

University of Massachusetts Amherst

From the Selected Works of Dong Wang

February, 2007

A High-Performance, Small-Scale Microarray for Expression Profiling of Many Samples in Arabidopsis-Pathogen Studies

Masano Sato

Raka M. Mitra

John Coller

Dong Wang, *University of Massachusetts - Amherst*

Natalie W. Spivey, et al.



Available at: https://works.bepress.com/dong_wang/4/

TECHNICAL ADVANCE

A high-performance, small-scale microarray for expression profiling of many samples in Arabidopsis–pathogen studies

Masanao Sato¹, Raka M. Mitra¹, John Coller², Dong Wang^{3,†}, Natalie W. Spivey³, Julia Dewdney⁴, Carine Denoux⁴, Jane Glazebrook¹ and Fumiaki Katagiri^{1,*}

¹Department of Plant Biology, Microbial and Plant Genomics Institute, University of Minnesota, 1500 Gortner Avenue, St Paul, MN 55108, USA,

²Stanford Functional Genomics Facility, 269 Campus Drive West, Stanford, CA 94305, USA,

³Developmental, Cell and Molecular Biology Group, Department of Biology, Box 91000, Duke University, Durham, NC 27708, USA, and

⁴Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA

Received 3 August 2006; revised 26 September 2006; accepted 5 October 2006.

*For correspondence (fax +1 612 624 6264; e-mail katagiri@umn.edu).

†Present address: 371 Serra Mall, Stanford University, Stanford, CA 94305, USA.

Summary

Studies of the behavior of biological systems often require monitoring of the expression of many genes in a large number of samples. While whole-genome arrays provide high-quality gene-expression profiles, their high cost generally limits the number of samples that can be studied. Although inexpensive small-scale arrays representing genes of interest could be used for many applications, it is challenging to obtain accurate measurements with conventional small-scale microarrays. We have developed a small-scale microarray system that yields highly accurate and reproducible expression measurements. This was achieved by implementing a stable gene-based quantile normalization method for array-to-array normalization, and a probe-printing design that allows use of a statistical model to correct for effects of print tips and uneven hybridization. The array measures expression values in a single sample, rather than ratios between two samples. This allows accurate comparisons among many samples. The array typically yielded correlation coefficients higher than 0.99 between technically duplicated samples. Accuracy was demonstrated by a correlation coefficient of 0.88 between expression ratios determined from this array and an Affymetrix GeneChip, by quantitative RT-PCR, and by spiking known amounts of specific RNAs into the RNA samples used for profiling. The array was used to compare the responses of wild-type, *rps2* and *ndr1* mutant plants to infection by a *Pseudomonas syringae* strain expressing *avrRpt2*. The results suggest that *ndr1* affects a defense-signaling pathway(s) in addition to the *RPS2*-dependent pathway, and indicate that the microarray is a powerful tool for systems analyses of the Arabidopsis disease-signaling network.

Keywords: expression profiling, systems analysis, calibration probe, stable genes-based quantile normalization, statistical model.

Introduction

In order to understand the behaviors of complex biological systems, it is crucial to collect large amounts of quantitative and context-dependent data, such as mRNA expression profiles, under highly controlled environments with suffi-

cient spatial and temporal resolution. It is likely that such data will need to be collected in numerous similar, but subtly different, experimental conditions, and that accurate measurements will be crucial for distinguishing small differences.

If a data-collection method is expensive, such a study involving analysis of many biological samples could be economically impossible. Therefore the economy and accuracy of highly parallel data-collection methods is a central issue for systems analysis.

There are two major applications of mRNA expression profiling. One is for gene discovery by the identification of candidate genes based on expression changes associated with a biological process of interest. The other is use of the expression levels of many genes as a detailed 'snapshot' that describes the state of a biological system under a particular set of conditions. These snapshots can be used to infer the network structure of a system when combined with specific perturbations to the system (Ideker *et al.*, 2001). For example, expression profiles provide detailed phenotypic descriptions (snapshots) of mutants (perturbations), and such snapshots can be used to infer the relationships among the mutant genes (Glazebrook *et al.*, 2003; Katagiri and Glazebrook, 2003). Snapshots over the course of time can reveal the dynamics of system behavior. For example, a temporal examination of expression profiles revealed that the path for differentiation of HL60 cells into neutrophil cells after induction with retinoic acid is different from the path after induction with DMSO, even though the final states as differentiated neutrophils are identical (Huang *et al.*, 2005). Snapshots can also be considered as highly parallel and sensitive assays in screens for mutations or chemicals with specific effects on a system.

For gene-discovery applications, it is important to monitor the expression of a large number of genes, but the number of biological samples to be profiled is often small. For these studies, commercially produced, high-density *in situ* oligonucleotide-synthesized arrays, such as those produced by Affymetrix or Nimblegen, are often a good option, as they have a deep coverage of the genome and provide high measurement accuracy. In contrast, for system snapshot applications it may be sufficient to monitor only a few hundred genes, but the number of biological samples needed could be very large. The use of high-density arrays for large numbers of samples can become prohibitively expensive.

A possible solution to this problem is to use inexpensive, small-scale custom arrays produced in-house that represent only the genes of interest. However, it is challenging to obtain accurate measurements with conventional small-scale microarrays, for several reasons. First, hybridization is typically done using a two-color method, in which mRNA samples from two biological samples are labeled with two different fluorophores and hybridized to a single array. The comparison of data from two biological samples that were not paired in a single hybridization results in a high error rate, because calculation of the expression ratio between unpaired samples involves at least two other measured values. In system snapshot applications, it is desirable to

compare many different samples in many different combinations, so the conventional two-color method is not ideal. Second, powerful statistical methods that are used to remove systematic errors from data obtained with large-scale arrays are unsuitable for small-scale microarrays. This is because these methods are based on the assumption that some statistical characteristics of measurements from a large number of randomly selected spots are the same, irrespective of the biological samples analysed or of particular local areas of an array. This assumption is generally false for small-scale microarrays, as the majority of the probes monitor genes for which mRNA levels change under the experimental conditions of interest. One example is normalization of data from different arrays. Normalization methods commonly used for data from large-scale microarrays, such as global normalization (GN; Affymetrix Microarray Suite User Guide, ver. 5 at <http://www.affymetrix.com/support/technical/manuals.affx>) and quantile normalization (QN; Bolstad *et al.*, 2003; Irizarry *et al.*, 2003), assume that the mean expression values or the expression value distributions of the genes, respectively, are the same among arrays, irrespective of biological samples. Therefore they cannot be used for data from small-scale microarrays. Another example is correction for effects of print tips and uneven hybridization, because this assumes a balanced distribution of the log ratio between two-color signals for a large number of spots printed by different print tips or for different areas of an array (Yang *et al.*, 2002). Due to the inability to apply these statistical correction methods, significant systematic errors cannot be removed from conventional small-scale microarray data.

We describe the design, production and use of a small-scale microarray printed with 576 long oligonucleotide probes for the study of Arabidopsis responses to pathogen interactions. The probe sequences consisted of one probe sequence for each of 464 pathogen-responsive genes and 107 genes with relatively stable expression levels, and five random sequences unrelated to the Arabidopsis genome sequence, to be used in spiking controls. We implemented several unconventional design components to improve the quality of the data obtained using this microarray, which we call a 'miniarray'. The expression values, rather than ratios, were measured by adding a common probe to each spot and using its signal to calibrate the amount of probe printed in each spot (Dudley *et al.*, 2002). Second, we dedicated 107 probes to genes expressed at relatively constant levels and used these together with a modified QN method, resulting in excellent array-to-array normalization. Third, we used an array-printing pattern that allows us to use a statistical model to estimate and correct systematic errors arising from effects of print tips and uneven hybridization. In practice, we obtained very high technical reproducibility and accuracy using the miniarray. All the features of the miniarray can be implemented easily for any small-scale printed microarray.

Highly reproducible and accurate small-scale microarrays should be helpful for many applications requiring measurement of modest numbers of mRNAs in large numbers of samples.

Results and discussion

Selection and evaluation of genes for the miniarray

We made the miniarray for studies of the Arabidopsis response to pathogen attack. We selected 337 Arabidopsis genes based on publicly available expression-profile data. The goals of the gene selection for the miniarray were: (i) the expression patterns of the selected genes should represent diverse expression patterns observed among all the Arabidopsis genes in pathogen-related experiments; (ii) the expression levels of the selected genes should allow accurate measurement. To achieve goal (i), the Arabidopsis genes that respond to pathogens were classified into groups according to the similarity in their expression patterns through many different pathogen-related experiments, and representatives of each group were selected. To achieve goal (ii), for the representatives of each group, genes with higher maximum expression levels and/or larger differences between their maximum and minimum expression levels were preferentially selected.

Expression-profile data in AffyWatch (<http://affymetrix.arabidopsis.info/AffyWatch.html>) were used for the evaluation of the gene selection. The data were generated using the ATH1 GeneChip, which covers most of the Arabidopsis protein-coding genes. In this data set, 7002 pathogen-responsive genes were identified, as described in Experimental procedures. The \log_2 expression ratios between treated and control samples were used in the following calculation. Among the 337 selected genes, 321 were found among the 7002 genes. For each of the 6681 excluded genes (7002 minus 321 genes), its closest representative in the 321 selected genes was identified according to the highest uncentered Pearson correlation. This distribution of the highest correlations is a measure of the degree of representation of all the expression patterns among the 7002 genes by a set of selected genes. Figure 1(a) shows the distribution of these best correlation values for the 6681 excluded genes (Miniarray, open bars). To increase the resolution in the high correlation range, the arc cosine value of the correlation value is used in Figure 1(a) (unit, radian; 0 and 1.57 radians for the correlations 1 and 0, respectively). Twenty sets of 321 genes were randomly sampled from the 7002 genes, and the 20 resulting distributions of the highest correlations were averaged and compared (random, shaded bars). Random sampling of a given number of genes should yield a set of genes very close to the optimal representation for the sample size when the sample size is sufficiently large. Only limited improvement in the distribution of the highest

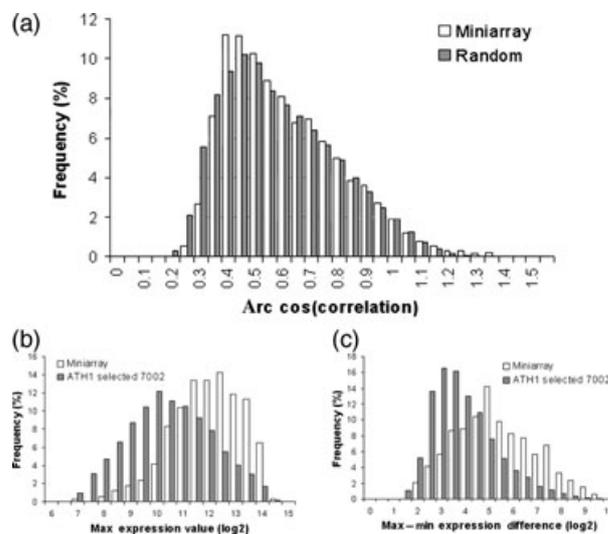


Figure 1. Pathogen-responsive genes for the miniarray.

(a) The distribution of the highest correlation between each of the 6681 excluded genes and the 321 genes selected for the miniarray (open bars) or 321 randomly selected genes (shaded bars). For the distribution with the 321 randomly selected genes, 20 random sets of 321 genes were generated and the distribution results averaged. The largest SEM among shaded bars for the random sets was 0.14% on the frequency scale. To increase the resolution in the high-correlation range, the arc cosine-transformed correlation was used (unit, radian) in the horizontal axis.

(b, c) The distribution of (b) maximum expression values; (c) difference between maximum and minimum expression values of each gene among the 321 genes selected for the miniarray (open bars) or among the population of 7002 genes (shaded bars). Expression values are in \log_2 -scale.

correlations was observed when a set of 963 (three times more) genes were randomly sampled (not shown), which indicates that the sample size of 321 is sufficiently large. The representation by our selection of 321 genes was very similar to the average of the 20 randomly selected sets of 321 genes, and therefore very close to the optimum. Our selected 321 genes are highly enriched for genes that have high maximum expression levels and/or large differences between the maximum and minimum expression levels compared with the original population of 7002 genes (Figure 1b,c). Thus we achieved goals (i) and (ii) in the gene selection for the miniarray.

We added 127 empirically chosen marker genes for various pathogen responses, resulting in 464 pathogen-responsive genes for the miniarray. In addition, we selected 107 normalization genes the expression levels of which were relatively stable among pathogen experiments, and that represented a wide range of expression levels. Thus a total of 571 Arabidopsis genes were selected for the miniarray (Table S1). Using Picky (Chou *et al.*, 2004), we selected a single probe sequence for each gene, as well as five random probe sequences not present in the Arabidopsis genome for use as RNA spiking controls (total 576 target probes). Picky selects probe sequences for similar thermodynamic

properties in hybridization and for low cross-hybridization probabilities.

Measurement of raw expression value

The amount of probe printed by an arrayer varies from spot to spot. Therefore information about the amount of probe DNA printed is needed to calibrate the signal from the target hybridized to each spot and to calculate the expression value. For this purpose, we used a calibration (Cal) probe ('common oligo reference' in Dudley *et al.*, 2002). A fixed amount of a Cal probe with another random sequence selected by Picky was mixed into each target probe solution before printing. This allowed the amount of target probe in each spot to be determined using the signal from hybridization of a Cy5-labeled oligonucleotide that is complementary to the Cal probe (Cy5-cCal). We define the raw expression value of a particular gene at a spot as the signal ratio between the Alexa Fluor 555-labeled target (green fluorescence) and the Cy5-cCal (red fluorescence). We used the median of the ratios at pixels in a spot to represent the ratio for that spot. To increase the accuracy of the signal measurement by the scanner, we used two different photo-multiplier tube voltages for the target signal (green), and combined them to obtain the raw expression value for each spot (Dudley *et al.*, 2002).

Other investigators have used a standard RNA sample common to all the experiments (Iyer *et al.*, 1999) or genomic DNA (Talaat *et al.*, 2002) as references. We did not want to use a standard RNA sample as the reference, because it is difficult to prepare a standard RNA sample with consistent quality and composition of mRNA species that can be used for many experiments over an indefinite period. In addition, if RNA species for some probes are scarce in the standard sample, so that the resultant signals are weak, measurements made using those probes will have large errors. We tried genomic DNA as a reference, but found that it was unsatisfactory. Signals from each spot for the target and the genomic DNA reference were highly correlated, clearly indicating that the genomic DNA did not accurately report the amount of probe DNA in each spot (not shown). As the same fluorescence values for two channels were obtained when an array was scanned in two channels either at the same time or separately, optical reasons (such as leak of excitation or emission light into the other channel, fluorescence resonance energy transfer) were excluded. We speculate that when target RNA hybridizes to the probe, only a small part of the target (50–70 nt) is occupied by the probe and the rest of the target RNA sequence is available for hybridization with genomic DNA. As the available part of the target sequence is generally much longer than the probe, the genomic DNA preferentially hybridizes to the target RNA.

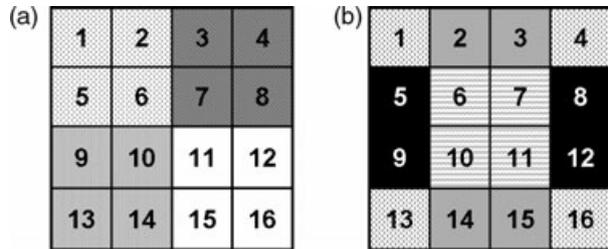


Figure 2. Probe-sharing sub-array groups.

Eight probe-sharing groups, each of which contains four sub-arrays. 72 probes were assigned to each of the probe-sharing sub-array groups and had duplicates in each of the four sub-arrays in their corresponding probe-sharing groups.

Overlapping sub-array groups sharing the same target probes

It is possible to estimate systematic errors, such as variation due to print tips and uneven hybridization, if replicates of each target probe are placed appropriately on the array. The miniarray had eight replicates of each of 576 target probes on a single array (4608 spots per array). A 4×4 print-tip block (16 print tips) was used to print the array (288 spots per print tip). A sector of the array that was printed by a single print tip is called a sub-array. The 16 print tips resulted in 16 sub-arrays, as shown in Figure 2. Each target probe was printed twice in each of four sub-arrays, such that the top half of a sub-array is exactly the same as the bottom half. Different shades of the sub-arrays in Figure 2 show eight different groups of four sub-arrays that share the same target probes. Seventy-two target probes were printed to sub-arrays 1, 2, 5 and 6 (Figure 2a) in duplicate, another 72 target probes were printed to sub-arrays 6, 7, 10 and 11 (Figure 2b) in duplicate, and so on. The probe-sharing groups are symmetrical and circularly overlapping. These sharing groups allow use of a linear model for the entire array, which enables estimation of systematic errors as described below. A simple procedure to generate this printing pattern is described in Appendix S1.

Use of a statistical model to remove systematic errors

The symmetrically overlapping probe-sharing sub-array group design allows us to fit the following linear model to the raw expression value for each spot:

$$S_{ij} = \mu + A_i + B_j + \varepsilon_{ij} \quad (\text{model1})$$

where S_{ij} denotes the \log_2 -transformed, raw expression value for the spot, μ denotes a constant, A_i and B_j denote the effects of i th gene and j th sub-array (or print tip), and ε_{ij} denotes the residual. For each gene, $\mu + A_i$ was defined as the estimated expression value, although addition of the constant μ is arbitrary considering the following array-to-array normalization process. The interaction $A:B_{ij}$ was

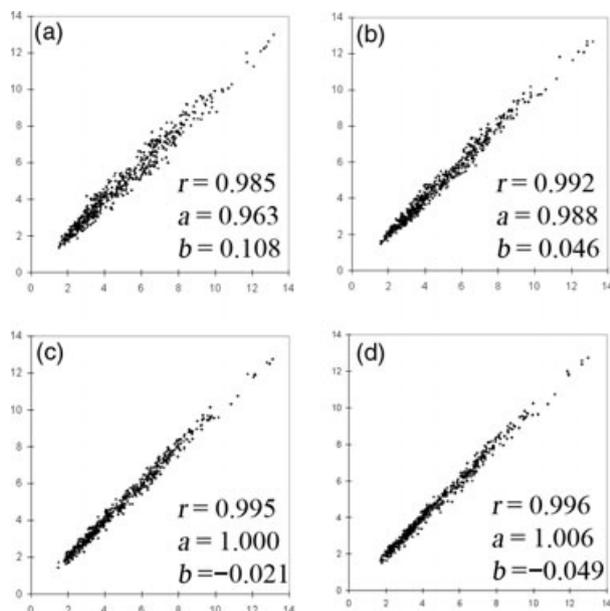


Figure 3. Application of statistical models improved technical reproducibility.

Estimated expression values for all genes from technical duplicates with the *Psm* sample were compared after four processing methods.

(a) For each probe, the data from one sub-array were randomly selected, and the average of the duplicates in the sub-array was used as the estimated expression value.

(b) The average of eight replicates for each probe was used as the estimated expression value.

(c) The estimated expression value was calculated using model 1.

(d) The estimated expression value was calculated using model 2.

r, Correlation coefficient; *a*, slope; *b*, *y*-intercept of the regression line.

ignored. Although ANOVA indicated that all A_i , B_j and $A_i B_j$ were significant, using the data for each of four arrays used in Figure 3 (not shown), the corrected Akaike's information criterion (AICc) values (Burnham and Anderson, 2002) were smaller with the model using only A_i and B_j than with the model including $A_i B_j$ (Table S2). A smaller AICc value indicates a better model in terms of the balance between goodness-of-fit and model complexity. Each sub-array specifies a particular spatial area, so model 1 can account for systematic errors arising from uneven hybridization at a spatial resolution corresponding to the sub-arrays, as well as those arising from print-tip effects.

As shown below, after fitting model 1 the spatial distribution of residuals within each sub-array was uneven. We added an arbitrary fourth-order smoothing function $f_j(x_j, y_j)$ specified for the j th sub-array to model 1, yielding model 2:

$$S_{ij} = \mu + A_i + B_j + f_j(x_j, y_j) + \varepsilon_{ij}, \text{ where}$$

$$f_j(x_j, y_j) = \sum_{p=1}^4 \sum_{q=0}^p a_{jpq} X_j^q Y_j^{p-q} \quad (\text{model 2})$$

x_j and y_j are spatial coordinates along the row and the column of the j th sub-array, with the center of the j th sub-array adjusted to (0, 0). Model 2 gave a lower AICc value than model 1 (Table S2), indicating that addition of the smoothing function is worthwhile. Again, $\mu + A_i$ was defined as the estimated expression value for each gene.

Array-to-array normalization method

Array-to-array normalization was performed by applying a modified QN to the expression values for the 107 normalization genes. We modified the QN (Bolstad *et al.*, 2003; Irizarry *et al.*, 2003) to compensate for the fact that the number of normalization genes is not very large. The expression values of the normalization genes were ranked for each array. The average of the expression values for each group of four genes with consecutive ranks was calculated (104 averaged values per array). These averaged values from each of the multiple arrays in an analysis set were subjected to QN. Expression values within the range of the averaged expression values were normalized proportionally according to the normalized values of two flanking averaged values. Although we chose normalization genes that covered a wide range of expression values, some pathogen-responsive genes had expression values outside the range. This may be unavoidable in our system, as some pathogen-inducible genes are expressed at extremely high levels. Expression values of these genes were normalized using a regression line fitted to the averaged expression values of normalization genes within $3 \log_2$ from the boundary of the range. The median of the averaged expression values of normalization genes was arbitrarily adjusted to 4.5 to make all the \log_2 -transformed expression values positive. We call this normalization method stable genes-based quantile normalization (SBQ).

The performance of SBQ was evaluated using ATH1 GeneChip data for three diverse pathogen treatments from separate AffyWatch experiments and their corresponding control treatments. For each treatment-control combination, the raw intensity data for the entire arrays in .CEL files were converted into the \log_2 -transformed expression values using RMAEXPRESS with either QN (whole-array QN) or no normalization. We used the whole-array QN output as the standard for properly normalized data. The expression values for the 535 genes which are common to the ATH1 GeneChip and the miniarray, and exclude genes subject to cross-hybridization, were extracted from the no-normalization output. The no-normalization output of the 535 genes was subjected to SBQ, QN, GN, or no normalization. Note that, at this stage, normalization methods were applied only to the data consisting of the biased set of 535 genes. After the normalization, for each gene the expression values from the replicates were averaged, the \log_2 expression ratio

between the treatment and control was calculated, and the difference between this \log_2 expression ratio and the \log_2 expression ratio similarly calculated from the whole-array QN output was determined. For each treatment, the square of the \log_2 expression ratio difference was calculated for each gene, and the average of the squared values through the genes is shown in Table 1. For all the treatments, SBQ resulted in the \log_2 expression ratio clearly closest to that determined from the whole-array QN output. As expected, QN and GN did not perform well. Even in the worst case [*Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*) treatment], SBQ yielded the average of the squared difference from the whole-array QN data of 0.017 (\log_2)², which is equivalent to approximately 9.5% difference in the expression ratio. We conclude that the performance of SBQ is satisfactory.

Importance of amplifying the target

For the miniarray, a modified Eberwine method (Phillips and Eberwine, 1996), a common target-labeling method for commercial microarray platforms, was used to linearly amplify the labeled target as cRNA. This resulted in excellent technical reproducibility and accuracy, as shown below. Initially, we tried labeling the target by directly incorporating a fluorescence-labeled nucleotide into the first cDNA strand, which is a target-labeling method commonly used for printed microarrays. In agreement with Stoyanova *et al.* (2004), the direct cDNA-labeling method resulted in poor technical reproducibility. Furthermore, it resulted in very poor accuracy based on comparison with results from Affymetrix GeneChips (data not shown). These results suggest that some data from microarray experiments that employed direct cDNA labeling may be inaccurate, and should be interpreted cautiously.

High technical reproducibility

We analysed data obtained from two biological samples. Leaf tissues of wild-type *Arabidopsis thaliana* accession Col-0 were harvested 1 day after infection by either the bacterial pathogen *P. syringae* pv. *maculicola* strain ES4326 at a dose of 1×10^5 cfu ml⁻¹ (*Psm* sample) or a mock treatment (mock sample). Each RNA sample was divided into two aliquots,

Table 1 Average of squared difference in \log_2 expression ratio from the whole-array quantile normalization (QN)

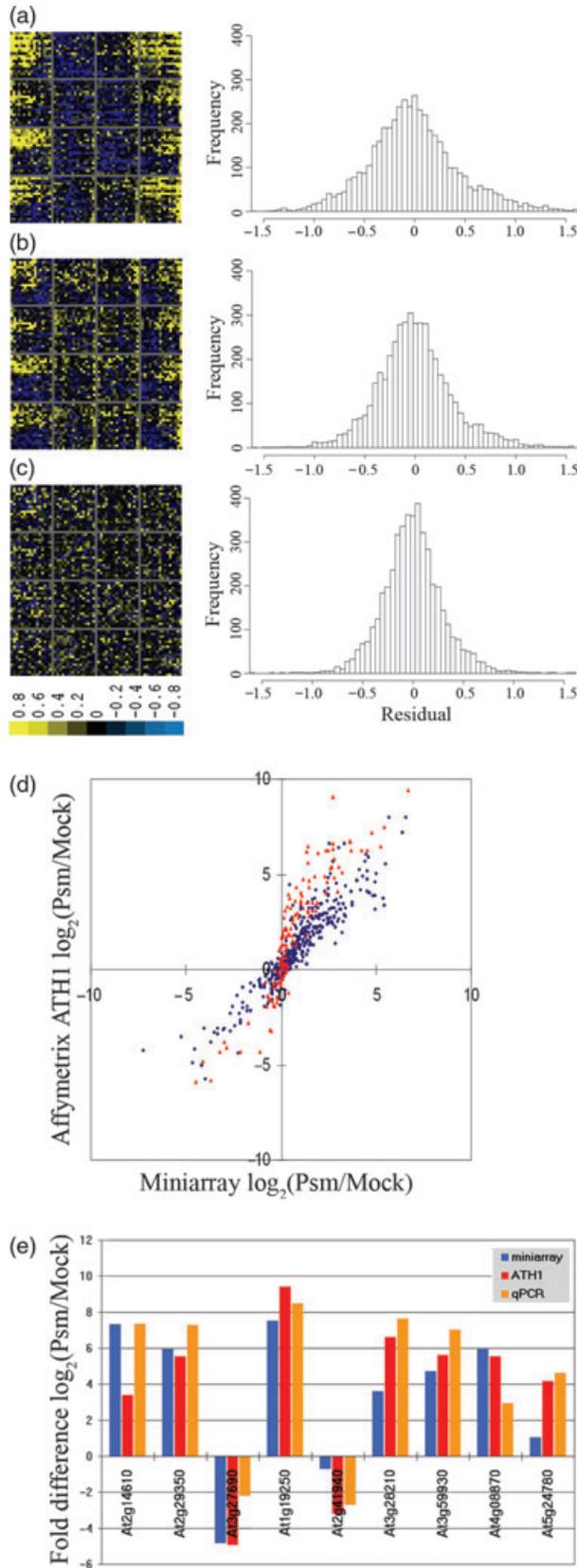
Sample compared (time)	SBQ	QN	GN	No normalization
<i>Pst</i> (24 h)	0.017	0.619	0.603	0.062
<i>Flg22</i> (4 h)	0.015	0.186	0.180	0.040
<i>Botrytis</i> (48 h)	0.009	0.471	0.448	0.009

SBQ, stable genes-based quantile normalization; GN, global normalization.

which were labeled separately and hybridized to the miniarrays. We refer to these comparisons as technical duplicates. To assess the effects of statistical methods on technical reproducibility, the spot-by-spot raw expression values were processed into estimated expression values for each gene in four different ways for the *Psm* sample (Figure 3). First, for each probe one print tip out of four was selected randomly, and the raw expression values from two spots printed by this print tip were averaged (this mimics a conventional small-scale array, Figure 3a); second, the raw expression values from eight replicated spots for each probe were averaged (Figure 3b); third and fourth, models 1 and 2 (Figure 3c,d respectively) were fitted to the raw expression values. After each processing method, the output data for the two arrays were normalized using SBQ. Comparison of Figure 3(a,b) shows that increasing the number of replicates from two to eight, and/or printing each probe with multiple print tips, improved the technical reproducibility as both the correlation coefficient (*r*) and the slope (*a*) of the regression line became closer to 1. Comparison of Figure 3(b,c) shows that fitting model 1 yielded better technical reproducibility than averaging. Comparison of Figure 3(c,d) shows that inclusion of the smoothing function does not obviously affect technical reproducibility. Similarly, comparison of the technical duplicates of the mock sample showed an improvement in technical reproducibility by increasing the number of spots per probe and by fitting the statistical models, as the correlation coefficients between the technical duplicates were 0.942, 0.966, 0.978 and 0.979 for duplicates by one print tip, average of eight spots, models 1 and 2, respectively (not shown). We conducted two additional comparisons of technical duplicates, which resulted in correlation coefficients of 0.996 and 0.991 after fitting model 2 (not shown). These high correlation coefficients between technical duplicates demonstrate the high technical reproducibility achieved by the miniarray.

The spatial distributions of the residuals on the miniarray vary depending on the processing methods used (Figure 4a–c, left panels). The uneven spatial residual distribution after averaging eight replicates (Figure 4a) is to some extent corrected by model 1 (Figure 4b), although an uneven residual distribution within each sub-array is still noticeable. Inclusion of the smoothing function in model 2 evened out the spatial distribution of residuals (Figure 4c). The right panels of Figure 4(a–c) show that the residual values were reduced by implementing model 1 (Figure 4b), and reduced further by inclusion of the smoothing function (Figure 4c). Based on this result that the distribution of the residuals is even using model 2, and the fact that model 2 gives a lower AICc than model 1, we decided to use model 2 as the routine method for calculation of the estimated expression value for each gene.

The eight replicates of each probe make the estimated expression values robust against partial loss of data. When



data for 922 randomly chosen spots (20% of the total) were removed, the correlation coefficient and the slope of the regression line of the estimated expression values obtained using the complete data and the depleted data were 0.9988 ± 0.0001 and 0.9999 ± 0.0018 , respectively (mean \pm SD, 100 simulations). Losing data from one area of an array is a common technical problem that can be caused by a bubble in the hybridization solution. To mimic this, data in a randomly selected 17×17 -spot area (289 spots) were removed. The correlation coefficient and the slope of the regression line of the estimated expression values obtained using the complete and the depleted data were 0.9998 ± 0.0001 and 0.9996 ± 0.0020 , respectively (mean \pm SD, 100 simulations). Therefore the estimated expression values remain reliable when some data are lost due to small defects in array printing or hybridization.

Based on the high level of technical reproducibility obtained, we conclude that it is unnecessary to run technical replicates when using the miniarray. If a conventional two-color method were used, a dye-swap technical replicate for each pair of samples would be necessary. Omission of technical replicates dramatically reduces the cost of experiments.

High measurement accuracy

To evaluate the accuracy of measurements, the estimated expression values obtained from the miniarray were compared with those obtained from the Affymetrix ATH1 GeneChip with the same *Psm* and mock samples. High technical reproducibility and accuracy of the ATH1 array have been demonstrated (Redman *et al.*, 2004). As differences in hybridization efficiencies of different probes for the same gene are expected, data generated using different microarray platforms must be compared as expression ratios rather than expression values. The \log_2 -transformed expression ratios between the *Psm* and mock samples were calculated using the mean values of the technical duplicates of each sample for the miniarray, and using single-array data for each sample from the ATH1 array. The expression ratios from the two platforms for the common 535 genes (represented by both dark blue and red dots) are shown in

Figure 4. Precision and accuracy of the miniarray.

(a–c) Spatial (left panels) and value (right panels) distributions of the residuals at spots in one of the *Psm* arrays: (a) when the average of eight replicates was used; (b) when model 1 was used; (c) when model 2 was used. Yellow and blue indicate positive and negative residuals, respectively (see color scale). (d) Comparison of expression ratios determined from measurements made by the miniarray and the Affymetrix ATH1 GeneChip. The \log_2 -transformed expression ratios between the *Psm* and mock samples are shown. Red dots, expression ratios of 140 genes that had the 140 lowest expression values measured by the miniarray in either *Psm* or mock samples among the 535 genes common to the two arrays; dark blue dots, expression ratios of the rest (395) of the 535 common genes.

(e) Comparison of expression ratios between *Psm* and mock samples for selected genes measured by the miniarray, ATH1 GeneChip and qRT-PCR.

Figure 4(d). Virtually no gene shows a clear qualitative discrepancy in expression ratios (for example, one gene induced in one platform and repressed in the other): the second (top left) and fourth (bottom right) quadrants of the plot are almost empty. The correlation coefficient and the slope of the regression line are 0.88 and 1.18, respectively. This level of agreement with GeneChip results is substantially higher than those reported for other custom microarray platforms (Barczak *et al.*, 2003; Schlingemann *et al.*, 2005).

The slope of the regression line, which is >1 , indicates that the expression ratio measured by the miniarray tends to be smaller than that measured by the ATH1 array. We hypothesized that this underestimate of the expression ratio by the miniarray is a result of the narrower dynamic range of the miniarray. The dynamic range is the range in which the relationship between the fluorescent signal and the amount of target in the hybridization solution is linear. In fact, when 140 genes showing the lowest expression values measured by the miniarray (red dots) were removed, the correlation coefficient and slope improved to 0.92 and 1.06, respectively (Figure 4d, dark blue dots only). Therefore the dynamic range of the miniarray measurement is narrower on the lower end than for the ATH1 array. The difference in dynamic range varies by gene, but a rough estimate is that the lower end of the miniarray dynamic range is approximately eight times ($3 \log_2$) higher than that of the ATH1 array.

For nine genes, we also compared the expression ratio determined by quantitative RT-PCR (qRT-PCR) (Figure 4e). These genes were chosen based on the diversity among expression level, induction or repression after *Psm* inoculation, degree of agreement between the microarray platforms and possibility of cross-hybridization. Although some expression ratios determined from the miniarray measurements were underestimated, due to its narrower dynamic range, overall data from both microarray platforms agreed with data from qRT-PCR.

We further tested the accuracy of the miniarray measurement using RNA spiking controls. The miniarray has five spiking control probes (1–5). Each of the spiking control RNAs 1, 3, 4, and 5 has an approximately 600-nt sequence unrelated to the Arabidopsis genome sequence, the same sequence as the corresponding spiking control probe, and a polyA tail. The spiking control RNAs 1, 3, 4 and 5 were mixed into $1 \mu\text{g}$ total RNA at 400, 40, 4 and 0.4 pg, respectively. As a negative control, spiking control RNA 2 was not included. Fifteen slides were hybridized with targets including the spiking controls. As shown in Figure 5, the linearity of the dose–signal relationships was excellent between spiking controls 3–5. The expression value for spiking control probe 1 appeared to have reached saturation. About 400 out of 571 genes measured in the experiment above were within the range covered by spiking controls 3–5.

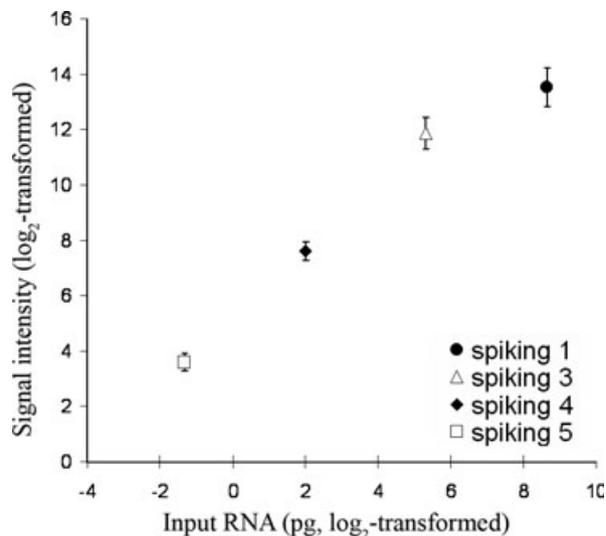


Figure 5. Expression value measurements for spiking controls. Amounts and mean of estimated expression values of spiking controls 1, 3, 4 and 5 from 15 miniarrays are plotted. Both axes are in \log_2 -scale. Error bars, SEM.

ndr1 affects a signaling pathway(s) in addition to the *RPS2*-dependent one

The miniarray was used to compare the responses of wild-type (*RPS2 NDR1*), *rps2-101C* and *ndr1-1* plants to a *Pst* strain expressing *avrRpt2* (*Pst/avrRpt2*). The resistance response is conditioned by the *RPS2* resistance gene when the bacterial strain expresses the corresponding avirulence gene *avrRpt2* (Bent *et al.*, 1994; Mindrinos *et al.*, 1994). The *ndr1* mutation blocks the resistance response mediated by *RPS2* (Century *et al.*, 1995). Leaves from the plants were collected 6 h after inoculation of *Pst/avrRpt2* at 1×10^8 cfu ml⁻¹. Two independent sets of experiments were performed. The expression data obtained from the miniarrays were analysed using the significance analysis of microarrays (SAM; Tusher *et al.*, 2001). In this experiment, another *ndr1-1* plant was included as the fourth genotype, the identity of which was unknown to the experimenter (*ndr1-1* blind). The profiles from the *ndr1-1* and *ndr1-1* blind plants were used empirically to set a critical false discovery rate (FDR) in SAM. Between the *ndr1-1* and *ndr1-1* blind plants, no gene was found significant at $\text{FDR} \leq 0.25$ or lower. In other pairwise comparisons of genotypes, a more conservative critical FDR of 0.2 was used to discover the genes with significantly different expression levels. The numbers of genes tested significantly different were 70 and 165 comparisons of wild type and *rps2-101C*, and wild type and *ndr1-1*, respectively (the fold change and *q* value for each comparison are listed in Table S3). The genes listed as significantly different between wild type

Table 2 The *ndr1*-specific genes

Gene ID	PAMPs-inducible?	Annotation
At1g02450		NPR1/NIM1-interacting protein 1 (NIMIN-1)
At1g03850	Y	Glutaredoxin family protein
At1g08830		Superoxide dismutase (Cu-Zn) (SODCC)/copper/zinc superoxide dismutase (CSD1)
At1g11310	Y	Seven transmembrane MLO family protein/MLO-like protein 2 (MLO2)
At1g21240		Wall-associated kinase, putative
At1g28480	Y	Glutaredoxin family protein
At1g45145	Y	Thioredoxin H-type 5 (TRX-H-5) (TOUL)
At1g59870		ABC transporter family protein
At1g74710		Isochorismate synthase 1 (ICS1)/isochorismate mutase
At1g75040		Pathogenesis-related protein 5 (PR-5)
At2g05940	Y	Protein kinase, putative
At2g18680	Y	Expressed protein
At2g19190	Y	Light-responsive receptor protein kinase/senescence-responsive receptor-like serine/threonine kinase, putative (SIRK)
At2g28400	Y	Expressed protein
At2g29350		Tropinone reductase, putative/tropine dehydrogenase, putative
At2g29460	Y	Glutathione <i>S</i> -transferase, putative
At2g30490	Y	<i>trans</i> -cinnamate 4-monooxygenase/cinnamic acid 4-hydroxylase (C4H) (CA4H)/cytochrome P ₄₅₀ 73 (CYP73) (CYP73A5)
At2g30550	Y	lipase class 3 family protein
At2g37040	Y	Phenylalanine ammonia-lyase 1 (PAL1)
At2g40140	Y	Zinc finger (CCCH-type) family protein
At2g42360	Y	Zinc finger (C3HC4-type RING finger) family protein
At2g44490		Glycosyl hydrolase family 1 protein
At2g45170		Autophagy 8e (APG8e)
At2g45570		Cytochrome P ₄₅₀ 76C2, putative (CYP76C2) (YLS6)
At3g09010	Y	Protein kinase family protein
At3g12580	Y	Heat shock protein 70, putative/HSP70, putative
At3g13950	Y	Expressed protein
At3g28510		AAA-type ATPase family protein
At3g28540		AAA-type ATPase family protein
At3g28930		avrRpt2-induced AIG2 protein (AIG2)
At3g52430		Phytoalexin-deficient 4 protein (PAD4)
At3g54420	Y	Class IV chitinase (CHIV)
At4g04490	Y	Protein kinase family protein
At4g14365	Y	Zinc finger (C3HC4-type RING finger) family protein/ankyrin repeat family protein
At4g23150	Y	Protein kinase family protein
At4g25900	Y	Aldose 1-epimerase family protein
At4g34230	Y	Cinnamyl-alcohol dehydrogenase, putative
At4g35180	Y	Amino acid transporter family protein
At4g36990	Y	Heat shock factor protein 4 (HSF4)/heat shock transcription factor 4 (HSTF4)
At4g39030		Enhanced disease susceptibility 5 (EDS5)/salicylic acid induction deficient 1 (SID1)
At4g39830	Y	L-ascorbate oxidase, putative
At4g39950		Cytochrome P ₄₅₀ 79B2, putative (CYP79B2)
At5g08240	Y	Expressed protein
At5g12030		17.7 kDa class II heat shock protein 17.6A (HSP17.7-CII)
At5g13320	Y	Auxin-responsive GH3 family protein
At5g22570		WRKY family transcription factor
At5g26920	Y	Calmodulin-binding protein
At5g47120	Y	Bax inhibitor-1 putative/BI-1 putative
At5g52760	Y	Heavy metal-associated domain-containing protein
At5g60800		Heavy metal-associated domain-containing protein

and *rps2-101C* were almost all different between wild type and *ndr1-1* (66 out of 70 genes; among the 66 genes, 22 were expressed at levels higher than the wild-type in both *rps2-101C* and *ndr1-1*, and 44 were expressed at lower levels in both *rps2-101C* and *ndr1-1*), which confirms that *ndr1-1* blocks the RPS2-mediated resistance pathway.

We were intrigued by the fact that substantially more genes were affected by *ndr1-1* than *rps2-101C* in comparison with the wild type. We speculated that *ndr1-1* affects some defense-signaling pathway(s) in addition to the RPS2-dependent pathway. We defined an *ndr1*-specific gene as a gene that was significantly different in the comparison between *ndr1-1* and wild type, but not between *rps2-101C* and wild type, and had lower expression levels in *ndr1-1* than the wild type. Fifty *ndr1*-specific genes were found among the 535 genes common between the miniarray and the ATH1 GeneChip (Table 2). A canonical pathogen-associated molecular patterns (PAMPs)-inducible marker gene *FRK1* (At2g19190; Asai *et al.*, 2002) was among the 50 genes, suggesting involvement of *NDR1* in the PAMPs-response pathway. To test this idea, we defined PAMPs-inducible genes as those for which expression levels were significantly increased in at least two of three PAMPs treatments: flagellin fragment 22 (flg22), bacterial lipopolysaccharides (LPS), and *Pst hrcC* (AffyWatch data) (He *et al.*, 2006). In this case, as the data did not allow us to determine the critical FDR empirically, a conventional critical FDR of 0.05 was used. Of the 535 genes common to both arrays, 121, including *FRK1*, were identified as PAMPs-inducible. Thirty-one of these 121 genes were *ndr1*-specific (Table S3). Fisher's exact test showed that the PAMPs-inducible genes were significantly enriched in the *ndr1*-specific genes ($P = 2.7 \times 10^{-10}$, two-tail). This result strongly suggests that *ndr1-1* affects at least part of the PAMPs response.

The notion that *ndr1* affects the PAMPs-response pathway is consistent with the prior observation that *ndr1-1* plants show higher susceptibility to the virulent strain *Pst* than the wild-type plant at early time points (Century *et al.*, 1995; Tao *et al.*, 2003). At early time points, the PAMPs-response pathway is compromised in *ndr1-1*, so the virulent strain grows better in *ndr1-1* plant than in wild type. Later, type III effectors delivered from the virulent strain inhibit the PAMPs-response pathway in wild-type plants, factors other than the PAMPs-response pathway become limiting on bacterial growth, and the difference in the bacterial counts between *ndr1-1* and wild type diminishes. RPS2 protein physically interacts with a negative regulator of the PAMPs response, RIN4 (Axtell and Staskawicz, 2003; Kim *et al.*, 2005; Mackey *et al.*, 2003). Although physical interactions between *NDR1* and RPS2/RIN4 proteins have not been demonstrated, *ndr1* may affect the PAMPs-response pathway via *RIN4*.

Concluding remarks

We have demonstrated that it is feasible to produce a highly reproducible and accurate, small-scale microarray printed with long oligonucleotide probes. Array-design components that enabled measurement of the expression value, use of SBQ, and use of powerful model-based statistics, as well as selection of the genes on the array, were critical for success. The designs and methods used for this miniarray could be easily implemented for other small-scale microarrays. To understand the behavior of a biological system, it is crucial to collect many system snapshots, such as those with different mutant backgrounds, with different concentrations of stimuli, at many different time points, etc. We expect that high-performance small-scale microarrays will make such projects involving system snapshot applications of microarrays much more economical.

Experimental procedures

All the oligonucleotide sequences and other information required to produce the miniarray, and all the microarray data obtained in this study, were submitted to Gene Expression Omnibus. Printed arrays are also available for use by other groups, provided compensation for materials and printing costs can be arranged. All the novel materials and Perl scripts used in the study, and detailed laboratory protocols, are available for non-profit research upon request. Refer to Appendix S1 for details of the methods.

Plants and bacteria

Arabidopsis was wild-type accession Columbia (Col-0), *ndr1-1* (Century *et al.*, 1995) or *rps2-101C* (Yu *et al.*, 1993), both in the Col-0 background. Plants were grown in a controlled environment chamber at 22°C, with 75% RH and a 12/12-h light/dark cycle. Leaves of 4-week-old plants were infiltrated with the bacterial strain *Psm* or *Pst/avrRpt2* at indicated doses using a needle-less syringe (Katagiri *et al.*, 2002).

Selection and evaluation of pathogen-responsive genes for the miniarray

Published expression-profile data (Glazebrook *et al.*, 2003; Tao *et al.*, 2003; van Wees *et al.*, 2003; Whitham *et al.*, 2003) generated using the AtGenome1 Affymetrix GeneChip® (Santa Clara, CA, USA), which has probes for approximately 8000 genes, and expression-profile data from experiments 'Oligogalacturonide treatment of seedlings', 'Response to *Erysiphe orontii* infection, time course experiment' and '*Botrytis cinerea* infection, 18 and 48 hpi' in the Integrated Microarray Database System (<http://ausubellab.mgh.harvard.edu/imds>), which were generated using the ATH1 GeneChip, were used for miniarray gene selection. As many of the data were from AtGenome1 experiments, we focused on the approximately 8000 genes represented by AtGenome1. In each experiment, the genes were classified into groups according to agglomerative hierarchical clustering based on the \log_2 -transformed expression ratios to appropriate controls, and they were labeled with arbitrary one-letter group names. The group names from multiple experiments for each gene were concatenated to make a string of letters representing the expression-pattern class of the gene defined in the

experimental set. Then, from each expression-pattern class, representative genes were selected that have high maximum expression levels and/or large differences between the maximum and minimum expression levels in the experimental set.

The data used for the evaluation were compiled from experiments 'AtGenExpress: Effect of ibuprofen, salicylic acid and daminozide on seedlings' (only the data for ibuprofen and salicylic acid were used), 'AtGenExpress: Pathogen Series: *Pseudomonas* half leaf injection', 'AtGenExpress: Response to bacterial-(LPS, HrpZ, Flg22) and oomycete-(NPP1) derived elicitors', 'AtGenExpress: Pathogen series: Response to *Botrytis cinerea* infection', 'AtGenExpress: Pathogen series: Response to *Erysiphe orontii* infection', 'AtGenExpress: Response to virulent, avirulent, typeIII-secretion system deficient and nonhost bacteria', 'AtGenExpress: Methyl jasmonate time course in wild type', 'AtGenExpress: Response to *Phytophthora infestans*', 'Impact of Type III effectors on plant defense responses', 'Hydrogen peroxide stress and Zat12 over-expression in Arabidopsis' and 'Transcriptome changes of Arabidopsis during pathogen and insect attack' in NASC AffyWatch. The .CEL files were pre-processed using RMAEXPRESS (<http://rmaexpress.bmbolstad.com>) using QN in each experiment to obtain the \log_2 -transformed expression values for each gene. Replicated array data were combined into single samples by averaging. Then the \log_2 expression values were subjected to QN through all the experiments. Note that this results in data for each sample having exactly the same distribution of \log_2 expression values. The genes that exceeded 3 \log_2 above the minimum value in at least one sample were selected (15 863 genes). The \log_2 expression values were floored at 1 \log_2 from the minimum value. In each experiment, the genes that show at least 1 \log_2 expression difference (twofold change) in at least one treated sample compared with the appropriate control were selected, and the genes that were selected in at least one experiment and that have corresponding AGI codes were designated pathogen-responsive (7002 genes). The ensuing evaluation procedure was performed as described in Results and Discussion.

Fabrication of the miniarray

Picky (Chou *et al.*, 2004) was used to select candidate probe sequences (50–70mers) for all Arabidopsis genes (The Arabidopsis Information Resource, <http://arabidopsis.org>) ATH1 ver. 012222004) and 10 500-nt random sequences. For each of the 464 pathogen-responsive genes and 107 normalization genes from Arabidopsis, one probe sequence was chosen based on its proximity to the 3' end of the transcript. Five spiking control probe sequences (1–5) were chosen from Picky-selected random sequences to use as RNA spiking controls.

The oligonucleotides for the target probes were suspended in Pronto! Universal Spotting Solution (Corning Life Sciences, Acton, MA, USA) at 0.9 mg ml^{-1} and aliquoted to six 96-well plates (designated 1A–6A). The Cal probe was added to each well at 0.25 mg ml^{-1} . The miniarray was printed on UltraGAPS slides (Corning Life Sciences) at the Stanford Functional Genomics Facility (Stanford, CA, USA).

Preparation of spiking control RNAs

We constructed plasmids pSP64bb1sp1, pSP64bb2sp3, pSP64bb2sp4 and pSP64bb2sp5 for *in vitro* transcription of spiking control RNAs 1, 3, 4 and 5, respectively. These sequences were submitted to GenBank (accessions DQ 480366, DQ 480367, DQ 480368 and DQ 480369, respectively). Each of these plasmids has

the SP6 promoter, a 579- or 584-nt backbone sequence, the corresponding spiking control probe sequence, and a polyA sequence, in this order in the vector pSP64poly(A) (Promega, Madison, WI, USA). The backbone sequences were made by concatenating nine of the Picky-selected random sequences. The plasmids were linearized by an *EcoRI* digestion at the end of the polyA sequence and transcribed *in vitro* with SP6 RNA polymerase to produce approximately 680-nt polyA RNAs that have the corresponding spiking probe sequences between the backbone sequence and the polyA tail. The spiking control RNAs were purified using the RNeasy mini kit (Qiagen, Valencia, CA, USA) after DNase treatment (Promega), and mixed into the target RNA samples.

Target preparation, miniarray hybridization and data collection

Total RNA was prepared using TRIzol (Invitrogen, Carlsbad, CA, USA) with the RNeasy mini kit (Qiagen) and multiple isopropanol precipitations. To amplify and label the target, AminoAllyl MessageAmp II aRNA Amplification kit (Ambion, Austin, TX, USA) and Alexa Fluor 555 carboxylic acid, succinimidyl ester (Invitrogen) were used. Five μg amplified RNA and 50 pmol of the Cy5-cCal oligonucleotide, in which the degree of labeling with fluorophore was adjusted to $\text{OD}_{260}/\text{OD}_{650} = 4.3$ by mixing with non-labeled cCal oligonucleotide, was included in 15 μl hybridization buffer (50% formamide, $5 \times \text{SSC}$, 0.1% SDS, 10 μg sheared salmon sperm DNA, (Eppendorf, Westbury, NY, USA). Slides were pre-treated according to Raghavachari *et al.* (2003), hybridized at 42°C for 24 h, and washed once in $2 \times \text{SSC}$, 0.1% SDS at 42°C for 5 min, twice in $1 \times \text{SSC}$ for 2 min at room temperature, and three times in $0.2 \times \text{SSC}$ for 2 min at room temperature. After drying by brief centrifugation, slides were scanned using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA, USA). Images were processed using GENEPIX ver. 6.0 to obtain raw intensity values for each spot. For each of the two colors, the median value for the background intensity was subtracted from the value of each pixel. Then the median value of the pixel-by-pixel ratios of the two-color intensity values (called the median of ratios by GENEPIX) was used as the raw signal value for each spot.

Data analyses

The raw signal intensity data were processed into the estimated expression values using a custom Perl script with *r* (<http://www.r-project.org>) as a module for model fitting. The processed data from multiple arrays were normalized by SBQ using a custom Perl script. Evaluation of SBQ used the data for three treatments, 24 h after infection of *Pst* infection (three replicates), 4 h after treatment with Flg22 peptide (three replicates), and 48 h after infection of *B. cinerea* (two replicates), and the corresponding controls from three experiments in AffyWatch: 'AtGenExpress: Response to virulent, avirulent, typeIII-secretion system deficient and nonhost bacteria', 'AtGenExpress: Response to bacterial-(LPS, HrpZ, Flg22) and oomycete-(NPP1) derived elicitors' and 'AtGenExpress: Pathogen series: Response to *Botrytis cinerea* infection', respectively. The evaluation of SBQ was performed as described in Results and Discussion.

The miniarray data for responses of wild-type, *rps2-101C* and *ndr1-1* plants to *Pst/avrRpt2* infection were normalized together by SBQ and analysed using SAM (two class-paired) (Tusher *et al.*, 2001) (<http://www-stat.stanford.edu/~tibs/SAM>). The PAMPs-responsive genes were identified using the flg22 and LPS data from 'AtGenExpress: Response to bacterial-(LPS, HrpZ, Flg22) and oomycete-(NPP1) derived elicitors' and *Pst hrcC* data from 'AtGenExpress: Response to virulent, avirulent, typeIII-secretion

system deficient and nonhost bacteria' in AffyWatch. The .CEL files corresponding to the treatments and the appropriate controls in each experiment were preprocessed using RMAEXPRESS with QN. The data from a treatment and the corresponding control were analysed using SAM (two class-paired, FDR = 0.05). For *flg22*, LPS and *Pst hrcC*, genes that showed expression levels significantly higher than the control at one or more time points were identified. Then the genes that were identified for at least two of the three PAMPs sources were defined as the PAMPs-inducible genes (1011 genes). From these 1011 genes, 121 genes that were among the 535 genes common to the miniarray and ATH1 GeneChip were used for comparison with the *ndr1*-specific genes.

Affymetrix ATH1 GeneChip experiment

Affymetrix experiments were performed as described in the Affymetrix technical manual. Total RNA (10 µg) was used for cDNA and cRNA synthesis with the Affymetrix GeneChip One-Cycle Target Labeling Kit. Hybridizations, washing, staining and scanning were performed at the University of Minnesota Microarray Facility (Minneapolis, MN, USA). RMAEXPRESS using QN was used to extract gene expression values from .CEL files.

qRT-PCR

qRT-PCR was performed using the Superscript III Platinum SYBR Green One-Step qRT-PCR kit (Invitrogen) and an Applied Biosystems 7500 Real Time PCR system (Applied Biosystems, Foster city, CA, USA) with the thermal cycling program: 48°C for 10 min, 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, 60°C for 1 min. Primer sequences used are shown in Table S4. Actin2 (At3g18780) was used as the internal reference gene. Data were analysed using the Applied Biosystems SEQUENCE DETECTION software (ver. 1.2), and normalized transcript levels were calculated as described (Livak and Schmittgen, 2001). The primer sequences are shown in Table S4.

Acknowledgements

We thank Zheng Jin Tu and the Minnesota Supercomputing Institute for access to and assistance with high-performance computers, Nick Hahn and the High Throughput Analysis Facility at University of Minnesota for operation of a liquid-handling robot, University of Minnesota Microarray Facility for hybridization of GeneChip arrays, Betsy M. Martinez-vaz and Arkady Khodursky for technical advice on microarray experiments, Hui-Hsien Chou for technical advice on Picky, Anton A. Sanderfoot for technical advice on DNA purification, Sanford Weisberg for statistics consulting, Shauna Somerville for help in selecting genes for the miniarray, Ken Vernick for letting us use his GenePix 4000B scanner, and Remco van Poecke for critical reading of the manuscript. This work was supported by a grant from the National Science Foundation, Arabidopsis 2010, grant number IBN-0419648, to J.G. and F.K. and a grant from the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2004-35301-14525, to F.K. R.M.M. was supported by a National Institutes of Health post-doctoral fellowship.

Note added in proof: Our suggestion that NDR1 acts through RIN4 is supported by a recent publication documenting physical interaction between these two proteins [Day et al. (2006) NDRI interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. *Plant Cell*, **18**, 2782–2791]

Supplementary Material

The following supplementary material is available for this article online:

Table S1 Genes represented by the miniarray

Table S2 Corrected Akaike's information criterion values for various models

Table S3 Miniarray expression profile results used in Table 2

Table S4 Primers used for qRT-PCR

Appendix S1 Detailed methods

This material is available as part of the online article from <http://www.blackwell-synergy.com>

References

- Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M. and Sheen, J. (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature*, **415**, 977–983.
- Axtell, M.J. and Staskawicz, B.J. (2003) Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell*, **112**, 369–377.
- Barczak, A., Rodriguez, M.W., Hanspers, K., Koth, L.L., Tai, Y.C., Bolstad, B.M., Speed, T.P. and Erle, D.J. (2003) Spotted long oligonucleotide arrays for human gene expression analysis. *Genome Res.* **13**, 1775–1785.
- Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J. and Staskawicz, B.J. (1994) *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science*, **265**, 1856–1860.
- Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
- Burnham, K.P. and Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. New York: Springer-Verlag.
- Century, K.S., Holub, E.B. and Staskawicz, B.J. (1995) *NDR1*, a locus of *Arabidopsis thaliana* that is required for disease resistance to both a bacterial and a fungal pathogen. *Proc. Natl Acad. Sci. USA*, **92**, 6597–6601.
- Chou, H.H., Hsia, A.P., Mooney, D.L. and Schnable, P.S. (2004) Picky: oligo microarray design for large genomes. *Bioinformatics*, **20**, 2893–2902.
- Dudley, A.M., Aach, J., Steffen, M.A. and Church, G.M. (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc. Natl Acad. Sci. USA*, **99**, 7554–7559.
- Glazebrook, J., Chen, W., Estes, B., Chang, H.S., Nawrath, C., Metraux, J.P., Zhu, T. and Katagiri, F. (2003) Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. *Plant J.* **34**, 217–228.
- He, P., Shan, L., Lin, N.C., Martin, G.B., Kemmerling, B., Nurnberger, T. and Sheen, J. (2006) Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in Arabidopsis innate immunity. *Cell*, **125**, 563–575.
- Huang, S., Eichler, G., Bar-Yam, Y. and Ingber, D.E. (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701.
- Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R. and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, **292**, 929–934.

- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- Iyer, V.R., Eisen, M.B., Ross, D.T. *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83–87.
- Katagiri, F. and Glazebrook, J. (2003) Local Context Finder (LCF) reveals multidimensional relationships among mRNA expression profiles of *Arabidopsis* responding to pathogen infection. *Proc. Natl Acad. Sci. USA*, **100**, 10842–10847.
- Katagiri, F., Thilmony, R. and He, S.Y. (2002) The *Arabidopsis thaliana*–*Pseudomonas syringae* interaction. In *The Arabidopsis Book* (Somerville, C.R. and Meyerowitz, E.M., eds). Rockville, MD, USA: American Society of Plant Biologists, pp. 1–35. doi:10.1199/tab.0009, www.aspb.org/publications/arabidopsis/.
- Kim, M.G., da Cunha, L., McFall, A.J., Belkadir, Y., DebRoy, S., Dangi, J.L. and Mackey, D. (2005) Two *Pseudomonas syringae* type III effectors inhibit RIN4-regulated basal defense in *Arabidopsis*. *Cell*, **121**, 749–759.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
- Mackey, D., Belkadir, Y., Alonso, J.M., Ecker, J.R. and Dangi, J.L. (2003) *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell*, **112**, 379–389.
- Mindrin, M., Katagiri, F., Yu, G.L. and Ausubel, F.M. (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell*, **78**, 1089–1099.
- Phillips, J. and Eberwine, J.H. (1996) Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods*, **10**, 283–288.
- Raghavachari, N., Bao, Y.P., Li, G., Xie, X. and Muller, U.R. (2003) Reduction of autofluorescence on DNA microarrays and slide surfaces by treatment with sodium borohydride. *Anal. Biochem.* **312**, 101–105.
- Redman, J.C., Haas, B.J., Tanimoto, G. and Town, C.D. (2004) Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J.* **38**, 545–561.
- Schlingemann, J., Habtemichael, N., Ittrich, C., Toedt, G., Kramer, H., Hambek, M., Knecht, R., Lichter, P., Stauber, R. and Hahn, M. (2005) Patient-based cross-platform comparison of oligonucleotide microarray expression profiles. *Lab. Invest.* **85**, 1024–1039.
- Stoyanova, R., Upson, J.J., Patriotis, C., Ross, E.A., Henske, E.P., Datta, K., Boman, B., Clapper, M.L., Knudson, A.G. and Bellacosa, A. (2004) Use of RNA amplification in the optimal characterization of global gene expression using cDNA microarrays. *J. Cell Physiol.* **201**, 359–365.
- Talaat, A.M., Howard, S.T., Hale, W.t., Lyons, R., Garner, H. and Johnston, S.A. (2002) Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res.* **30**, e104.
- Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G. and Katagiri, F. (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*, **15**, 317–330.
- Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
- van Wees, S.C., Chang, H.S., Zhu, T. and Glazebrook, J. (2003) Characterization of the early response of *Arabidopsis* to *Alternaria brassicicola* infection using expression profiling. *Plant Physiol.* **132**, 606–617.
- Whitham, S.A., Quan, S., Chang, H.S., Cooper, B., Estes, B., Zhu, T., Wang, X. and Hou, Y.M. (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *Plant J.* **33**, 271–283.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15.
- Yu, G.L., Katagiri, F. and Ausubel, F.M. (1993) *Arabidopsis* mutations at the *RPS2* locus result in loss of resistance to *Pseudomonas syringae* strains expressing the avirulence gene *avrRpt2*. *Mol. Plant–Microbe Interact.* **6**, 434–443.

Data depositions

Gene Expression Omnibus accessions: GPL3638, GSE4632, GSM103772, GSM103773, GSM103774, GSM103775, GSE4429, GSM99793, GSM99794, GSE5308, GSM124754, GSM124755, GSM124756, GSM124757, GSM124758, GSM124759, GSM124760, GSM124761.
GenBank accessions: DQ480366, DQ480367, DQ480368, DQ480369.