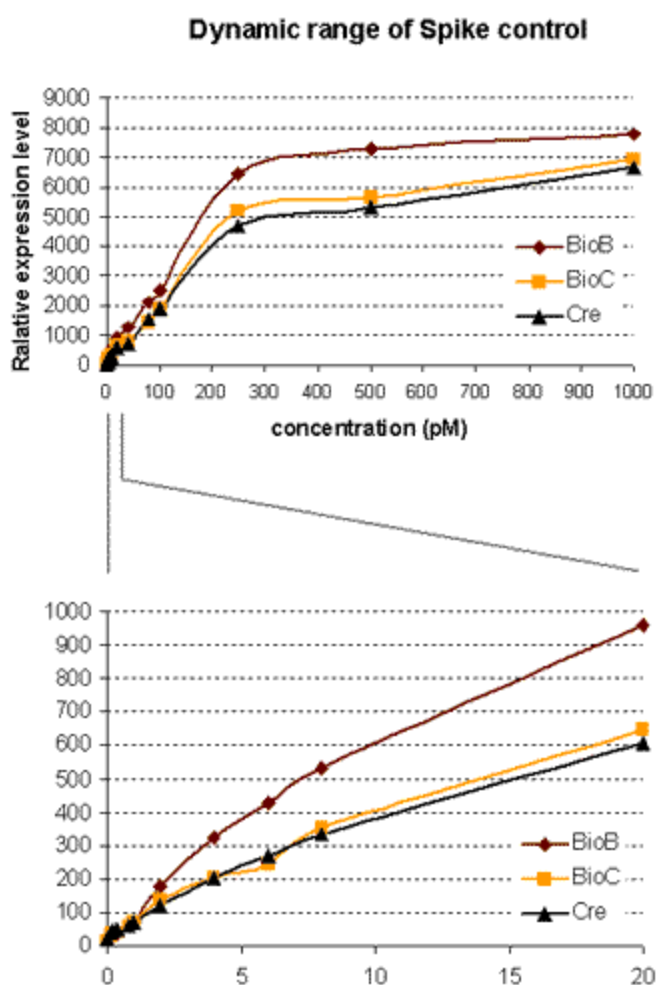


## Arabidopsis Whole Genome Array

The custom Arabidopsis whole genome array contains 42 Affymetrix control probe sets, 3 transgene control probe sets, 4 QC probe sets and 26412 probe sets representing 26367 Arabidopsis genes. The probe is 25mer oligonucleotide in length selected from individual exon region of the genes that allow differentiating genes with alternative splicing as well as gene family. On average, each gene contains 15 perfect match probes and mismatch probes are completely eliminated. The probes were selected from the 3' end of each exon.

## Sensitivity and Dynamic Range

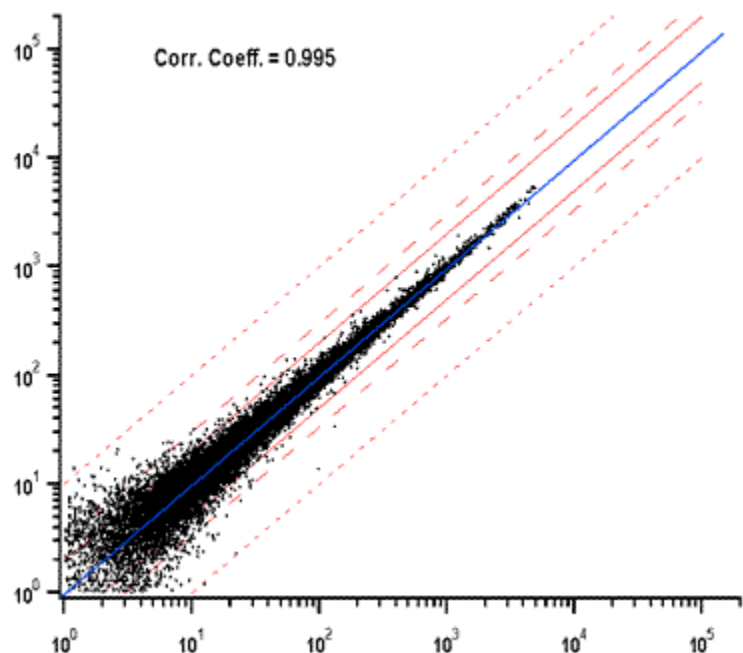
To determine the dynamic range of the 26K Arabidopsis whole genome array, an equal molar of spike control mixture including BioB, BioC, BioD and CreX was prepared and serial dilution of spike control mixture was mixed with 10 mg Arabidopsis cRNAs and hybridized to the 26K Arabidopsis whole genome array. Data from BioD were excluded from calculation due to higher background value even in the absence of BioD spike-in.



As shown in the figure, the linear dynamic range is between 0.4 pM and 250 pM and there is greater than 500 fold linearity for the Arabidopsis 26K whole genome array.

## Reproducibility

To measure reproducibility of expression profiling data, cRNAs prepared from the same sample were split to two and hybridized to the 26K Arabidopsis genome array. False positives are defined based on experiments in which cRNAs are split, hybridized to the 26K Arabidopsis genome arrays and the results are compared. A false positive is indicated if a probe is scored quantitatively as changing by at least two fold and the expression level is greater than 25. Scatter plots from duplicate samples suggest that false positive rate is very low with 0.23% on average).

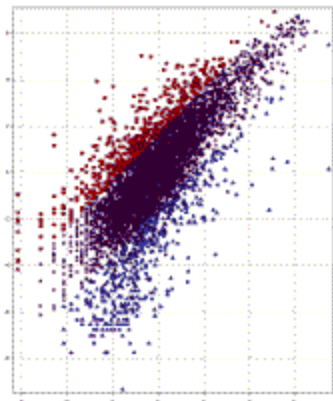


| sample ID | Number of FP | Number of genes | False positive rate |
|-----------|--------------|-----------------|---------------------|
| AD002001  | 59           | 26369           | 0.22                |
| AD002003  | 81           | 26369           | 0.31                |
| AD002005  | 18           | 26369           | 0.07                |
| AD002009  | 126          | 26369           | 0.48                |
| AD002011  | 36           | 26369           | 0.14                |
| AD002013  | 46           | 26369           | 0.17                |

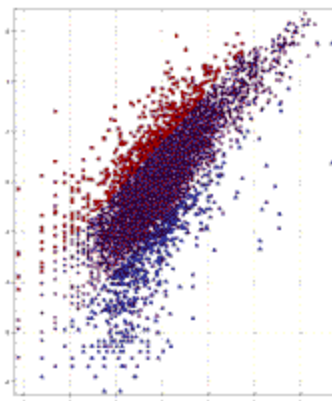
## Correlation with Arabidopsis 8K GeneChip

In order to compare the data generated from the commercial 8K GeneChip, 7582 common genes on both the Arabidopsis 8K and 26K genome arrays were selected and correlation coefficient was calculated. If data from these genome arrays are highly correlated, the correlation coefficient will be close to 1. If they are not correlated then the correlation coefficient will be close to 0. The mean correlation coefficient was  $0.85 \pm 0.02$  among 23 different tissues. These results suggest that gene expression profiling results using either the 8K and 26K are comparable despite the differences in probe sequences, feature sizes, probe set design and computational algorithms between these two genome arrays.

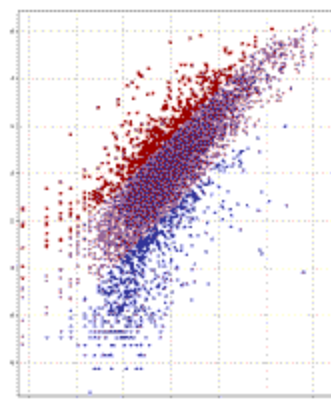
The differences between the Arabidopsis 8K and 26K GeneChip microarrays are mainly for the low expressed genes. Among 12 samples, we consistently observed higher signals for genes which are low expressors, as shown in the scatter plots. Note how the curve at the low end is biased towards the 26K GeneChip (X-axis).



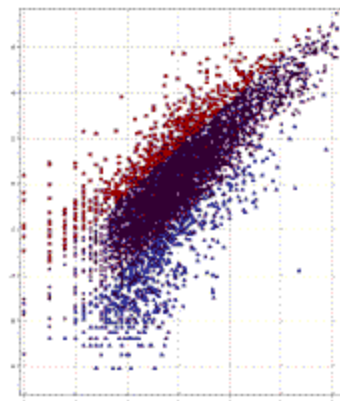
4d seedling 1



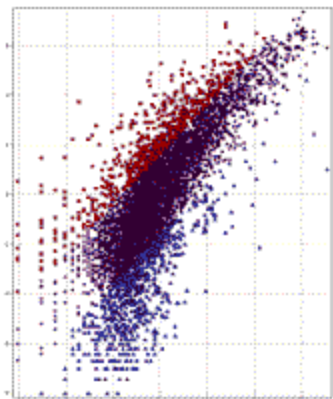
4d seedling 2



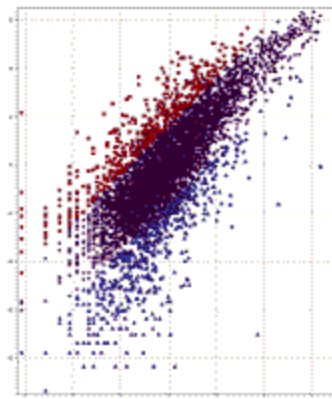
5wk root



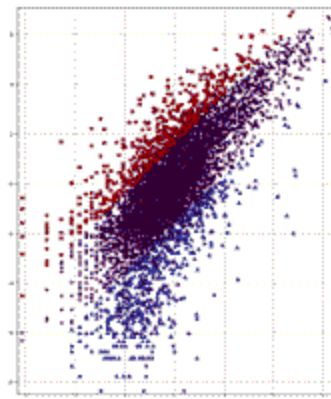
Adult root



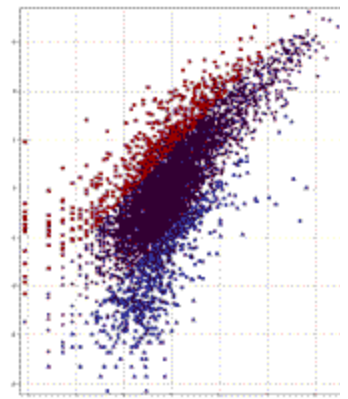
Inflo stem 1



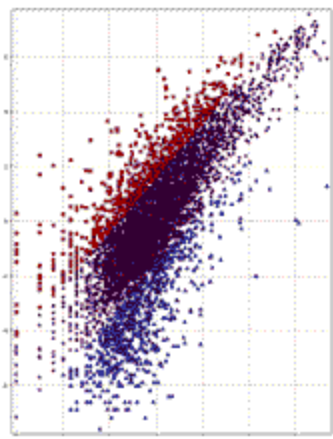
Inflo stem 2



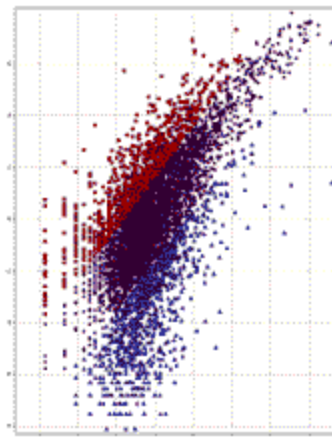
Flower



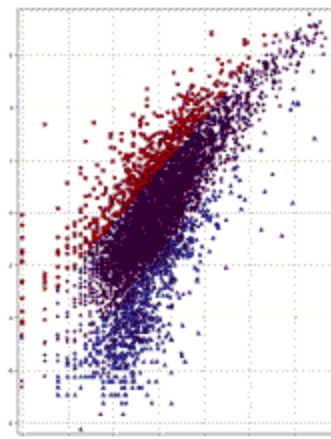
Silique



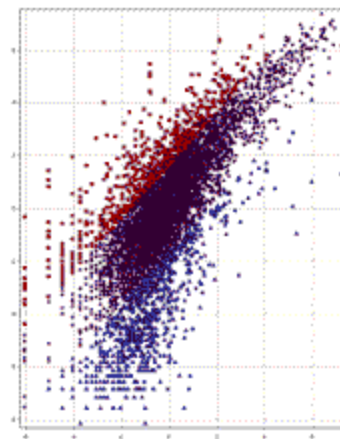
2wk leaf 1



2wk leaf 2



Adult leaf



Senescent leaf