Table 8. Probe set to gene ID reduction

Number of probe set ids per gene	Number of genes	Percent of genes
1	10,553	70.07
2	2,758	18.31
3	1,136	7.54
4	407	2.70
5	128	0.85
6	46	0.31
7	13	0.09
8	9	0.06
9	5	0.03
10	3	0.02
11	1	0.01
12	0	0.00
13	1	0.01
Total	15,060	

HGU133A

Number of probe set ids per gene	Number of genes	Percent of genes
1	8,276	81.91
2	1,326	13.12
3	381	3.77
4	83	0.82
5	18	0.18
6	11	0.11
7	6	0.06
8	3	0.03
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
Total	10,104	

HGU95AV2

Number of probe set ids per gene	Number of genes	Percent of genes
1	5,670	89.8
2	516	8.17
3	102	1.62
4	18	0.29
5	2	0.03
6	4	0.06
7	1	0.02
8	1	0.02
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
Total	6,314	

HU6800

The distribution of probe sets per gene in the three Affymetrix chip types used on the data sets in the paper are shown. The data displayed is binned by the number of probes per gene. The majority of the overrepresentation arises from two or three probes per gene. In our analyses, we chose the maximally expressed probe as the single representative of the corresponding gene.