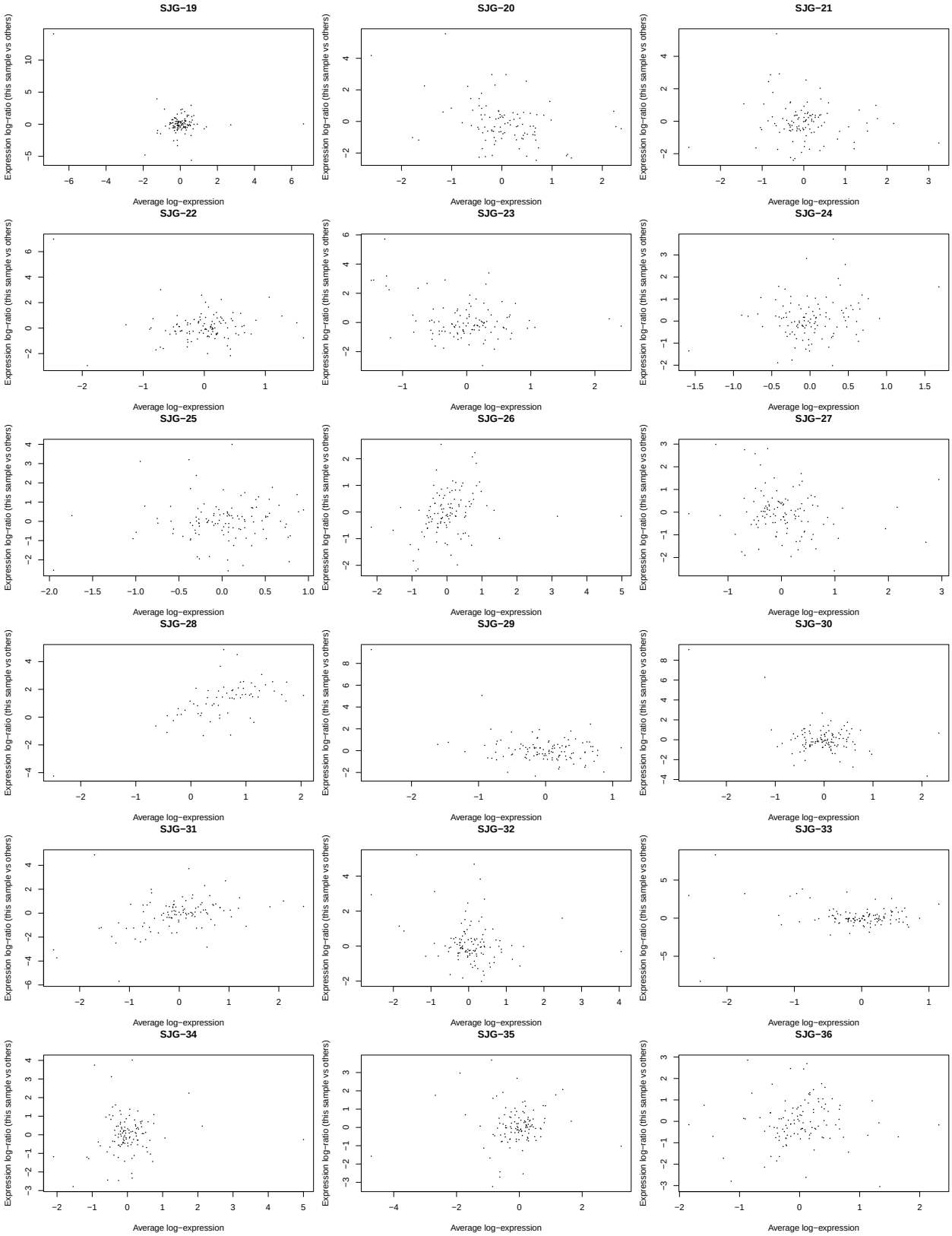

SJG−19
SJG−20
SJG−21
SJG−22
SJG−23
SJG−24
SJG−25
SJG−26
SJG−27
SJG−28
SJG−29
SJG−30
SJG−31
SJG−32
SJG−33
SJG−34
SJG−35
SJG−36
Expression log−ratio (this sample vs others)
Average log−expression

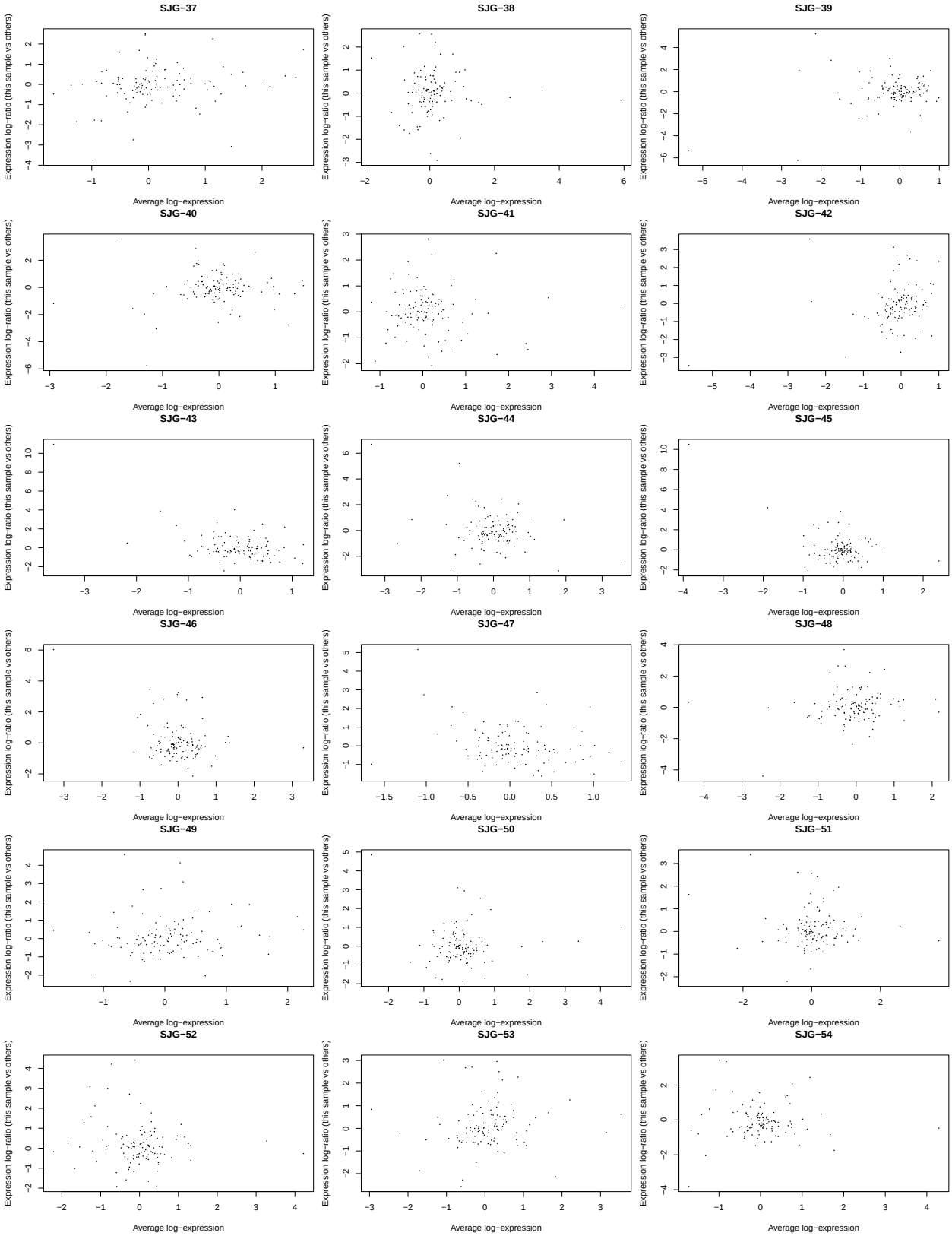

SJG−37
SJG−38
SJG−39
SJG−40
SJG−41
SJG−42
SJG−43
SJG−44
SJG−45
SJG−46
SJG−47
SJG−48
SJG−49
SJG−50
SJG−51
SJG−52
SJG−53
SJG−54
Expression log−ratio (this sample vs others)
Average log−expression

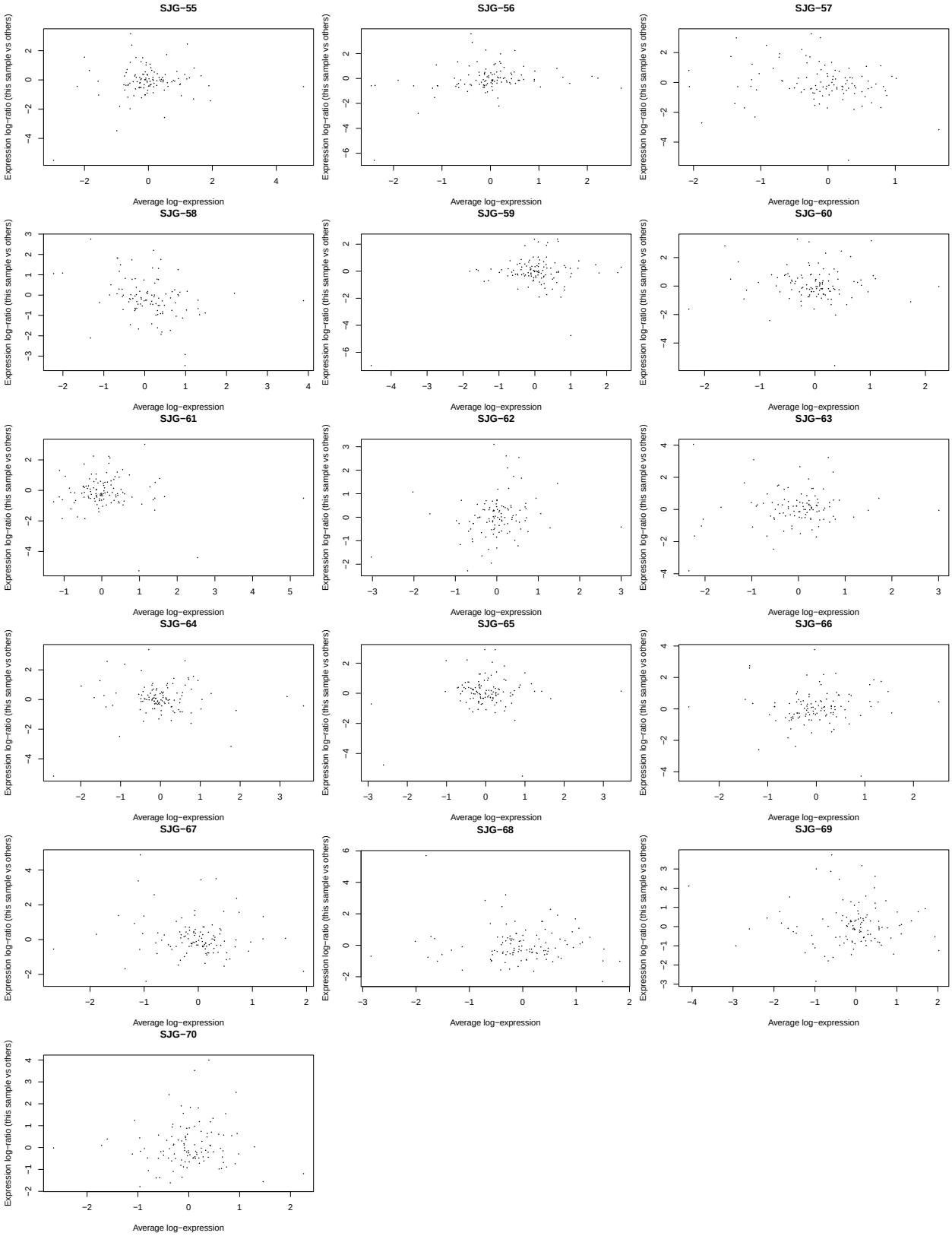

SJG−55
SJG−56
SJG−57
SJG−58
SJG−59
SJG−60
SJG−61
SJG−62
SJG−63
SJG−64
SJG−65
SJG−66
SJG−67
SJG−68
SJG−69
SJG−70
Expression log−ratio (this sample vs others)
Average log−expression

Loading control normalisation

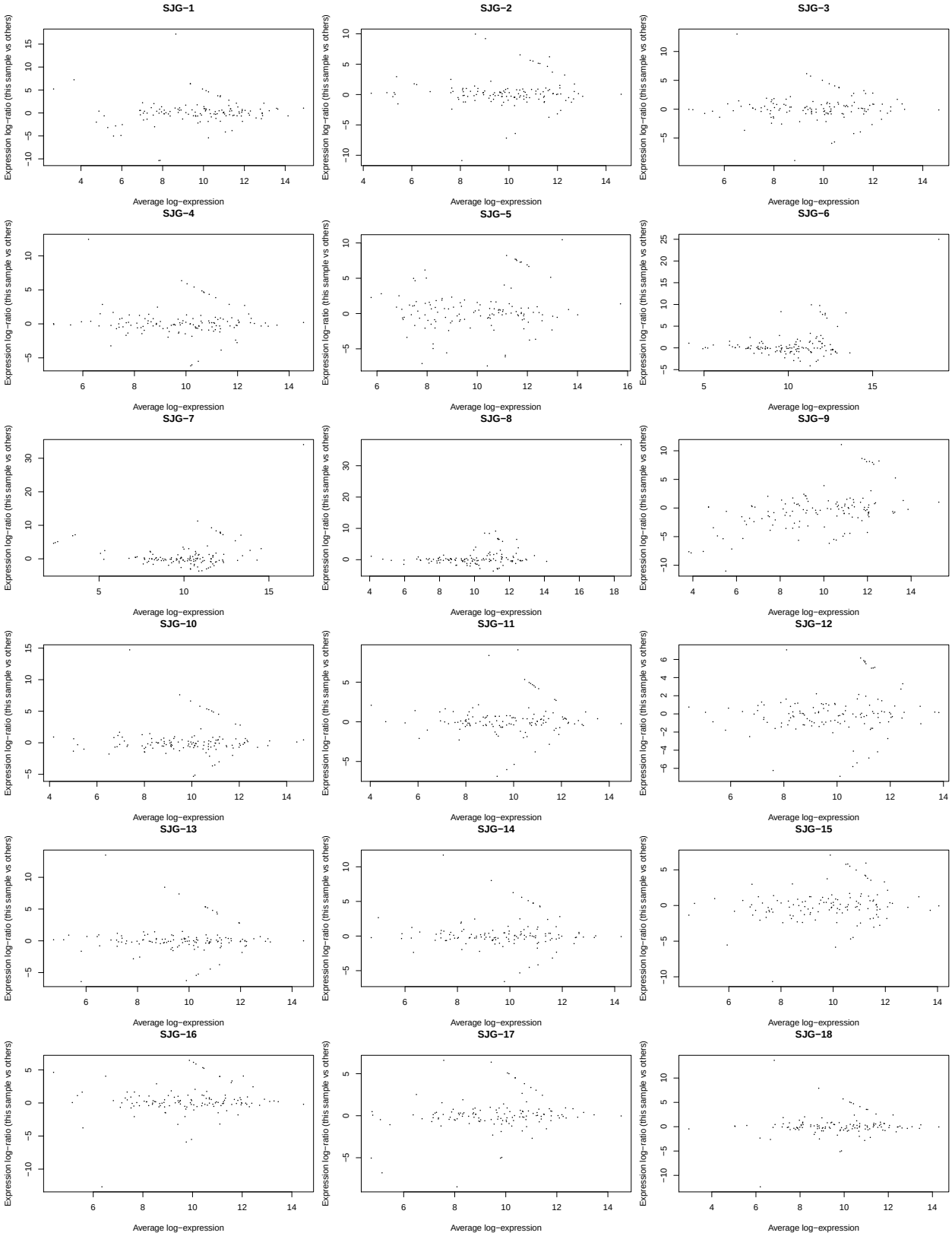


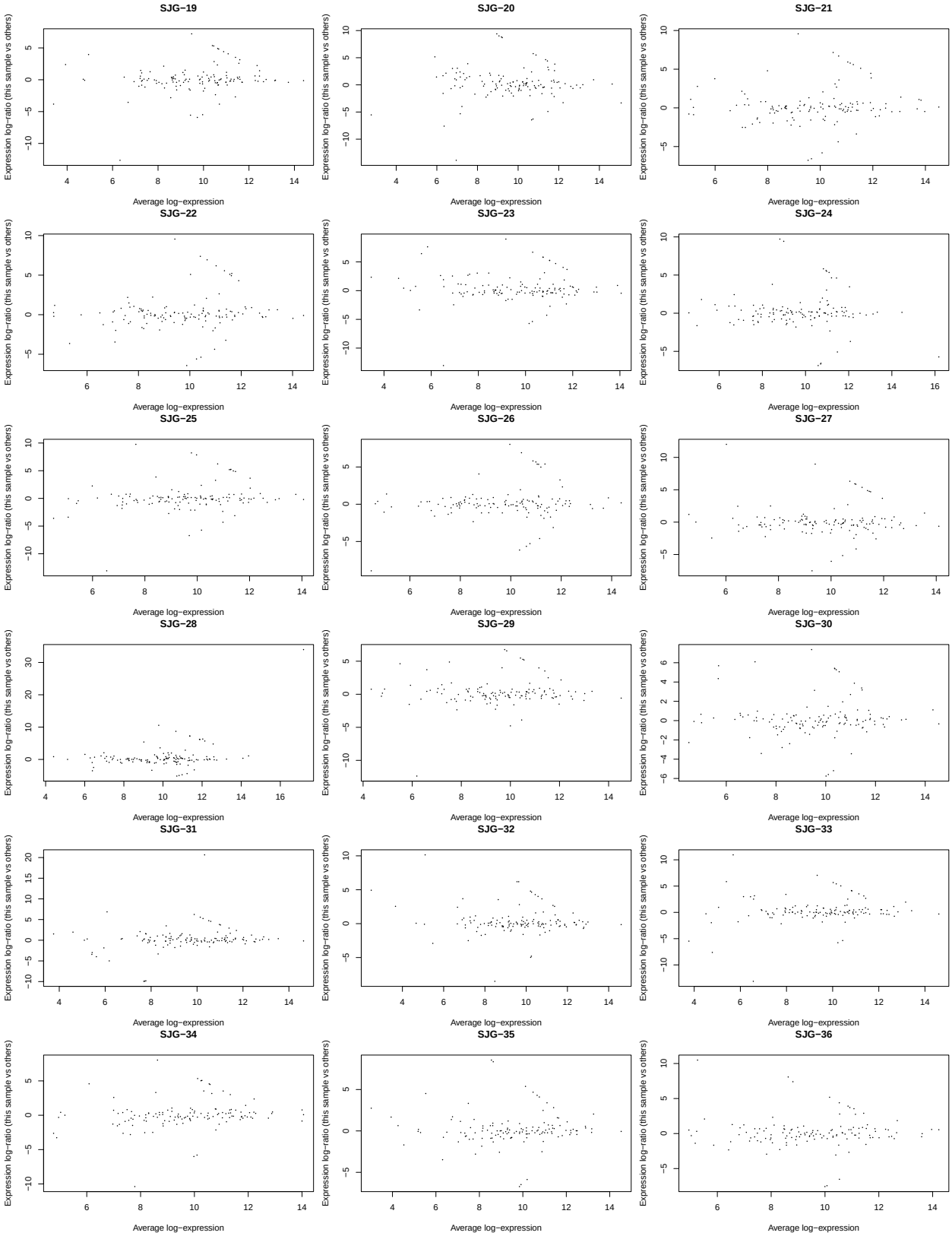

SJG−19
SJG−20
SJG−21
SJG−22
SJG−23
SJG−24
SJG−25
SJG−26
SJG−27
SJG−28
SJG−29
SJG−30
SJG−31
SJG−32
SJG−33
SJG−34
SJG−35
SJG−36
Expression log−ratio (this sample vs others)
Average log−expression

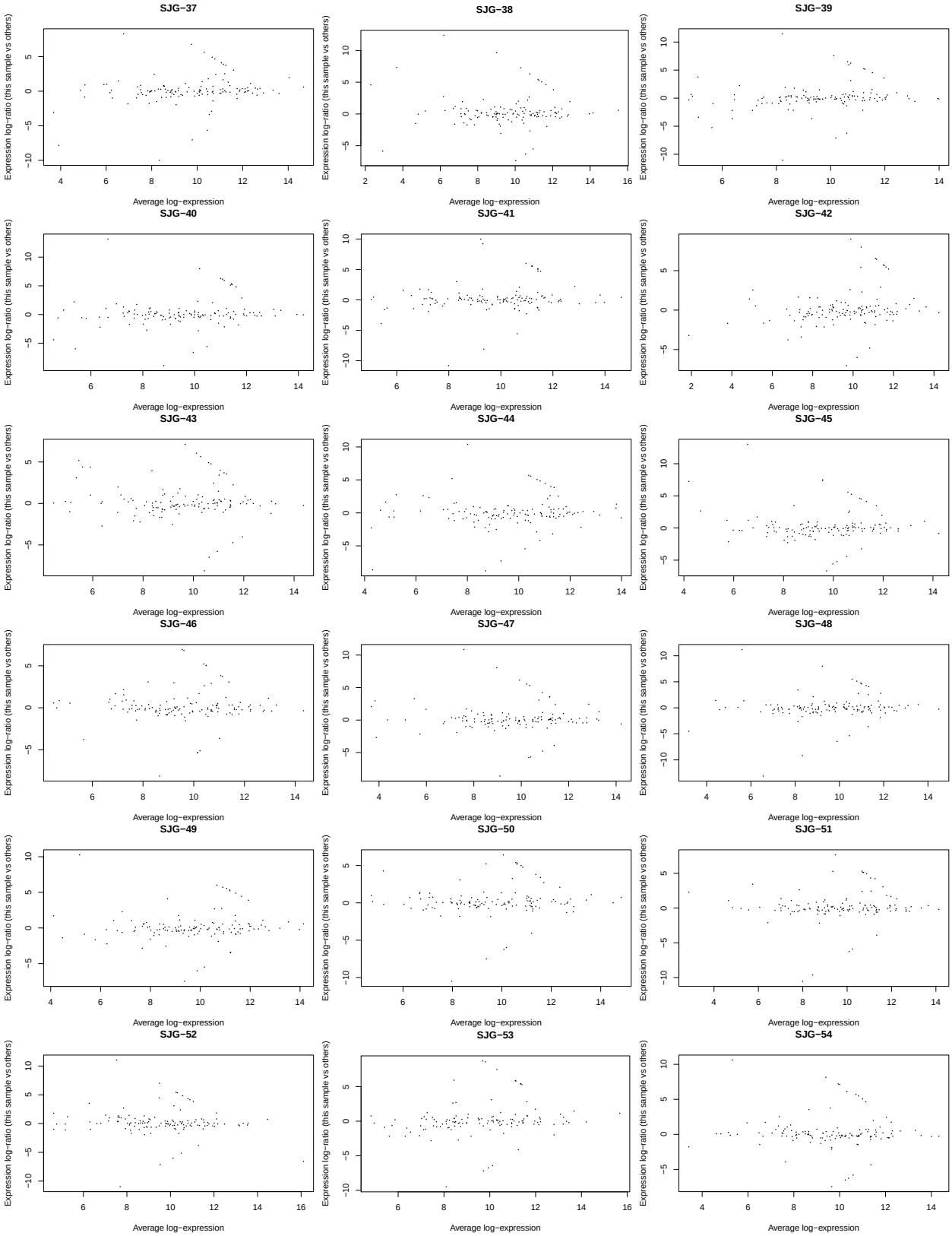

SJG−37
SJG−38
SJG−39
SJG−40
SJG−41
SJG−42
SJG−43
SJG−44
SJG−45
SJG−46
SJG−47
SJG−48
SJG−49
SJG−50
SJG−51
SJG−52
SJG−53
SJG−54
Expression log−ratio (this sample vs others)
Average log−expression

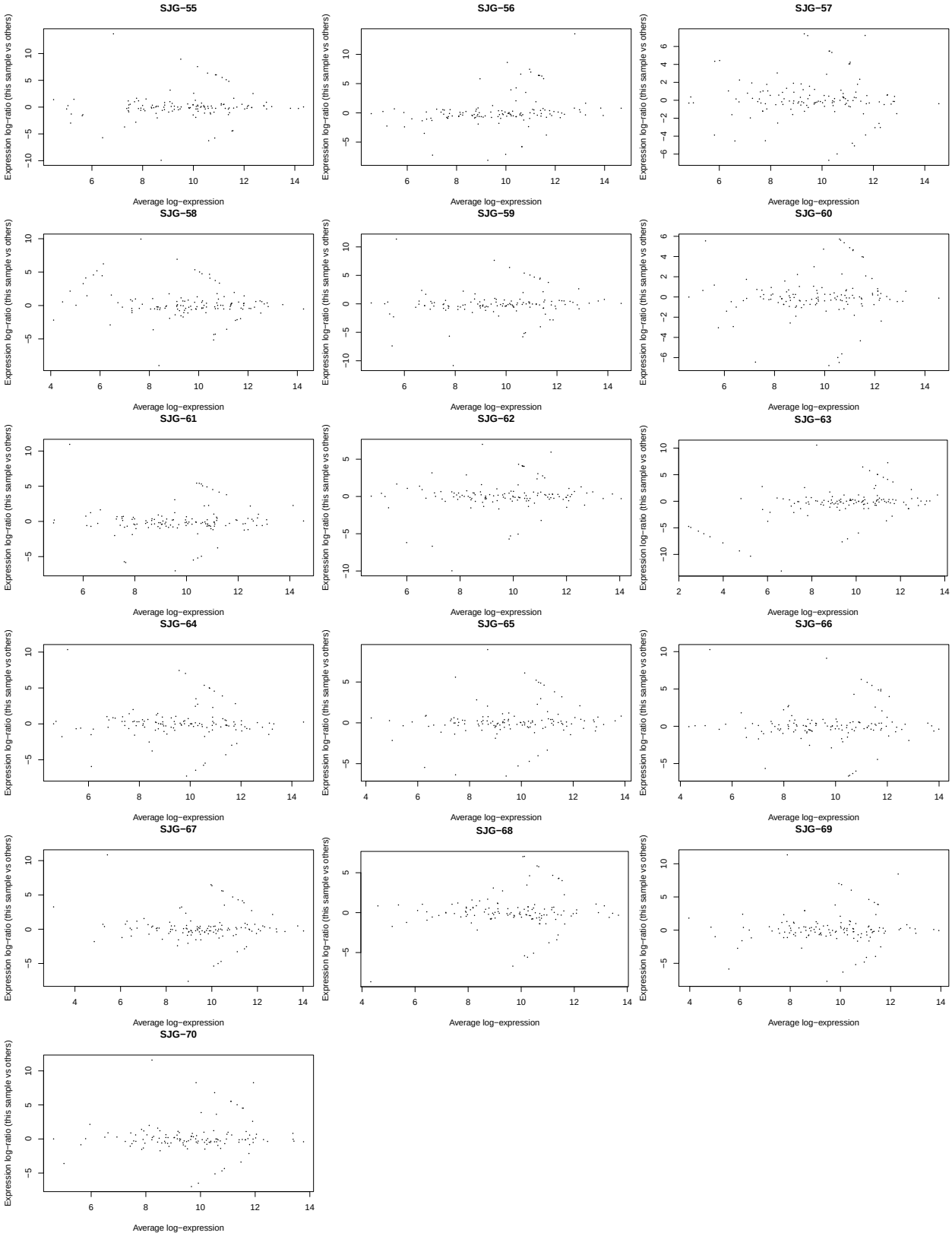

SJG−55
SJG−56
SJG−57
SJG−58
SJG−59
SJG−60
SJG−61
SJG−62
SJG−63
SJG−64
SJG−65
SJG−66
SJG−67
SJG−68
SJG−69
SJG−70
Expression log−ratio (this sample vs others)
Average log−expression

Global rank invariant



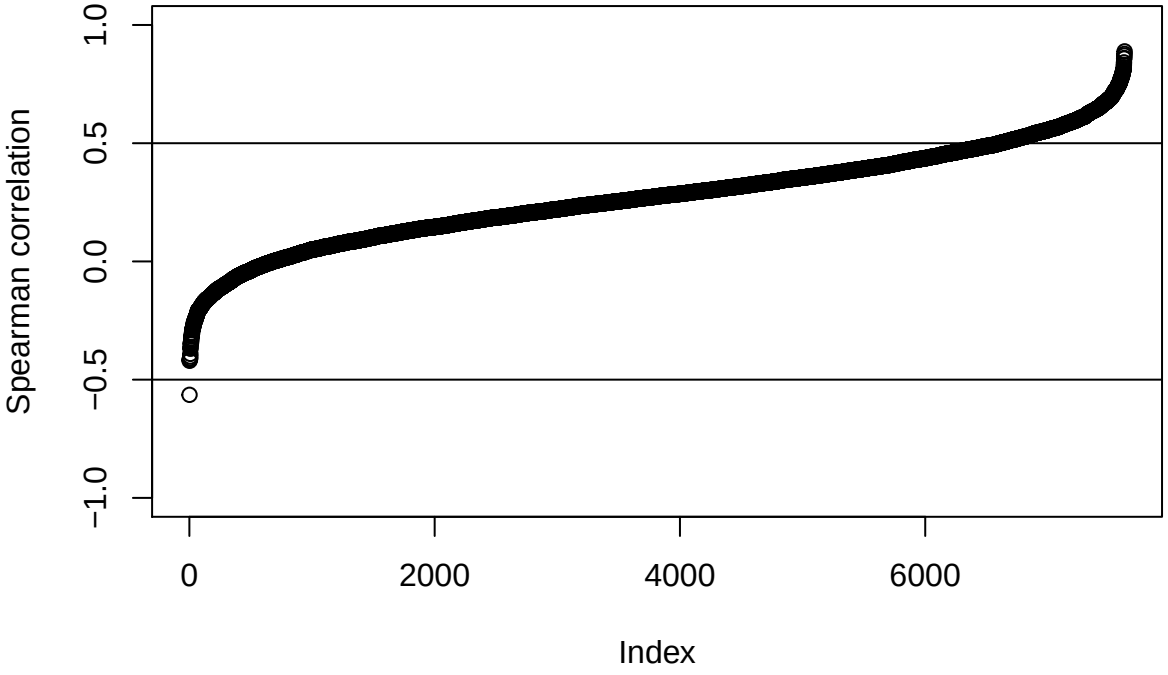

SJG−19
SJG−20
SJG−21
SJG−22
SJG−23
SJG−24
SJG−25
SJG−26
SJG−27
SJG−28
SJG−29
SJG−30
SJG−31
SJG−32
SJG−33
SJG−34
SJG−35
SJG−36
Expression log−ratio (this sample vs others)
Average log−expression


SJG−37
SJG−38
SJG−39
SJG−40
SJG−41
SJG−42
SJG−43
SJG−44
SJG−45
SJG−46
SJG−47
SJG−48
SJG−49
SJG−50
SJG−51
SJG−52
SJG−53
SJG−54
Expression log−ratio (this sample vs others)
Average log−expression

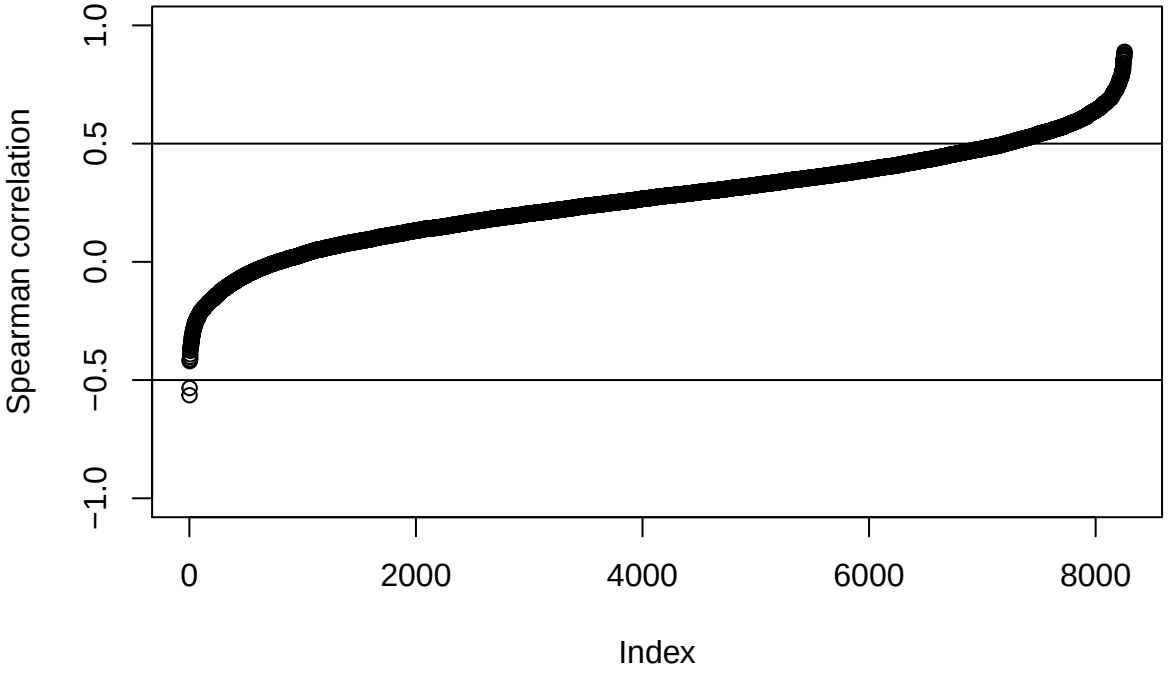

SJG−55
SJG−56
SJG−57
SJG−58
SJG−59
SJG−60
SJG−61
SJG−62
SJG−63
SJG−64
SJG−65
SJG−66
SJG−67
SJG−68
SJG−69
SJG−70
Expression log−ratio (this sample vs others)
Average log−expression

Pairwise correlation coefficient plots

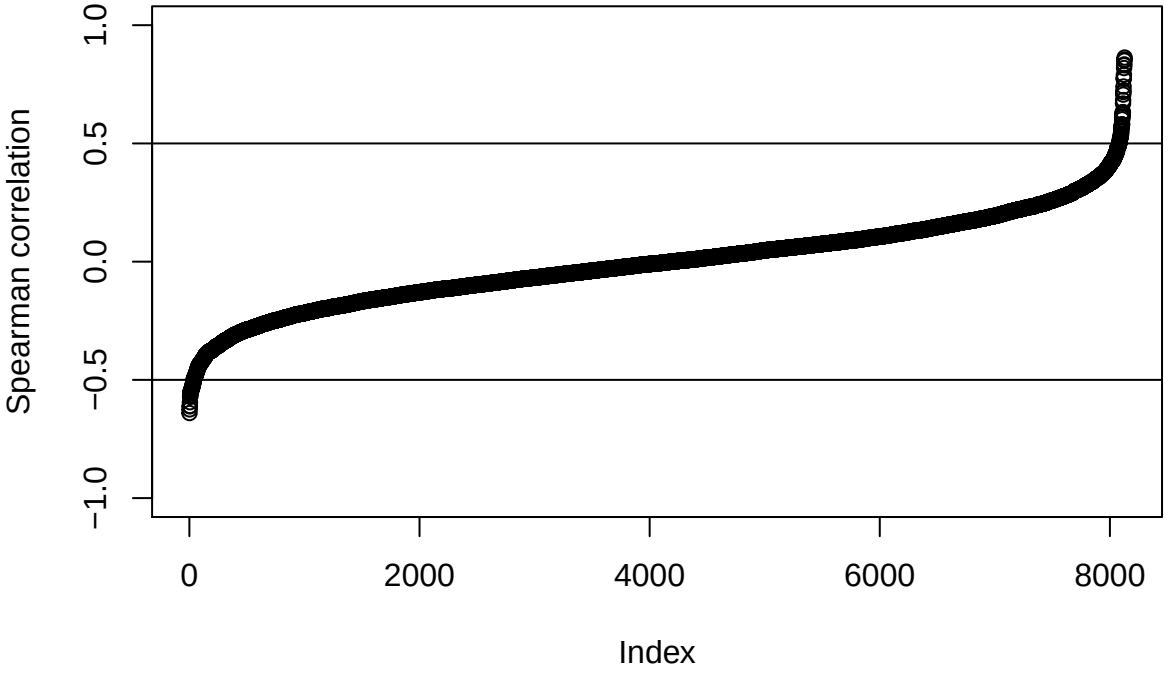
Raw data



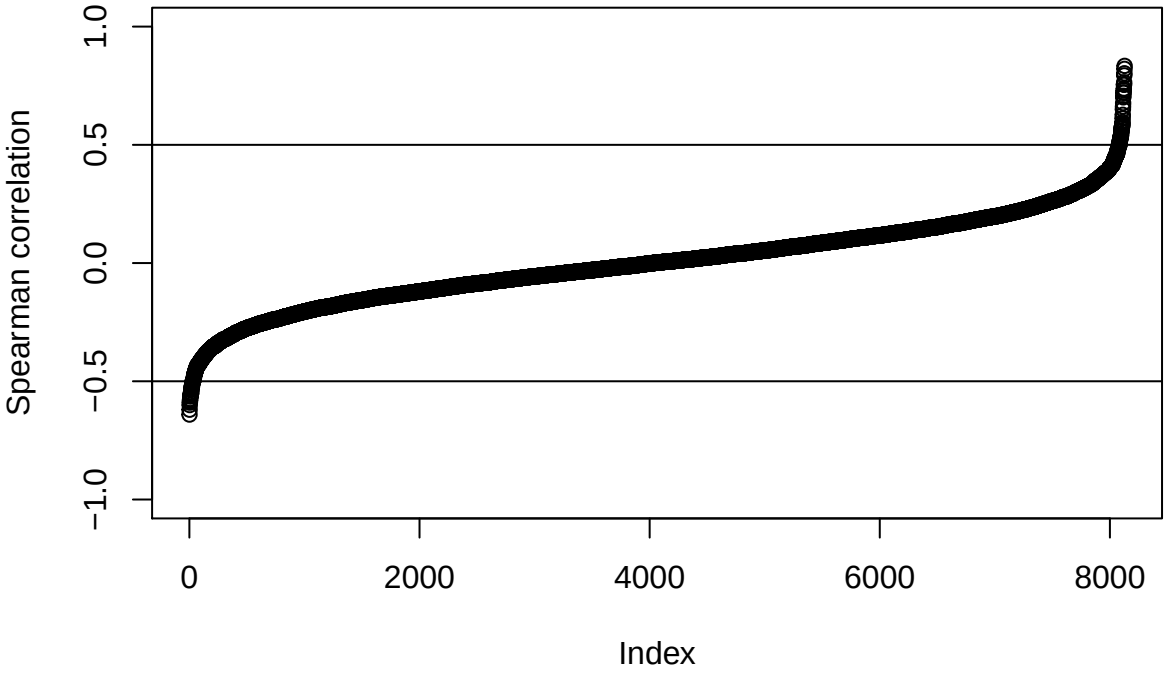
Median normalisation



Loading control normalisation



Global rank invariant



Normalised data

Average replicates then normalise

Normalise data using Global rank invariant and remove the 5 replicates with much poorer performance than their corresponding pair. Write out the data into a csv file. The steps are as follows:

1. Remove the 5 replicates that performed poorly. There were 5 samples were 1 replicate performed much worse than the other replicate.
2. Take the mean of the RFI values from the two replicates, for each sample. For samples where 1 replicate was removed, the RFI from the remaining replicate will be used.
3. Use Global rank invariant to normalise the data.
4. Write to a csv file.

Normalise both replicates

Normalise both replicates. We first look at how many NA's each AB has:

Antibody.Name	countNAs
Beta Catenin_P Ser33, Ser37, Thr41	70
Jak2_P Tyr1007, Tyr1008	70
MEK1/2	70
FLT3_P Tyr591	65
HER2/ErbB2_P Tyr1221, Tyr1222	65
JAK1_P Tyr1022, Tyr1023	65
Lyn_P Tyr507	65
N-Cadherin (D4R1H) XP	65
NDRG1	65
S6 Ribosomal Protein	65
Smad2_P Ser465, Ser467; Smad3_P Ser423, Ser425	65
Zap-70	65
Bad_P Ser112	28
Caspase 7_Cleaved Asp198	28
PKC pan_P bII Ser660	28
Caspase 3_Cleaved Asp175	23
Cyclin E1	15
BRCA1	14
Caveolin-1	14
cdc2_P Tyr15	14
cdc25c_P Ser216	14

Looking at the number of NA values, I have set the threshold at 30, removing all ABs with more than 30 NA values. The steps are:

1. Remove the 5 replicates that performed poorly. There were 5 samples were 1 replicate performed much worse than the other replicate.
2. Remove ABs with more than 30 missing values.
3. Use Global rank invariant to normalise the data.
4. Write to a tsv file.

References

Gandolfo, Luke C., and Terence P. Speed. 2018. "RLE Plots: Visualizing Unwanted Variation in High Dimensional Data." *PLOS ONE* 13 (2): e0191629. doi:10.1371/journal.pone.0191629.

Liu, Wenbin, Zhenlin Ju, Yiling Lu, Gordon B. Mills, and Rehan Akbani. 2014. "A Comprehensive Comparison of Normalization Methods for Loading Control and Variance Stabilization of Reverse-Phase Protein Array Data , A Comprehensive Comparison of Normalization Methods for Loading Control and Variance Stabilization of Reverse-Phase Protein Array Data." *Cancer Informatics* 13 (January): CIN.S13329. doi:10.4137/CIN.S13329.

Pelz, Carl R., Molly Kulesz-Martin, Grover Bagby, and Rosalie C. Sears. 2008. "Global Rank-Invariant Set Normalization (GRSN) to Reduce Systematic Distortions in Microarray Data." *BMC Bioinformatics* 9 (1): 520. doi:10.1186/1471-2105-9-520.

Wachter, Astrid, Stephan Bernhardt, Tim Beissbarth, and Ulrike Korf. 2015. "Analysis of Reverse Phase Protein Array Data: From Experimental Design Towards Targeted Biomarker Discovery." *Microarrays* 4 (4): 520–39. doi:10.3390/microarrays4040520.