Supplementary Material – Methods

Microarray analysis

The eleven hybridizations were conducted by the German Resource Centre for Genome Research GmbH (now ATLAS Biolabs GmbH; both in Berlin, Germany) following the manufacturer's protocols (for details see Szakasits et al., 2008, in press). Affymetrix CEL files were read into the R statistical analysis environment (www.r-project.org) using the *affy* package of the Bioconductor suite (www.bioconductor.org). As 10–40% of probe sets are affected by updated gene annotation, chips were processed with current TAIR v8 probe-set annotation (Dai et al., 2005). Probe sequence specific 'background correction' (Wu et al., 2004) was performed using routines available in the Bioconductor *gcrma* package. Using the 'affinity' model, while 'MM' probes were employed for the determination of affinity parameters, only 'PM' probes were used for the probe-specific background correction. An inspection of exploratory pairwise scatter and 'MA' plots confirmed the need for inter-chip normalization. The thus required explicit normalization steps made a subtraction of the heuristic estimate for optical instrument background as offered in gcrma unnecessary. Defaults were used for all other *gcrma* parameters. As an examination of pairwise quantile-quantile plots showed only random fluctuations, inter-chip normalization could be achieved using quantile-quantile normalization (Bolstad et al., 2003). See 'Low-level microarray analysis and diagnostic plots' section of the Online Supplement for diagnostic plots and figures.

After normalization, robust summaries of probe set signals were obtained for each gene using an iterative weighted least squares fit of a linear probe level model (Bolstad, 2004) through the *fitPLM* function of the Bioconductor package *affyPLM*. This process automatically identifies unreliable chip areas and correspondingly downweights outlier probes. See Online Supplement for figures.

The normalized data on $\log_2$ scale were then fitted gene by gene with a linear model including hybridization batch effects, using the *lmFit* function (Smyth, 2004) of the Bioconductor package *limma*. The result-tables also include *q*-values as indicators of significance of contrasts after correction for multiple testing controlling the False Discovery Rate (Benjamini and Hochberg, 1995). For the statistical tests, individual gene variances have been moderated using an Empirical Bayes approach that draws strength from transferring variance characteristics from the set of all genes to the test for each individual gene (Smyth, 2004). Tests were restricted to a subset of 71 genes involved in UDP-glucuronic acid and *myo*-inositol metabolism according to the pathways provided by the MapMan project (Thimm *et al*., 2004). This considerably increases the statistical power of the testing procedure as it reduces the necessary correction for massive multiple testing.

References

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discoveryrate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B **57**, 289–300.

Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., Watson, S.J., and Meng, F. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. **33,** e175

Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F. and Spencer, F. (2004) A model based background adjustment for oligonucleotide expression arrays. J. Am. Stat. Assoc. **99**, 909–917.

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 bias and variance. Bioinformatics **19**, 185–193.

Bolstad, B.M. (2004) Low Level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization. PhD Dissertation. University of California, Berkeley.

Smyth, G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Gene. Mol. Biol. **3**, Article 3.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. **37**, 914-939