Application of One Sided t-tests and a Generalized Experiment Wise Error Rate to High-Density Oligonucleotide Microarray Experiments: An Example Using Arabidopsis

W. M. Muir[1], J. Romero-Severson[2], S. D. Rider Jr.[1], A. Simons[1], and J. Ogas[1]
[1]*Purdue University and* [2]*University of Notre Dame*

Abstract: **Motivation**: A formidable challenge in the analysis of microarray data is the identification of those genes that exhibit differential expression. The objectives of this research were to examine the utility of simple ANOVA, one sided t tests, natural log transformation, and a generalized experiment wise error rate methodology for analysis of such experiments. As a test case, we analyzed a Affymetrix GeneChip microarray experiment designed to test for the effect of a CHD3 chromatin remodeling factor, PICKLE, and an inhibitor of the plant hormone gibberellin (GA), on the expression of 8256 *Arabidopsis thaliana genes*.
Results: The GFWER(k) is defined as the probability of rejecting k or more true null hypothesis at a given p level. Computing probabilities by GFWER(k) was shown to be simple to apply and, depending on the value of k, can greatly increase power. A k value as small as 2 or 3 was concluded to be adequate for large or small experiments respectively. A one sided t-test along with GFWER(2)=.05 identified 43 genes as exhibiting PICKLE-dependent expression. Expression of all 43 genes was re-examined by qRT-PCR, of which 36 (83.7%) were confirmed to exhibit PICKLE-dependent expression.

Key words: False discovery rate, microarray, power, *t*-test.

1. Introduction

The advent of inexpensive microarray technology has enabled individual laboratories to easily obtain a global perspective on the expression pattern of thousands of genes. This powerful technology has allowed investigators to diagnose early cancers (Kim, J. W. and Wang, X. W., 2003; Zhang *et al.*, 2003), discover genes that contribute to quantitative traits (Gu *et al.*, 2002), and detect coordinated gene regulation during pivotal developmental events such as embryogenesis and sexual maturation (Girke *et al.*, 2000; Lo *et al.*, 2003; Ruuska *et al.*, 2002).

The first generation microarrays were generally based on two dye methodologies. These cDNA microarray experiments involve hybridizing two mRNA

samples, each of which has been converted into cDNA and labelled with its own fluorophore, on a single glass slide that has been spotted with 10,000-20,000 cDNA probes. In contrast, more recent high-density oligonucleotide microarrays, such as those offered by Affymetrix , provide direct information about the expression levels in an mRNA sample and can have a much higher density (Yang and Speed, 2002).

The majority of methodologies for microarray analysis have been developed for two dye spotted arrays (Kerr *et al.*, 2000; Kerr and Churchill, 2001; Lee *et al.*, 2003, Nguyan *et al.*, 2004, for review see Quackenbush, 2001 and Yang and Speed, 2002). Unfortunately these two-dye spotted arrays also pose other statistical issues, such as normalization to correct for dye bias. Furthermore if more than 2 treatments are used, it is not possible to compare all treatments on the same chip thus necessitating an Incomplete Block Design (IBD) type design (Kerr and Churchill, 2001). As such, special experimental designs, such as the reference and rotational design are needed for correct analysis (Kerr and Churchill, 2001; Quackenbush, 2001 and Yang and Speed, 2002).

In contrast, oligonucleotide microarrays use a single dye technology and pose some advantages, including a greatly increased density of genes and simplified experimental design because treatment effects are tested independently on each chip, eliminating the need for IBD designs. Nevertheless, statistical issues remain, such as normality of residuals, homogeneity of residual variance, correlation of errors within an array, and correlation of biological samples across arrays.

Mixed model methods for analysis of microarray experiment, proposed by Wolfinger *et al.* (2001), solves most of these issues (see Craig *et al.*, 2003 for review). However, the complexity of analysis dramatically increases with these advanced methods. Unfortunately, many of the current practitioners of microarray technology do not possess the mathematical expertise necessary to meaningfully employ these methods. On the other hand ANOVA is a tool that is easy to implement with methods common to most researchers. Kerr and Churchill (2001) conclude that "The analysis of variance (ANOVA) is a natural tool for studying data from experiments with multiple categorical factors".

The first objective of this research was to examine the utility of simple ANOVA for analysis of replicated oligonucleotide microarrays experiments. The motivation was given eloquently by Kerr and Churchill (2001) who stated "An advantage of model based data analysis such as ANOVA is that a model helps the analyst explore the data. If one finds a model inadequate, discovering *why* it is inadequate can help the analyst identify sources of variation and bias." A secondary objective of this study was to show how using a one sided *t*-test can be used to increase power. The final objective was to introduce an alternative method to increase power by accepting a base number of false positive with high

probability.

The ANOVA is particularly suited to analyzing data from microarray experiments that employ a replicated factorial arrangement of treatments. An example of such an experimental design is one in which the investigator looks at gene expression in wild-type and mutant plants in the presence or absence of an added chemical. Many microarray studies incorporate this type of experimental design, e.g. the response of genes in nontumorigenic and tumorigenic tissues to different concentrations of toxic or therapeutic drugs (Lundquist *et al.*, 2002; Martinez *et al.*, 2002) or the response of genes from different tissues to estrogen or other hormones (Abe *et al.*, 2003; Faccioli *et al.*, 2002; Fujita *et al.*, 2003; Goda *et al.*, 2002). This design easily extends into any number of genotypes (or tissues) by any number of developmental time points (or biochemical exposures).

The primary biological objective of this research was to understand how a CHD3-chromatin remodeling factor, *PICKLE*, and a plant growth regulator, gibberellin (GA), regulate gene expression during germination of Arabidopsis seeds (Rider *et al.*, 2003). PICKLE is necessary for repression of embryonic traits in Arabidopsis (Ogas *et al.*, 1997). Expression of the embryonic state in pickle seedlings is inhibited by the plant growth regulator gibberellin (GA) and is enhanced by application of uniconazole-P, an inhibitor of GA biosynthesis (Izumi *et al.*, 1985; Ogas *et al.*, 1997). Specifically, gene expression was examined in wild-type and *pickle* seeds grown in the absence and presence of 10^{-8} M uniconazole-P. Thus the genotypes were 'wild type' vs. the *pickle* mutant, and biochemical exposure was to either 10^{-8} M uniconazole-P or no uniconazole-P during seed germination.

Our working hypothesis was that PICKLE functions during germination to repress genes that promote embryonic identity. In support of such a hypothesis, the transcript levels of two positive regulators of embryogenesis, *LEAFY COTYLEDON1 (LEC1)* and *LEAFY COTYLEDON2 (LEC2)* (Lotan *et al.*, 1998; Stone *et al.*, 2001), are elevated during germination of pickle seedlings (Ogas *et al.*, 1999; Rider *et al.*, 2003). Our interest was to find new genes that exhibited *PICKLE*-dependent expression, i.e. were up regulated. As such, we had a natural one sided test.

2. Biological Methods

Seeds and tissues from the Arabidopsis *pickle-1* mutant (in a Columbia ecotype background) and wild-type Columbia were used for all investigations. Plants were grown as described previously (Ogas *et al.*, 1997; Rider *et al.*, 2003).

The Affymetrix GeneChip Arabidopsis Genome Array[1] contained 8256 sets of

---

[1]part no. 510429, Affymetrix, Santa Clara, CA

oligos representing approximately 30% of the *Arabidopsis thaliana* transcriptome. A 2×2 factorial arrangement of treatments were examined. The first treatment was genotype (*pickle* mutant vs. wild type), the second treatment was uniconazole (applied vs. control), the treatment combinations were designated pkl, Upkl, wt and Uwt (*pickle* mutant untreated, *pickle* mutant treated with uniconazole-P, wild type untreated, and wild type treated with uniconazole-P) were each represented by four biological replicates ($n = 4$) for a total of 16 chips (Rider *et al.*, 2003).

3. StatisticalL Methods

3.1 The ANOVA, partitions, and transformations

The model for the completely randomized design (CRD) associated with the k-th spot (or gene) is $Y_{ij}^k = \mu + \tau_i^k + \epsilon_{(i)j}^k$ where Y_{ij}^k is the expression (or log transform) for the k-th gene, in the j-th replicate of the i-th treatment; μ^k is the overall mean; τ_i^k is the effect of the i-th treatment on that gene, and $\epsilon_{(i)j}^k$ is random residual. For maximum information treatment effects are further partitioned into main effects and interactions. The partitioning should be reduced to single degree of freedom tests by use of orthogonal contrasts. Because the ANOVA must be completed for each spot on the array, methods to automate the test are needed. To accomplish this goal, we use the well-known result that any single degree of F tests can equivalently be constructed as a t-test (Gill, 1978). A simple t-test for any contrast can be computed with the means procedure in SAS or in any standard spreadsheet, such as Excel. The t-test also offers the advantage of being able to test for a one sided alternative. In some experiments, as in this one, the researchers may only be interested in genes that are either up or down regulated, as a result, the power to detect those genes will be greatly increased.

For expression type data, the variance is usually correlated with the mean, violating a critical assumption for the ANOVA. For such data, transforming to logs will usually correct this problem. Interpretation of log transformed data also better meets the interest of the biologist as significant differences are interpreted as being significant ratios on a non-transformed basis, i.e. the difference between logs of numbers is the same as the log of a ratio. A log base 2 is interpreted as fold change, while base 10 is interpreted as orders of magnitude difference. Natural logs have not been widely used for array data but perhaps represents the most valid biological interpretation due to kinetics. A common rate equation in chemistry is where the rate of change in product (∂Y) per unit of time (∂t) is proportional (c) to the product (Y), thus $\partial Y = cY\partial t$. The solution to this differential equation is $Y = ce^t$. Therefore by taking natural logs, the expression is linearized into a rate equation, $\ln(Y) = \ln c + t$. If t is constant across biological replications, then variation in expression is due to linear differences in the rate

constant c, the gene regulatory factor. Differences due to treatments are then interpreted as linear differences in gene regulatory factors (rate constants).

The vast majority of array data will require such a transformation, however, curiously these data better met the assumption when non-transformed. To check this assumption for any data, compute the within gene variance for each gene (the residual error variance in the ANOVA), then plot that against the average expression level for that gene. Any slope significantly different from zero (a zero slope is parallel to the x axis) indicates that the data require a transformation before the analysis proceeds.

For a given gene, because each treatment combination was randomized onto each of 4 biological replicates, the experiment as detailed above is a 2×2 factorial arrangement of treatments in a completely randomized design (CRD). The ANOVA for this design with treatment effects partitioned is given in Table 1.

Table 1: ANOVA table with partitions.

Source of Variation	Degrees of Freedom	Mean Square
Treatments	$t - 1$	MS(T)
Genotypes (C_1)	1	MS(C_1)
Inhibitor (C_2)	1	MS(C_2)
Interaction of	1	MS(C_3)
Genotype x Inhibitor (C_3)		
Within Error	$t(r - 1)$	MS(E)

The mean squares for the partitions can be found using the following formula along with the contrast coefficients given in Table 2

Table 2: Coefficients for partitions of treatment effects.

Treatment	Treatment Combination		Contrast Coefficients (C_{mj})		
	Genotype	Inhibitor	C_{1j}	C_{2j}	C_{3j}
1- pkl	*pickle*	None	1	1	1
2- Upkl	*pickle*	Uniconazole	1	−1	−1
3- wt	wild type	None	−1	1	−1
4- Uwt	wild type	Uniconazole	−1	−1	1

$$MS(C_m) = r(\sum_j C_{mj}\bar{Y}_{ij})^2 I(\sum_j)C_{mj}^2). \qquad (3.1)$$

The F test, which is distributed as F with 1 and $t(r-1)$ degrees of freedom, is then computed as the ratio of $F = MS(C_m)/MS(E)$. This test is equivalently computed as t

$$T_m = \frac{\sum_j C_{mj} \bar{Y}_{ij\cdot}}{\sqrt{\frac{1}{r} MS(E) \sum_j C_{mj}^2}} \tag{3.2}$$

which has $t(r-1)$ degrees of freedom. From these formula it is easy to verify that the calculated value $F = t^2$ and from tables one can verify corresponding critical values, i.e. $F_{1,t(r-1)} = (t_{t(r-1)})^2$. However, when calculated as a t-test the sign of the contrast is preserved, thus allowing a one tailed test. This approach will extend to any contrast for any number of treatments, provided the sum of the coefficients for that contrast is zero. To be orthogonal with other contrasts the sum of the cross products must also sum to zero.

For this analysis, our hypothesis was that one or more genes existed for which the expression level was elevated in pickle mutants, regardless of uniconazole treatment. This hypothesis was based on an expression pattern similar to that of *LEC1* and *LEC2*. Thus the primary contrast of interest was the main effect of genotype (C_1). Because we were only looking for a similar pattern (up regulation), the power to detect up regulated genes increased. Use of prior information to increase power is more cost effective than increasing the number of biological replicates. In other experiments additional contrasts may be of equal or greater importance, this may be particularly true of the interaction of genotypes with uniconazole treatment (C_3), which test the hypothesis that application of uniconazole has a different effect on one genotype than the other.

The critical value of t depends on a number of factors, including one- vs. two-sided alternatives, degrees of freedom (df) for estimation of error variance, and acceptable type I error rates. Choosing an acceptable Type I error rate is discussed in the next section.

3.2 Generalized experiment wise error rate (GFWER(k))

Experimenters have long recognized that if a comparison wise type I error rate (CWER) is used across a great number of tests, a large proportion of declared significant differences would be false. For example analysis of array data involves thousands of comparisons, consequently, if a per comparison error rate of 0.05 were used for our analysis, more than 413 of the 8256 tests would be expected to be declared significant by chance alone. The most widely used approaches to control Type I errors in multiple tests is based on controlling the family wise Type I error rate (FWER) (Fernando *et al.*, 2004). The FWER is the probability of rejecting one or more true null hypotheses, i.e. the probability of accepting

one or more false positives. A common method for controlling the FWER is the Bonferroni or Sidak (1967) adjustments.

However, the FWER with those adjustments is too conservative if the cost of false negatives is high relative to the cost of false positives, i.e. they sacrifice power to avoid accepting false positives. Methods have been developed to address this issue by allowing for some false positives among those declared significant, such as the false discovery rate (FDR, Benjamini and Hochberg, 1995; Reiner. *et al.*, 2003; see Nguyen, 2005 for general discussion on this issue). Alternatives to the FDR have since been proposed that take into account the expected number of false null hypothesis and other modifications (see Fernando *et al.*, 2004 for review). However, all methods used to estimate an FDR make assumptions about the distribution of truly expressed genes. As a result, the FDR will either be too liberal or conservative.

Here we present an alternative that does not attempt to establish an FDR. Rather the method is an extension of the FWER methodology to allow for a higher family wise error rate. The development is as follows: Assume a strictly null distribution from which N independent test statistics are computed, from which N independent decisions are made at the same critical threshold level. The probability that any one decision is incorrect is p. An incorrect decision is defined as rejecting a true null hypothesis. With multiple tests, the probability of exactly m incorrect and $N - m$ correct decisions is

$$P(m = \text{ incorrect } |N = \text{ decisions }) = \binom{N}{m} p^m (1 - p)^{N-m} \qquad (3.3)$$

The usual FWER $= \xi(1)$ is the probability of rejecting 1 or more true null hypotheses found as:

$$\xi(1) = \sum_{m=1}^{N} p^m (1 - p)^{N-m} \qquad (3.4)$$

or equivalently, 1 minus the probability of no incorrect decisions,

$$\xi(1) = 1 - (1 - p)^N \qquad (3.5)$$

which is Sidak's (1967) equation. The value of p per comparison (CWER) is found such that the $\xi(1)$ is achieved, i.e.

$$p = 1 - e^{[\ln\{1 - \xi(1)\}/N]} \qquad (3.6)$$

Stated in the reverse, there is a 1-FWER probability of *no* incorrect decisions among the N decisions made, i.e.

$$\omega(k) = \sum_{m=1}^{k-1} \binom{N}{m} p^m (1-p)^{N-m} \tag{3.7}$$

A generalization of this procedure is to divide the total probability of making Type I errors into parts associated with how many errors are likely to be made at a given probability. Among the N decisions made, define $\xi(k)$ as the probability of rejecting k or more true null hypotheses and $\omega(k)$ as the probability of rejecting fewer than k true null hypotheses, $\xi(k) + \omega(k) = 1$, where

$$\xi(k) = \sum_{m=k}^{N} p^m (1-p)^{N-m} \tag{3.8}$$

$$\omega(k) = \sum_{m=0}^{k-1} p^m (1-p)^{N-m} \tag{3.9}$$

If for a given k, the value for $\xi(k)$ is set to a small value, then among those tests declared significant, one accepts that there will be a high probability of $k-1$ false positives plus a low probability of k or more false positives. Therefore, a new type of error rate is defined as $\text{GFWER}(k)$, which is strictly the probability of making k or more incorrect decisions at a given level of p, and ignores the probability of less than k Type I errors. The latter type of errors are considered acceptable in order to gain power and decrease the Type II error rate. For a more general development of the generalized family wise error rate see van der Laan (2004). The $\text{GFWER}(k)$ cannot be solved for directly, but solutions can be found by iteration. SAS source code used to compute adjusted p values for any $\xi(k)$ and N is given at our web site. However, Equations (3.8) and (3.9) can also be approximated by the normal as follows: If X is binomial with n trials and probability of success p, then

$$P[X > r] \approx \Phi\left(\frac{r - np}{\sqrt{np(1-p)}}\right),$$

where Φ is the cumulative distribution of standard normal distribution.

Tables 3 and 4 give p values for, respectively, a one- and two-tailed alternative, and $\xi(k) = .05$. Associated critical values of t are given in Tables 5 and 6 for experiments with 6 and 60 df for estimating error variance. Note that for all k values, the critical value of t for a one-sided test is between 12 and 13% smaller, with corresponding increases in power. Table 3 shows that by allowing for 2 or more false positives in a 1 sided t-test increases the adjusted p value by 6.5 times, and thereby also increasing the power of the test. Results presented in Tables 3

show that for the range of N examined (i.e. $N > 1000$) the ratios of p values for GFWER(k) to that of GFWER(1) are independent of N. Thus, for such chips, once the Sidak p values are found, GFWER(k) can be found by multiplication using the constants given in the table.

Table 3: Adjusted p values for $k = 1$ to 5, chips of size 1,000 and 50,000 and a one-tailed GFWER(k)=5%.

k	Number of Tests			
	1,000		50,000	
	p-value$\times 10^6$	Ratio*	p-value $\times 10^6$	Ratio*
1	51.29		1.025	
2	335.02	6.5	6.70	6.5
3	783.41	15.3	15.66	15.3
4	1320.01	25.7	26.38	25.7
5	1913.31	37.3	38.23	37.3

* Ratio of p-values to that of GFWER(1)

Table 4: Adjusted p-values for $k = 1$ to 5, chips of size 1,000 and 50,000 and a two-tailed GFWER(k)=5%.

k	Number of Tests	
	1,000	50,000
	p-value$\times 10^6$	p-value $\times 10^6$
1	25.63	.521
2	167.5	3.35
3	391.7	7.83
4	660.0	13.19
5	956.65	19.15

An important issue is what value of k should one use. The value of k should be set as small as possible without sacrificing too much power. For an experiment of a given size, the rate at which power increases is dependent on the critical value of t. Examination of Tables 5 and 6 shows that the greatest decrease in the critical value of t, with either large or small experiment, comes from increasing k from 1 to 2. For large experiments increasing k beyond 2, or for small increasing k beyond 3, brings about much smaller incremental decreases in t. From these results, some general guidelines can be deduced for choice of k. Regardless of the number of spots on a chip, a k value of 2 or 3 should be adequate for large and small experiments respectively.

Table 5: The six genes for which the qRT-PCR assay detected no expression in untreated wild type seed. Transcripts were detected in untreated *pickle* seeds. Transcripts were also detected for both wild type and *pickle* seeds (Uwt and Upkl) germinated in the presence of uniconazole-*p*, thus permitting calculation of Upickle fold change relative to Uwt. The mean values from the arrays are included for illustration. #Pr is the number of times Affymetrix Microarray Suite software (v. 5.0) labeled a gene 'present' for the 16 gene chips used for this investigation.

AGI Code		Mean values (4 chips)				qRT-PCR fold change				Putative ID/function
	#Pr	wt	pickle	Uwt	Upickle	wt	pickle	Uwt	Upickle	
At3g16410	16	4535	14225	3630	18074	−	+	1	204.3	Jacalin type lectin
At4g27140	15	1000	2365	417	3214	−	+	1	8.95	2S1 seed storage protein
At1g67330	4	170	701	97	1255	−	+	1	3.75	uncharacterized
At5g13930	16	18459	37623	18471	46304	−	+	1	1.75	TT4/chalcone synthase
At3g23220	16	1611	2606	1364	2560	−	+	1	1.43	ERF1/transcription factor
At1g09750	16	2005	3200	2178	5193	−	+	1	0.58	nucleoid-like protein

Table 6: Presence of uniconazole-*p* increases derepression of *PICKLE*-dependent genes in *pickle* seedlings.

	qPCR Ratios			Putative Function
AGI Code	pkl/Wt	Upk/Wt	Upk/pkl	
At5g01600	2.90	8.82	3.04	maturation
At3g16420	4.75	11.41	2.40	defense
At1g73190	1.79	2.96	1.65	maturation
At2g28790	1.77	2.90	1.64	
At1g20620	2.33	3.51	1.50	maturation
At5g54740	2.97	4.10	1.43	
At4g19810	2.43	3.33	1.37	
At3g52500	5.55	7.55	1.36	
At3g16430	1.55	2.08	1.34	defense
At1g05510	1.88	2.53	1.34	
At3g16460	4.64	5.22	1.13	defense
At4g08685	2.61	2.78	1.06	
At2g35810	4.24	4.24	1.00	
At2g19590	2.50	2.24	0.90	
At4g37410	2.38	1.95	0.82	
At5g12030	5.36	2.68	0.50	desiccation

4. Biological Verification: QRT-PCR Analysis

Those genes found significant with ANOVA were re-analyzed using qRT-PCR to compare results. The qRT-PCR method, while more precise than the chip analysis, is still subject to error. The method is based on PCR amplification of mRNA in the sample until a pre-determined threshold is obtained. Because the amplification is a doubling with each cycle, the accuracy of the method is questionable if there exists less than a 2 fold difference in mRNA between the two treatments. qRT-PCR is also subject to biological variability between samples and should therefore also be replicated and treated to statistical analysis. However, replicated qRT-PCR analysis for each gene would be extremely expensive and time consuming. Therefore within the limitations of this experiment, and recognizing those limitations, we defined confirmation of *PICKLE*-dependent expression as a two-fold or greater increase in expression level of a given gene in *pickle* versus wild-type seed when grown in either the absence or presence of uniconazole-P. qRT-PCR was used to compare transcript levels in *pickle* versus wild-type seed grown in the absence of uniconazole-P as well as transcript levels in *pickle* versus wild-type seed grown in the presence of uniconazole-P.

Quantitative RT-PCR was performed on an ABI sequence detection system using RNA from one of the biological replicates previously generated (Rider *et al.*, 2003). Oligonucleotide primer sequences and primer concentrations used are listed in supplementary Table 2S available at the web site.

5. Results and Discussion

5.1 Statistical issues

For this experiment, we used $\xi(2) = .05$. Allowing for one false positive raised the adjusted p value from 6.21×10^{-6} to 4.1×10^{-5} and correspondingly increased the power of the test The ANOVA method selected 43 genes, less than one of which was expected to be a false positive based on the experimentwise selection criteria that we employed ($8256 \times 4.1 \times 10^{-5} = .33$). Our qRT-PCR analysis supported 36 of the 43 genes (Figure 1). A surprising result of this study was that qRT-PCR did not detect transcripts in wild-type seeds for 6 of the 43 genes identified as having expression differences based on analysis of the array data (Table 5). Although this observation is consistent with the hypothesis that PICKLE represses expression of these genes in wild-type seeds to facilitate the developmental transition from embryo to seedling, the array expression values did not suggest absence of transcripts in wild-type seeds.
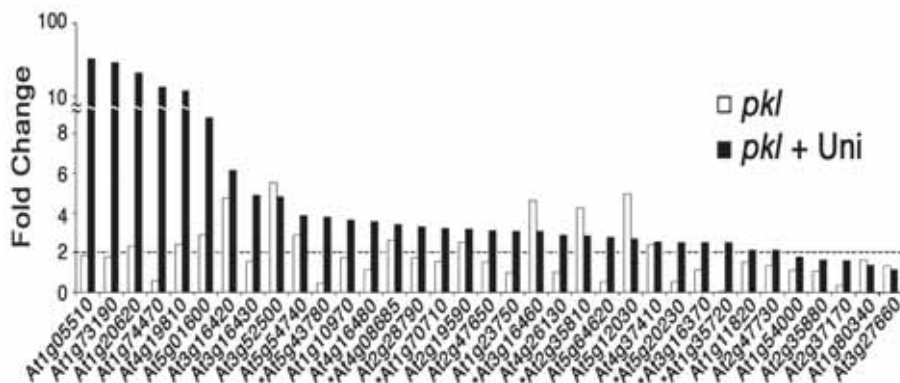
Figure 1. qRT-PCR results for genes selected. The Arabidopsis Genome Initiative (AGI) code for each gene is provided along the x-axis. The y-axis represents the fold change in pickle normalized to wild type expression levels. The white bars denote fold change in pickle normalized to wild-type expression levels imbibed in the absence in uniconazole-P (pickle/wt) whereas the black bars denote fold change in pickle normalized to wild-type expression levels imbibed in the presence of 10-8 M uniconazole-P (Upickle/Uwt). Only those genes whose expression was detected in all four cases are shown here. Expression analysis of those genes for which expression was not detected in untreated wild-type seeds can be found in Table 5. Genes that were also identified by the MFC method (Rider *et al.*, 2003) are indicated by an asterisk

There are at least two possible explanations for the elevated number of observed false positives. Affymetrix constructed this GeneChip when the sequence of the Arabidopsis genome was only partially completed. Inflated expression values for some oligos may have arisen from cross hybridization to unintended targets. In fact, two of the false positives were false because qRT-PCR detected no expression in germinating seeds under any condition. Alternatively, as previously discussed, the discrepancy may be due to different criteria used to determine success for each method. The qRT-PCR data should only be viewed as supporting evidence, not confirmatory.

5.2 Biological Inferences

PICKLE is necessary to repress expression of embryonic traits in Arabidopsis seedlings. Previous analysis of genes that exhibit *PICKLE*-dependent repression identified genes associated with various stages of seed development. ANOVA identified genes associated with seed development, including 2S albumin genes, HSP17.6, and several lectin-like genes (Guerche *et al.*, 1990; Lenman *et al.*, 1993; Ruuska *et al.*, 2002; Sun *et al.*, 2001). In all, 10 of the genes (28%) identified and confirmed by qRT-PCR analysis were associated with embryo development

or exhibit sequence similarity to genes involved in embryo development (Table 1S, available at the web site). Additional studies will be necessary to determine if the other 26 genes that showed *PICKLE*-dependent expression in the germinating seed are also involved in some aspect of embryo development. Previous expression analysis did not suggest a specific role for uniconazole-P in increasing penetrance of the pickle root phenotype in *pickle* seedlings (Rider *et al.*, 2003). This analysis revealed that the extent of derepression of many of the genes that exhibit *PICKLE*-dependent repression is enhanced by the presence of uniconazole-P (Figure 1, black bars versus white bars). The magnitude of this enhancement, however, was often due in large part to the fact that the presence of uniconazole-P resulted in decreased expression of the gene relative to wild-type seed imbibed in the absence of uniconazole-P (data not shown).

In order to examine the effect of combining the *pickle* mutation with exposure to uniconazole-P, we compared the fold change values of genes in *pickle* versus wt seedlings (pkl/wt) and the fold change values of genes in *pickle* treated with uniconazole-P versus wt seedlings (Upkl/wt) as determined by qRT-PCR (Table 6). In order to make this analysis comparable to previous analysis of the dataset, either the ratio pkl/wt or the ratio Upkl/wt or both had to be ≥ 2 for a gene to be included in this analysis. Genes for which a transcript was not detected in wild-type seedlings were excluded from this analysis. Sixteen genes identified with ANOVA met these expression criteria. We found that the presence of uniconazole-P did increase expression of many of these genes in *pickle* seedlings; the transcript level of 10 genes increased 33% or more when *pickle* seeds were imbibed in the presence of uniconazole-P. In contrast, a previous analysis of the same array data identified no genes for which the corresponding transcript was increased by treatment of *pickle* seeds with uniconazole-P (Rider *et al.*, 2003).

Uniconazole-P increases the probability that primary roots of the *pickle* mutant will express embryonic differentiation traits (Ogas *et al.*, 1997). Genes associated with seed development exhibit elevated expression in *pickle* seedlings, suggesting that the expression of these genes contributes to the ability of *pickle* seedling to express embryonic traits after germination (Rider *et al.*, 2003). The discovery that the presence of uniconazole-P enhances the expression of 10 genes in *pickle* seedlings, 5 of which (50%) are involved in seed development or exhibit sequence similarity to genes involved in seed development, suggests for the first time that the increased penetrance of embryonic traits in *pickle* seedlings treated with uniconazole-P may be mediated in part through changes in gene expression. Specifically, our results are consistent with the hypothesis that GA acts in concert with PICKLE during germination to repress expression of genes that promote embryonic traits. Further characterization of the genes identified here may facilitate subsequent genetic and biochemical analysis of the GA signal

transduction pathway that mediates this response.

5.3 Utility

We have shown that a simple ANOVA method can identify a manageable number of candidate genes for differential expression from a gene expression array, most of which were real. Although we only applied the approach to an experiment that incorporated a simple class-by-treatment design, it is applicable to any full factorial design and is computationally straightforward. Previous analysis of the array data employed a modified fold change (MFC) approach (Rider *et al.*, 2003) and failed to detect many of the genes identified by ANOVA. In addition, our current analysis demonstrates that treatment of *pickle* seedlings with uniconazole-P enhances the derepression of *PICKLE*-dependent genes during germination. These results reinforce the power of ANOVA versus a method that emphasizes fold-change.

The practical utility of the GFWER(k) method is derived from allowing the user to influence the number of genes identified by selecting the appropriate value for k, the number of false positives allowed above the threshold significance level. A critical question is what value of k will result in the greatest increase in power with the lowest number of Type I errors. A simple power analysis showed that regardless of the number of spots on a chip, a k value of 2 or 3 should be adequate for large and small experiments respectively. Although the GFWER(k) and FDR are closely related and greatly increase the power of the experiment by relaxing the Type 1 error rate, the application of the GFWER(k) does not attempt to project an FDR, rather, we only set the maximum number of false positives under the null hypothesis. Calculations for an exact FDR would require knowledge of 1) the number of truly expressed genes, 2) the signal to noise ratio, and 3) their distribution. Without knowing these factors, the FDR as calculated by any of the current methods is an approximation. As a result, the GFWER(k) may be more or less conservative than FDR methods, depending on the particular experiment. However, the GFWER(k) is constant and independent of the experiment, which in itself is appealing. This gives rise to another interesting difference between the methods. The expected number of false positives can be determined *a priori* with the GFWER(k) because the rate is independent of the data, whereas with the FDR (and newer methods as reviewed by Fernando *et al.*, 2004) calculations are dependent on the data and one has to wait until the list is generated to determine what the expected number of false positives will be. This difference could be critical in the planning stage of an experiment.

Acknowledgements

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **15**, 63-78.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate — A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289-300.

Black M. A., and Doerge R. W. (2002) Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiments. *Bioinformatics* **18**, 1609-1616.

Craig B. A., Black M. A., and Doerge R. W. (2003). Gene expression data: The technology and statistical analysis. *J. Agricultural Biological and Environmental Statistics* **8**, 1-28.

Faccioli, P., Lagonigro, M., Cecco, L., de, Stanca, A., Alberici, R., Terzi, V., and de Cecco, L. (2002). Analysis of differential expression of barley ESTs during cold acclimatization using microarray technology. *Plant Biol.* **4**, 630-639.

Fernando, R. L., Nettleton, D., Southey, B. R., Dekkers, J. C. M., Rothschild, M. F., and Soller M. (2004). Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611-619.

Fujita, N., Jaye, D. L., Kajita, M., Geigerman, C., Moreno, C. S., and Wade, P. A. (2003). MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell* **113**, 207-219.

Gill, J. L. (1978). *Design and Analysis of Experiments in the Animal and Medial Sciences. Volume 1.* The Iowa State University Press.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C., and Ohlrogge, J. (2000). Microarray analysis of developing Arabidopsis seeds. *Plant Physiol* **124**, 1570-1581.

Goda, H., Shimada, Y., Asami, T., Fujioka, S., and Yoshida, S. (2002). Microarray analysis of brassinosteroid-regulated genes in arabidopsis. *Plant Physiol* **130**, 1319-1334.

Gu, C. C., Rao, D. C., Stormo, G., Hicks, C., T., and Province, M. A. (2002). Role of gene expression microarray analysis in finding complex disease. *Genes. Genet Epidemiol* **23**, 37-56.

Guerche, P., Tire, C., De Sa, F. G., De Clercq, A., Van Montagu, M., and Krebbers, E. (1990). Differential expression of the arabidopsis 2S albumin genes and the effect of increasing gene family Size. *Plant Cell* **2**, 469-478.

Izumi, K., Kamiya, Y., Sakurai, A., Oshio, H., and Takahashi, N. (1985). Studies of sites of action of a new plant growth retardant (E)-1-(4-chlorophenyl)-4,4-dimethyl-2-(1,2,4-triazol-1-yl)-1-penten-3-ol (S-3307) and comparative effects of its stereoisomers in a cell-free system from Cucurbita maxima. *Plant Cell Physiol* **26**, 821-827.

Kerr M. K., andChurchill G. A. (2001). Statistical design and the analysis of gene expression microarray data. *Genetical Research* **77**, 123-128.

Kerr M. K., Martin M., and Churchill G. A. (2000). Analysis of variance for gene expression microarray data. *J. Computational Biology* **7**, 819-837.

Kim, J. W., and Wang, X. W. (2003). Gene expression profiling of preneoplastic liver disease and liver cancer: A new era for improved early detection and treatment of these deadly diseases? *Carcinogenesis* **24**, 363-369.

Lee, M. T., Whitmore, G. A., Yukhananov, R. Y. (2003). Analysis of unbalanced microarray data. *Journal of Data Science* **1**, 103-121.

Lenman, M., Falk, A., Rodin, J., Hoglund, A. S., Ek, B., and Rask, L. (1993). Differential expression of myrosinase gene families. *Plant Physiol* **103**, 703-711.

Lo, J., Lee, S., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W., Wang, W., Wen, Z., and Peng, J. (2003). 15,000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. *Genome Research* **13**, 455-466.

Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., Yamagishi, K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195-1205.

Lundquist, H., Oredsson, S., and Akesson, B. (2002). Effect of quercetin on gene expression in human cells as measured by microarrays — a pilot study. Paper presented at: Health promoting compounds in vegetables and fruit. Proceedings of a workshop in Karrebaeksminde, Denmark, 6 8 November, 2002. DIAS Report, Horticulture. 2002, No.29, 77 80; 10 ref. (Danmarks JordbrugsForskning; Tjele; Denmark).

Martinez, J. M., Afshari, C. A., Bushel, P. R., Masuda, A., Takahashi, T., and Walker, N. J. (2002). Differential toxicogenomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin in malignant and nonmalignant human airway epithelial cells. *Toxicol Sci.* **69**, 409-423.

Nguyen, D. V., Wang, N. and Carroll, R. J. (2004). Evaluation of missing value estimation for microarray data. *Journal of Data Science* **2**, 347-370.

Nguyen, D. V. (2005). A unified computational framework to compare direct and sequential false discovery rate algorithms for exploratory DNA microarray studies. *J. Data Science* (In Press).

Nitin, J., Thatte, j. Braciale, T., Ley, K., O'Connell, M. and Lee. J. K. (2003). Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* **19**, 1945-1951.

Ogas, J., Cheng, J. C., Sung, Z. R., and Somerville, C. (1997). Cellular differentiation regulated by gibberellin in the arabidopsis thaliana pickle mutant. *Science* **277**, 91-94.

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in arabidopsis. *Proc. Natl. Acad. Sci. USA* **96**, 13839-13844.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nature Reviews Genetics* **2**, 418-429.

Reiner, A., Yekutieli, D., and Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375.

Rider, S. D., Henderson, J. T., Jerome, R. E., Edenberg, H. J., Romero-Severson, J., and Ogas, J. (2003). Coordinate repression of regulators of embryonic identity by PICKLE during germination in arabidopsis. *Plant J.* **35**, 33-43.

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* **14**, 1191-1206.

Sidak, Z. (1967). Rectangular confidence regions for means of multivariate normal distributions. *J. Amer. Statis. Asso.* **62**, 626-636.

Sokal, R. R., and Rohlf, F. J. (1995). *Biometry, Third edition*. W. H. Freeman and Company.

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. USA* **98**, 11806-11811.

Storey, J. D. and Tibshiran, R. (2003a). Statistical significance for genome wide studies. *Proc. Natl. Acad. Sci, USA* **100**, 9440-9445.

Storey, J. D. and R. Tibshirani, R. (2003b). SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In *The Analysis of Gene Expression Data: Methods and Software* (Edited by G. Parmigiani, E. S. Garrett, R. A. Irizarry, S.L . Zeger). Springer.

Sun, W., Bernard, C., van de Cotte, B., Van Montagu, M., and Verbruggen, N. (2001). At-HSP17.6A, encoding a small heat-shock protein in Arabidopsis, can enhance osmotolerance upon overexpression. *Plant J.* **27**, 407-415.

van der Laan, M. J., Dudoit, S. and Pollard, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*[2]

Wolfinger R. D., Gibson G., Wolfinger E. D., Bennett L., Hamadeh H., Bushel P., Afshari C., Paules, R. S. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *J. Computational Biology* **8**, 625-637.

Yang, Y. H. and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nature Reviews Genetics* **3**, 579-588.

Zhang, H., Yu, C. Y., and Singer, B. (2003). Cell and tumor classification using gene expression data: Construction of forests. *Proc. Natl. Acad. Sci. USA* **100**, 4168-4172.

W. M. Muir
Department of Animal Sciences
915 W. State Street
Purdue University
W. Lafayette IN 47907-2054, USA

J. Romero-Severson
University of Notre Dame
327 Galvin Life Sciences
Notre Dame, IN 46556-5645. USA

S.D. Rider Jr.
Department of Biochemistry
175 S. University Ave.
Purdue University
West Lafayette, IN 47907-2063. USA

---

[2]See http://www.bepress.com/sagmb/vol3/iss1/art14/

A. Simons
Department of Biochemistry
175 S. University Ave.
Purdue University
West Lafayette, IN 47907-2063. USA

J. Ogas
Department of Biochemistry
175 S. University Ave.
Purdue University
West Lafayette, IN 47907-2063. USA