

Validation of NSF supported rice 45k array with light vs dark experiment

Pamela Ronald (PI), UCD, pcronald@ucdavis.edu

Robin Buell (co-PI), TIGR, rhuell@tigr.org

H.H. Chou (co-PI), ISU, hhchou@iastate.edu

David Rocke (co-PI), UCD, dmrocke@ucdavis.edu

Patrick Schnable (co-PI), ISU, schnable@iastate.edu

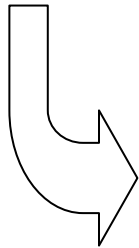
Procedure for validation of 45K microarray data

Statistical validation of microarray

- Well designed experimental plan
- Check error rate (e.g., control oligos such as hygromycin)
- Check reproducibility (Correlation Coefficient test)
- Find sources of variation (ANOVA test)



Isolate significant genes with p-value (<0.05) and fold change (2 fold)



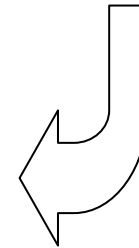
Functional validation by screening defective phenotypes



Functional classification and comparison with other transcriptional profiling data

- Functional classification with slimGO (gene ontology)
- Comparison with other expression profiling data (20k vs 45k or 45k vs digital northern based on the number of ESTs)
- Confirmation by experiments (e.g., northern or Q-PCR)

Finding Biological meaning



Screening phenotypes of candidate genes with rice insertional mutants

- What is the cut off number of significant genes?
- How much false positive results do the microarray have?

**By multiple insertional lines or homozygous lines for validation of phenotypes
By comparing with the previous results of orthologous genes in other plants**

Materials and Methods for validation

Platform of NSF rice whole genome (45K) oligo array

Whole genome array consist of two slides which have spotted 43,482 oligos designed for 45,116 TIGR V3 rice gene models. Of them, 40,279 unique oligo probes are targeted for single gene and 3,203 shared oligo probes and targeted for 4,837 additional genes. In addition, this array includes 163 oligos targeted for chloroplast and mitochondria and 8 transgenes including *GUS*, *BAR*, and *Hph* which are non rice genes. Especially, 456 hygromycin oligos are randomly spotted as spiking controls. These oligos are good positive or negative control depending on experimental conditions.

Oligo design by PICKY 2

Using Picky 2.0 (Chou et al., 2004) we have designed oligo probes for 45,116 gene models (out of 61,420 total target sequences available in TIGR V3 rice gene set release), using a stringency of less than 17 nt exactly match and 10 degree minimum separation of nontarget and target hybridization temperatures.

Growing condition and RNA isolation

Rice seedlings (cv. Kitaake, Nipponbare, IR24, and TP309) were germinated at dark condition for 3-4 days and then grown at light condition (greenhouse, 30°C) and dark condition (dark incubator, 28°C). After 10 days, leaf grown at both condition were harvested. Total RNA was isolated from light or dark treated leaves using TRIZOL Reagent (Invitrogen) and then purified by Qiagen midi kit. Finally, at least 2ug mRNAs were isolated by Qiagen mRNA isolation kit. All RNAs were quantified and qualified by Nanodrop and Agilent Bioanalyzer.

Microarrays and probe preparation

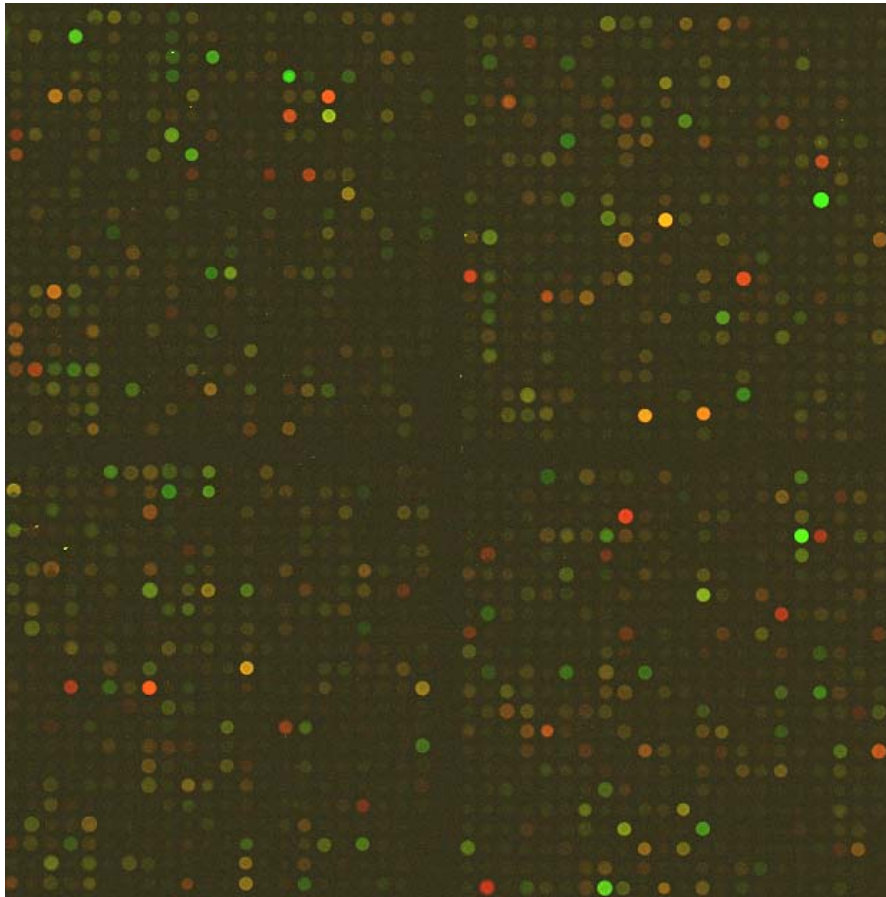
500ng of mRNA from light and dark leaves was labeled with Cy3 and Cy5 using Amersham indirect labeling kit. At least 50 pmoles of labeled cDNA were hybridized onto each slide of NSF 45K rice arrays and washed by Tecan HS4800 Automated Slide Hybridization Station. Microarray data acquisition and analysis. Gene expression data was acquired by measuring the fluorescent intensity of each spot at both 532nm (Cy3) and 635nm (Cy5) with the GenePix Pro 4000B scanner (Axon Instrument). Raw data (GenePix Results Files, GPR) was loaded into R-based LMgene package (Rocke, et al., 2004) and median values were background corrected and then normalized using Lowess normalization. The median intensity value of both dye swaps was averaged and used to determine fold induction or repression.

Large scale of rice T-DNA insertional mutant pool

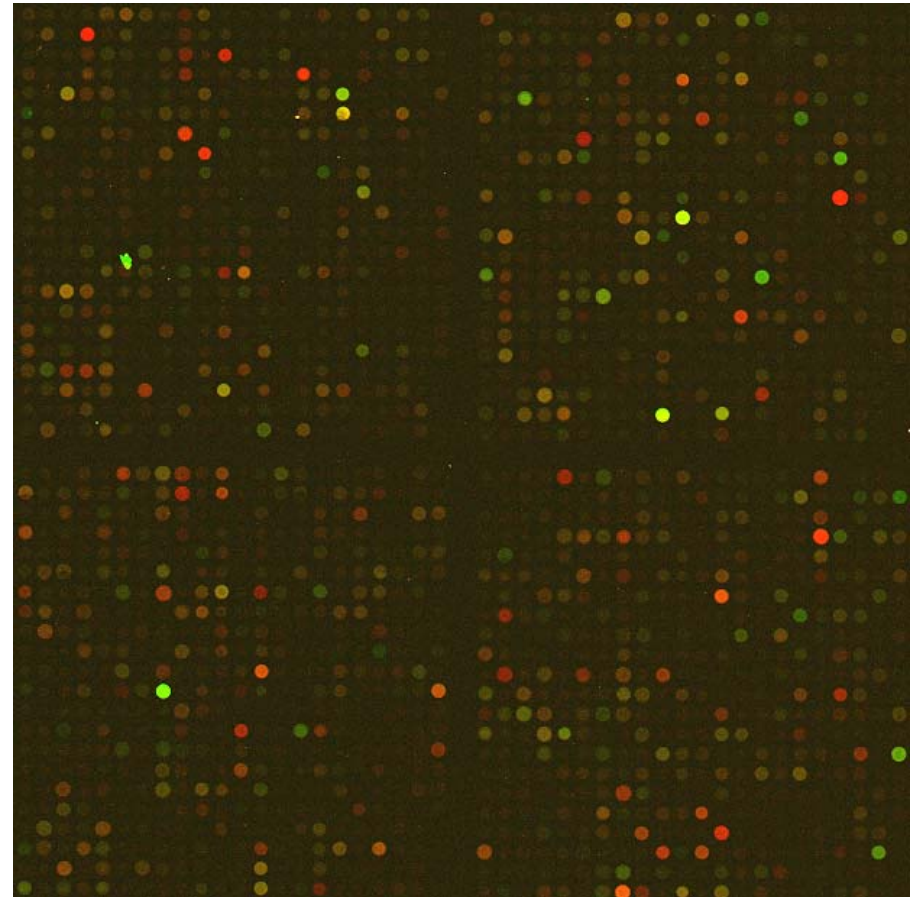
Gyn An's group (POSTECH in Korea) generated huge scale of rice T-DNA insertion mutants. Until now, more than 100,000 mutant are generated and of them franking sequences of at least 50,000 lines have been isolated and all information of these sequences are available in rice genome web browser (<http://signal.salk.edu/ricemsearch.4.html>) (An et al., 2005).

Reproducibility of 45k array through dye swapping

Dark (cy3 green) light (cy5 red)

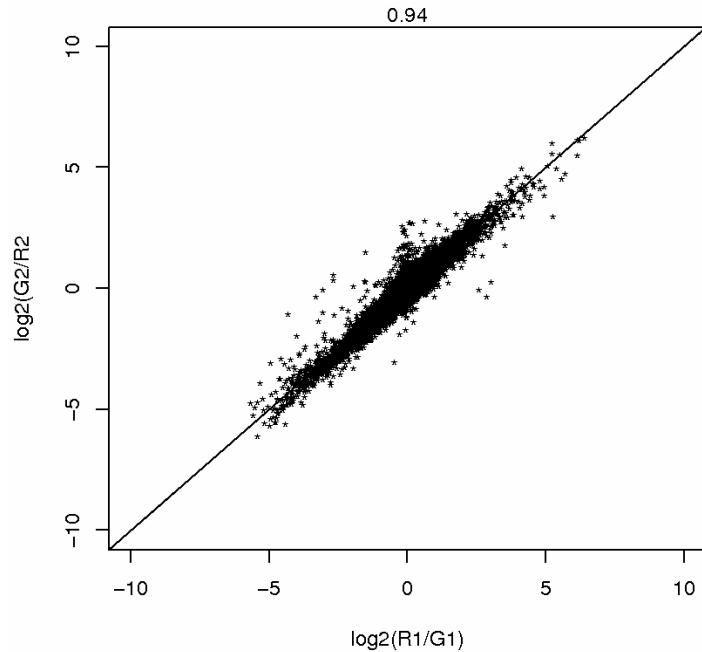


Light (cy3 green) dark (cy5 red)

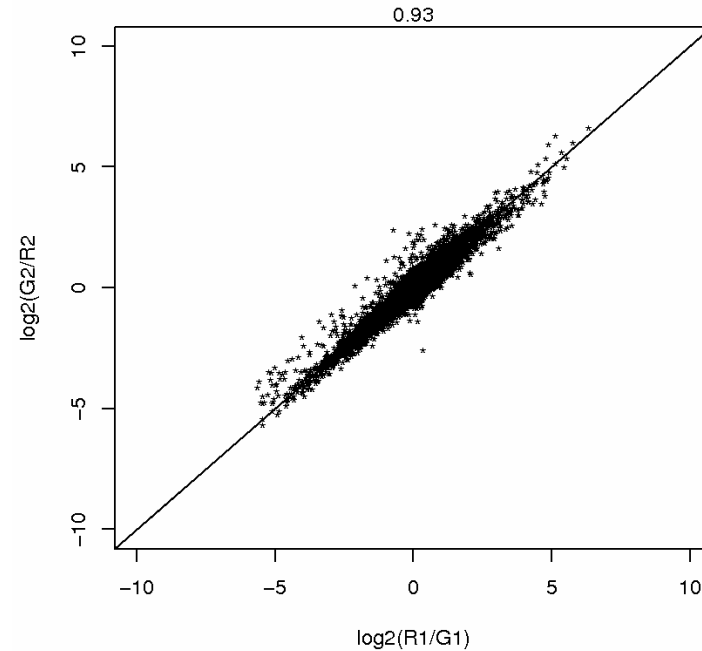


Statistical reproducibility of 45k array

Slide A_sample 1 vs sample 2



Slide B_sample 1 vs sample 2



Correlation coefficients between dye-swapped replicates indicate reproducibility and 45k array showed more than 90% reproducibility by these calculations.

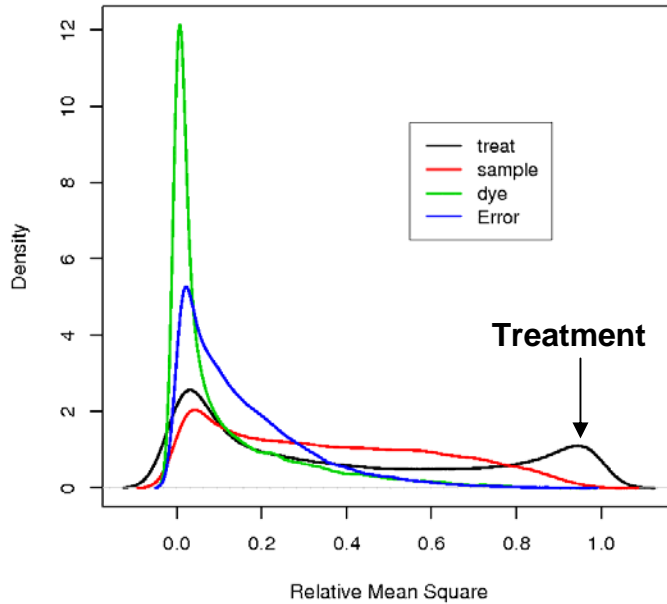
Error rate of 45k array with hph oligos

	total		error >2 fold		error >4 fold		total error rate (%)	
	45K_A	45K_B	45K_A	45K_B	45K_A	45K_B	>2 fold	> 4 fold
Hygromycin	240	216	3	0	0	0	0.66 (3/456)	0
Empty	192	2505	0	0	0	0	0	0

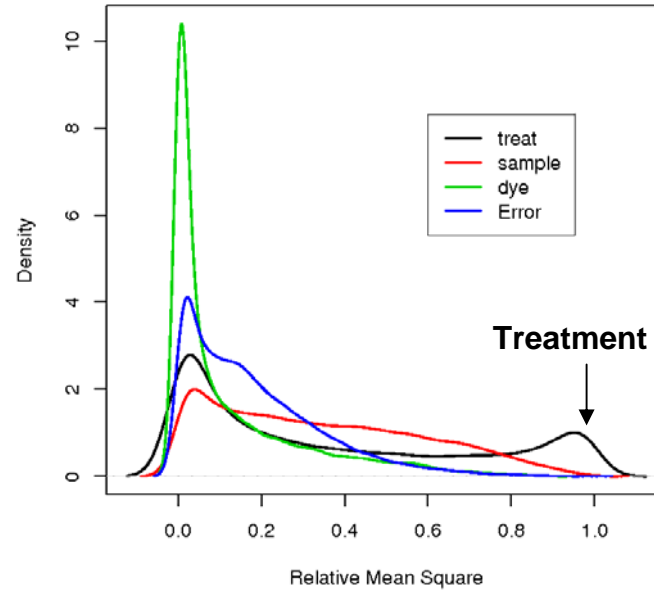
Error rate test using Hph oligos which are randomly distributed showed that 45k array have less than 1% of false positive for showing at least 2 fold changes. These hph oligos are used for negative control to compare wild vs wild or transgenic vs transgenic but in case of comparing wild vs transgenic having hph genes they are positive controls. Empty spots are unspotted controls.

Main Source of variation

45k_slide_A



45k_slide_B



Because main variation between *cys3* and *cys5* resulted from treatment not samples or dye, the data from these experiments are meaningful.

Summary

1. NSF 45k array were printed on two slides due to the limitation of spotting per slide.
2. Generate highly reproducible results between dye-swapped samples.
3. Error rate test with control oligos showed that this array had very low portion (less than 1%) of false positive showing at least two fold.
4. Main source of variation came from the treatment factor but not from dyes or samples.
5. We are validating the functions of candidate genes selected from this microarray by using rice knockout lines.
6. As the cost is \$110 per set, the array will be cost-effective for doing large scale experiments.

Reference

- Rocke DM. 2004. Design and Analysis of Experiments with High Throughput Biological Assay Data. *Sem. Cell Dev. Biol.*, 15, 708-713.
- 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.
- An, G., Jeong, D.H., Jung, K.H. and Lee, S. 2005. Reverse genetic approaches for functional genomics of rice. *Plant Mol Biol*, 59, 111-123.