# Supplementary Methods

November 24, 2011

## 1 Examination of removal of non G-stack probes with above-average magnitude

We have shown in this paper that the effect of removal of G-stack probes are more pronounced than that of randomly chosen probes. It could be argued that this effect is because at least some G-stack probes are more efficiently hybridising with the transcripts in question and eliminating them reduces the overall signal which is mistaken as a bias.

It is therefore important to compare this data with the effect of removing non G-stack probes with very high levels of hybridisation. This section aims to demonstrate the effects of removal of non G-stack probes which are considered as highly expressed probes. Namely as they have the highest probe intensities among the members of their probe set.

The selection of probe sets for this section is termed as Group E2 in which two probes per probe sets are removed and examined against Group A2. For Group E2, 2,124 probe sets have been chosen randomly that do not have any G-stack probe (2,124 is the number of probe sets in Group A2 that each have two G-stack probes). The two probes from each of the probe set in Group E2 are not randomly chosen and are selected using the following method:

- As we are focussing on the data set GSE1869, the median of the 25 intensity values (for each CEL file in GSE1869) of each probe is computed.
- For each probe set, two probes are selected that have highest median intensity values within the probe set.

The two probes in each probe set described above are removed from Group E2 and a modified CDF is generated. Plots of difference in expression level, correlation and fold change values between A2 and E2 are depicted in figures S1, S2, and S3 respectively.

Figure S1 shows that after removal of two non G-stack probes with above-average intensity values there is consistent reduction in expression values, which is consistent given that the probes removed have the highest intensity. However, the removal of G-stack probes give a more complicated pattern of change, particularly for larger expression levels. Figures S2 and S3 also show a linear pattern in the difference in fold change and correlation values. These figures show that the clear biases seen in Group A2 have almost disappeared in Group E2.

We note that the same analysis was repeated where two non G-stack probes that have average intensity values are removed. Again, it was found that there was clear difference between this choice of probe removal and A2 (data not shown).

## Figure Legends

Figure S1: Plot comparing change in expression values in the data set GSE1869, before and after removal of two probes, of the probe sets in groups A2 and E2. The central values represent the median while the error bars indicate the upper and lower quartiles. The width of error bars reflect the number of data points (widest for the most numerous, shortest for the least).

Figure S2: Scatter plots comparing the change in fold change, before and after removal of two probes, of the probe sets in a) group A2 b) group E2.

Figure S3: Plot comparing the change in correlation, before and after removal of two probes, of the probe sets in groups A2 and group E2. The central points indicate the median. The error bars indicate the upper and lower quartiles. The width of error bars reflect the number of data points (widest for the most numerous, shortest for the least).

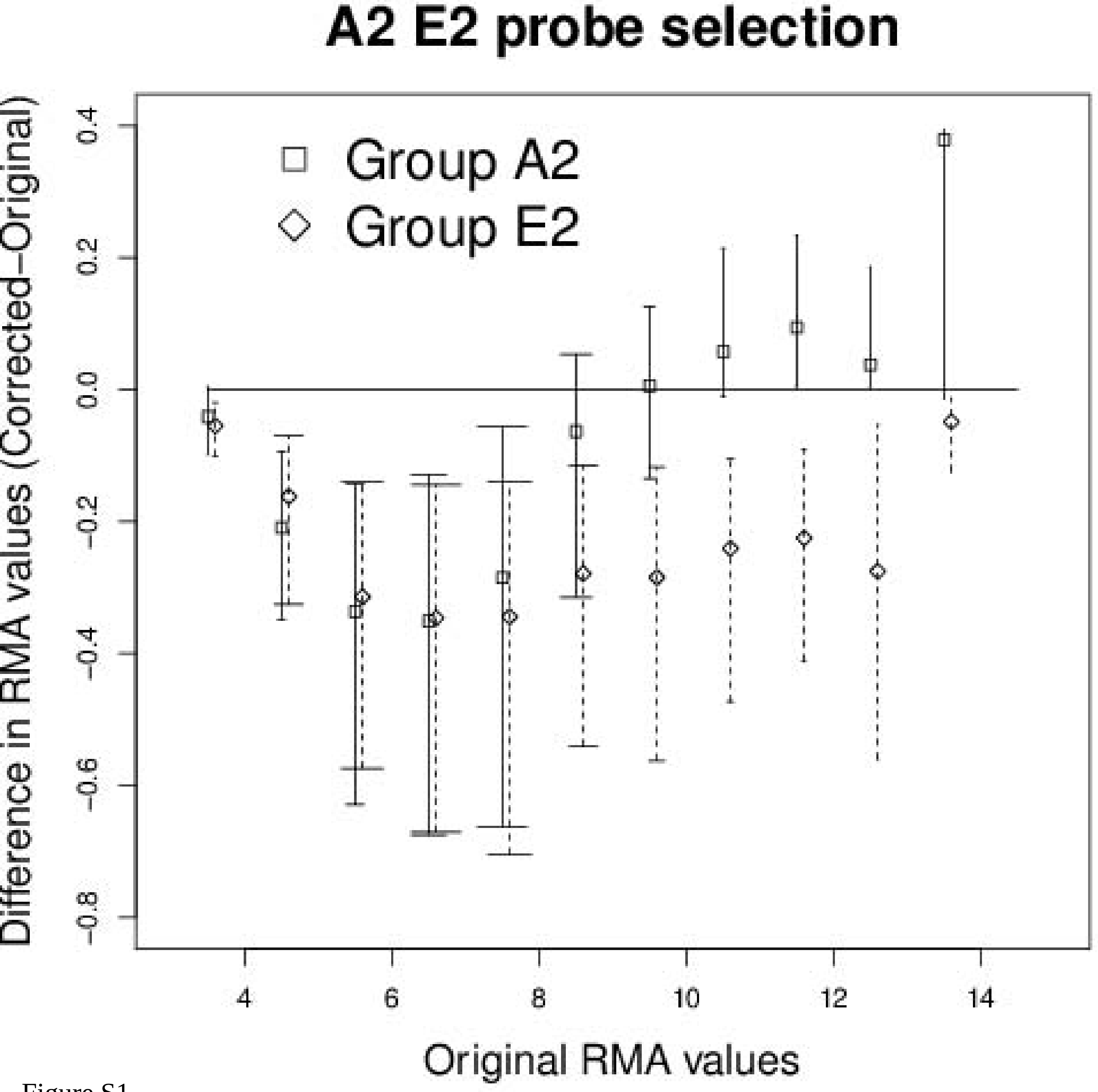
A2 E2 probe selection
Group A2
Group E2
Difference in RMA values (Corrected–Original)
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
4
6
8
10
12
14
Original RMA values

Figure S1

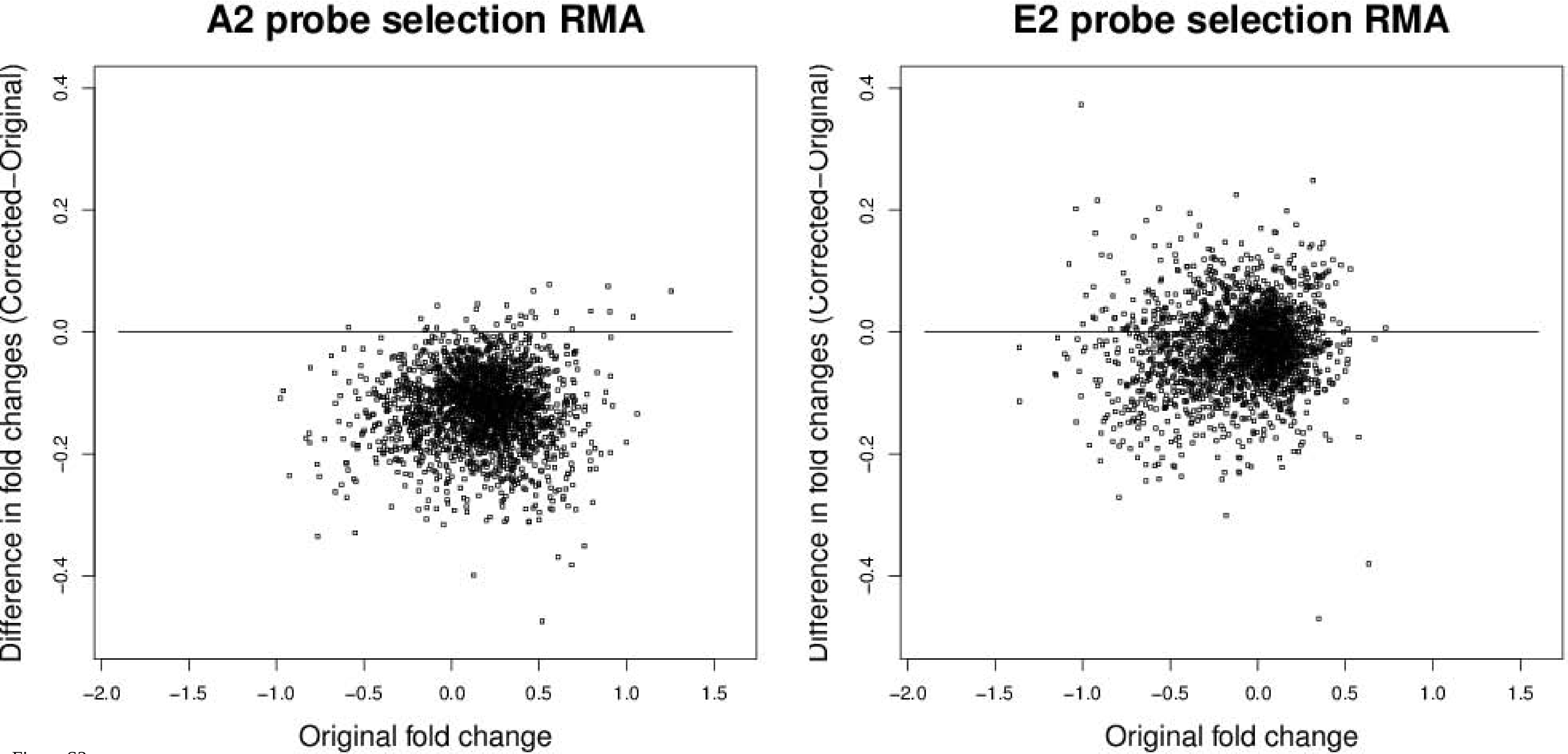
A2 probe selection RMA
E2 probe selection RMA
Difference in fold changes (Corrected–Original)
Original fold change
0.4
0.2
0.0
-0.2
-0.4
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

Figure S2

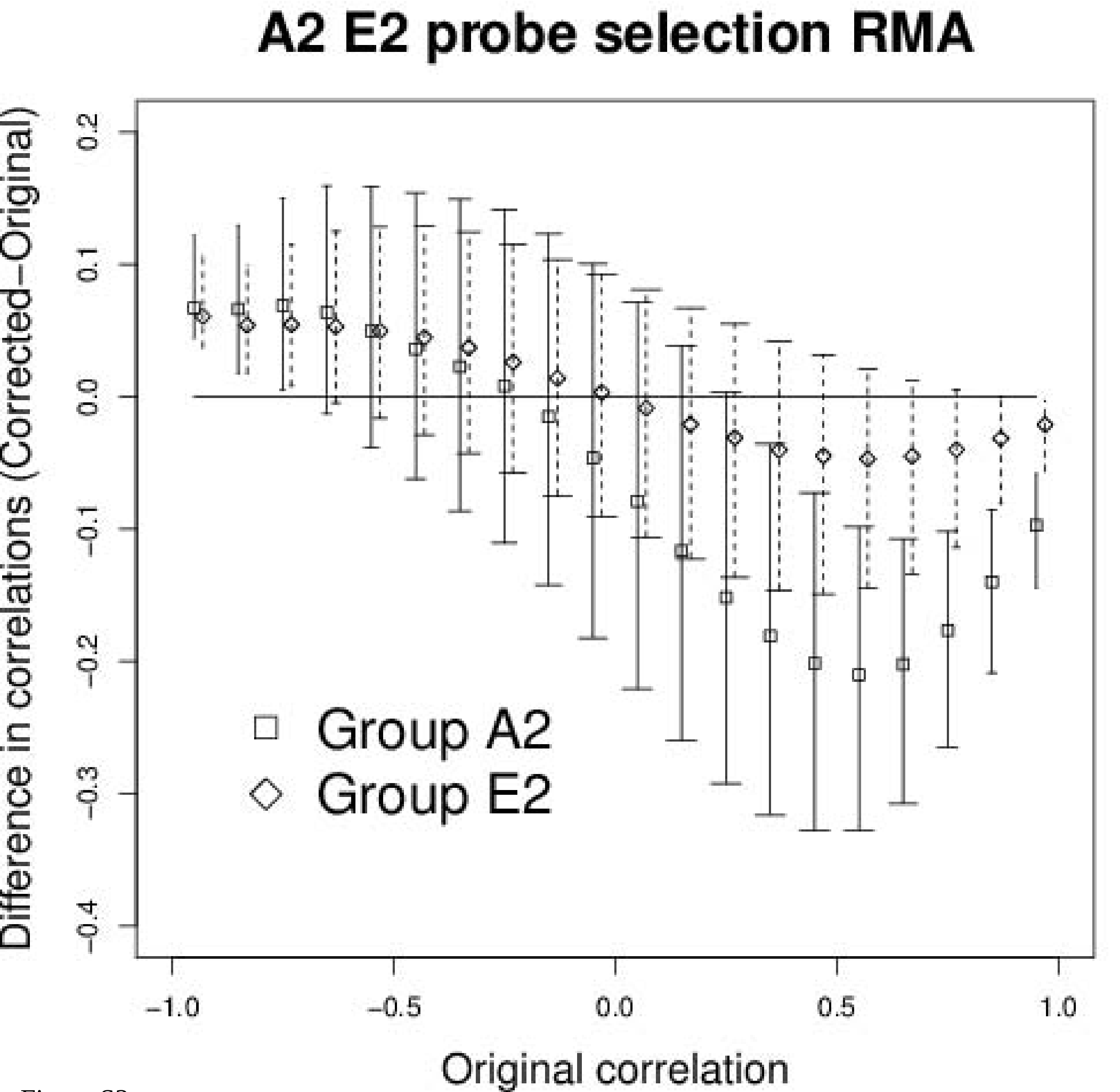
A2 E2 probe selection RMA
Difference in correlations (Corrected−Original)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−1.0
−0.5
0.0
0.5
1.0
Original correlation
Group A2
Group E2

Figure S3