Research article

Assessment of differential gene expression in human peripheral nerve injury

Yuanyuan Xiao†[1], Mark R Segal†[2], Douglas Rabert[3], Andrew H Ahn[4], Praveen Anand[5], Lakshmi Sangameswaran[6], Donglei Hu[1] and C Anthony Hunt*[1]

Address: [1]Department of Biopharmaceutical Sciences, University of California, San Francisco, San Francisco, CA 94143, USA, [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94143, USA, [3]Neurobiology Unit, Roche Bioscience, Palo Alto, CA 94304, USA, [4]Department of Anatomy and Neurology, University of California, San Francisco, San Francisco, CA 94143, USA, [5]Peripheral Neuropathy Unit, Division of Neuroscience and Psychological Medicine, Imperial College of Science, Technology and Medicine, Hammersmith Hospital, London W12 ONN, UK and [6]Pherin Pharmaceuticals, Mountain View, CA 94043

E-mail: Yuanyuan Xiao - yxiao@itsa.ucsf.edu; Mark R Segal - mark@biostat.ucsf.edu; Douglas Rabert - douglas.rabert@roche.com; Andrew H Ahn - aahn@itsa.ucsf.edu; Praveen Anand - p.anand@ic.ac.uk; Lakshmi Sangameswaran - lakshmi@pherin.com; Donglei Hu - donghu@itsa.ucsf.edu; C Anthony Hunt* - hunt@itsa.ucsf.edu

*Corresponding author   †Equal contributors

Abstract

Background: Microarray technology is a powerful methodology for identifying differentially expressed genes. However, when thousands of genes in a microarray data set are evaluated simultaneously by fold changes and significance tests, the probability of detecting false positives rises sharply. In this first microarray study of brachial plexus injury, we applied and compared the performance of two recently proposed algorithms for tackling this multiple testing problem, Significance Analysis of Microarrays (SAM) and Westfall and Young step down adjusted *p* values, as well as *t*-statistics and Welch statistics, in specifying differential gene expression under different biological states.

Results: Using SAM based on *t* statistics, we identified 73 significant genes, which fall into different functional categories, such as cytokines / neurotrophin, myelin function and signal transduction. Interestingly, all but one gene were down-regulated in the patients. Using Welch statistics in conjunction with SAM, we identified an additional set of up-regulated genes, several of which are engaged in transcription and translation regulation. In contrast, the Westfall and Young algorithm identified only one gene using a conventional significance level of 0.05.

Conclusion: In coping with multiple testing problems, Family-wise type I error rate (FWER) and false discovery rate (FDR) are different expressions of Type I error rates. The Westfall and Young algorithm controls FWER. In the context of this microarray study, it is, seemingly, too conservative. In contrast, SAM, by controlling FDR, provides a promising alternative. In this instance, genes selected by SAM were shown to be biologically meaningful.

Background

Injuries to nerve or tissue cause an immediate sensation of pain, or acute pain, which resolves with the resolution of the injury. Less commonly, these injuries produce a heterogeneous group of pathological states that produce chronic pain that is unresponsive to the analgesics used to treat acute pain. It would be therefore of great interest to map the characteristic features of gene expression following nerve injuries, as they may reveal mechanisms of chronic pain and genetic targets for the design of novel therapies. In this study we apply the powerful microarray technology to study the injuries of the brachial plexus, a typical form of nerve injury from motorcycle or workplace incidences [1]. Since this is largely a paper focusing on methodology, we refer the readers to our website [2] for details of the brachial plexus injuries.

The goal of this study, detection of condition or state-dependent differentially expressed genes, is an important problem in the context of microarrays. Fold changes and *t* tests have been applied to microarray datasets to yield lists of genes that have undergone altered expression under a certain significance threshold. Although these methods are useful in identifying potential gene targets, they are best applied to small sets of hypothesis tests. In the context of microarray experiments where thousands of genes are commonly evaluated simultaneously, the probability of detecting false positives rises sharply. Two recently proposed algorithms seeking to address this multiple testing problem, Significance Analysis of Microarray (SAM) and Westfall and Young step-down adjusted *p* values, will be investigated in our study. We note at the outset that our investigation of methods for detection of such differentially expressed genes has several limitations. Firstly, it is based solely on the brachial plexus injury data which, in turn, is limited by small sample sizes (10 injured patients and 3 controls). However, such small sample sizes are characteristic of many current microarray studies, in view of technologic expense, so that exploration in this setting is purposeful. Secondly, our investigation is largely empiric, there being no overarching theory by which to arbitrate competing approaches. Nonetheless, it is the frequently observed heteroskedasticity of gene expression measurements that motivates our extension of the SAM methodology (see below) to Welch statistics.

SAM was designed to evaluate the significance of changes in gene expression between different biological states [3]. SAM assigns a score to each gene (here and below what we refer to as genes are actually probe sets on the chip that target full length cDNAs or ESTs) as an index of the relative difference between groups. A gene is labeled as significant if its score surpasses a threshold. SAM terms the percentage of genes identified by chance the false discovery rate (FDR) and estimates this quantity by recourse to permutation. The threshold can be adjusted to identify different sizes of sets of putatively significant genes, and FDRs are modified accordingly. We note here that this differs from the formulation and sequential control of FDRs as originally proposed by Benjamini and Hochberg [4]; see Storey [5] and Storey and Tibshirani [6] for discussion of the distinctions and considerations pertinent to microarray data analysis. Throughout, we will use FDR in the sense of SAM. Tusher, *et al.* used SAM to study the transcriptional response of lymphoblastoid cells to ionizing radiation, and they were successful in identifying 34 biologically meaningful genes. A different algorithm, proposed by Westfall and Young and adapted by Dudoit *et al.*[7] for microarrays, tackles the multiple testing problem from a different angle. It controls the family-wise Type I error rate (FWER), which is the probability of having at least one false positive among all hypotheses tested. It provides less conservative control than the traditional Bonferroni, Šidák, and Holm corrections and is able to handle the dependence (e.g. co-regulation) among gene expression profiles. In a study that tested whether alterations in the expression of genes related to high-density lipoprotein metabolism affects the expression of other genes, multiple-testing methods identified 5 genes that showed significant altered expression levels in scavenger receptor BI transgenic mice and 4 genes in apolipoprotein AI-knockout mice [8]. Further discussions, additional approaches and comparisons can be found at [9].

In the comparison of expression levels of more than 5000 probe sets between normal subjects and brachial plexus injury patients, we adopt both SAM and Westfall & Young algorithms using regular *t* statistics assuming equal variances and *t* statistics assuming unequal variances (hereafter called Welch statistics). The application of Welch statistics is motivated by the work of Thomas, *et al.*[10]. In their paper, they applied Welch tests on a leukemia dataset [11] and demonstrated the importance of allowing for unequal variances. Following a similar approach, we have identified a set of known and novel genes that largely fall into six categories: cytokines / neurotrophine, myelin function, signal transduction, cytoskeleton, transcription / translation and others. Some genes are confirmed by literature; others still require explanation and future research. We also compared the performance of SAM and Westfall & Young algorithms, as well as *t* statistics and Welch statistics. In this study, we hope to provide a framework for future experimental research in avulsion injuries and pain. In addition, we illustrated the use of appropriate multiple testing methods for microarrays in monitoring differential gene expression between different biological states.

Results

Data displays for test statistics

Figure 1 is a scatter plot of the intensity readings for a patient (P15) and the corresponding readings of a control (C1). The considerable deviation from the identity line indicates that gene expression is substantially disturbed as a consequence of the avulsion injury. For comparison, the scatter plots for a pair of control and a pair of patient data sets clearly show less deviation.

The frequency distribution of *t* statistics and the normal Q-Q plot for *t*-statistics are provided in Figure 2. Both the histogram and the Q-Q plot show that the distribution of *t*-statistics is not normal. However, in our study we are more interested in identifying genes with substantial *t*-statistics than in testing the distribution of *t*-statistics. The genes having *t*-statistics that deviate markedly from the bulk of the observations were labeled as potentially having biological significance and may deserve further investigation [8]. Plots of *t*-statistics against average intensities and different components of *t*-statistics against each other (Figure 3) are especially useful in identifying genes with substantial statistics and in studying the features of these genes [7]. The average intensity for each gene is the average of the mean expression level for the patient and the control groups. In Figure 3, large *t*-statistics were easily visualized and labeled in red. They correspond to the same genes identified as outliers in the Q-Q plot (Figure 2). The six genes that have the largest *t*-statistics do so by virtue of having denominators close to zero, implying near constant expression levels. Further, their expression levels are observed to be very low. Consequently, these genes might be of little biological interest. In order to more systematically evaluate differential expression and complement these graphical approaches, we next apply SAM.

Application of SAM: t-SAM and Welch-SAM

t-SAM

Experimental outline

We used the SAM methodology to identify genes with statistically significant changes in expression. For the ith gene a modified *t*-statistic, $d(i)$, was used to quantitate differential expression.

$$d(i) = \frac{\bar{x}_{1i} - \bar{x}_{2i}}{S(i) + S_0}$$

All quantities here and below are defined in the Methods section. The s_0 factor was added to help ensure that the variance of $d(i)$ is independent of gene expression level and to mitigate problems caused by small denominators. A similar strategy derived from a Bayesian approach is detailed in [12]. In order to select an appropriate s_0 value for this dataset, we plotted Δ against FDR using different values of s_0. Figure 4 shows that calculating s_0 according to the Tusher, *et al.* prescription yields a value equal to the
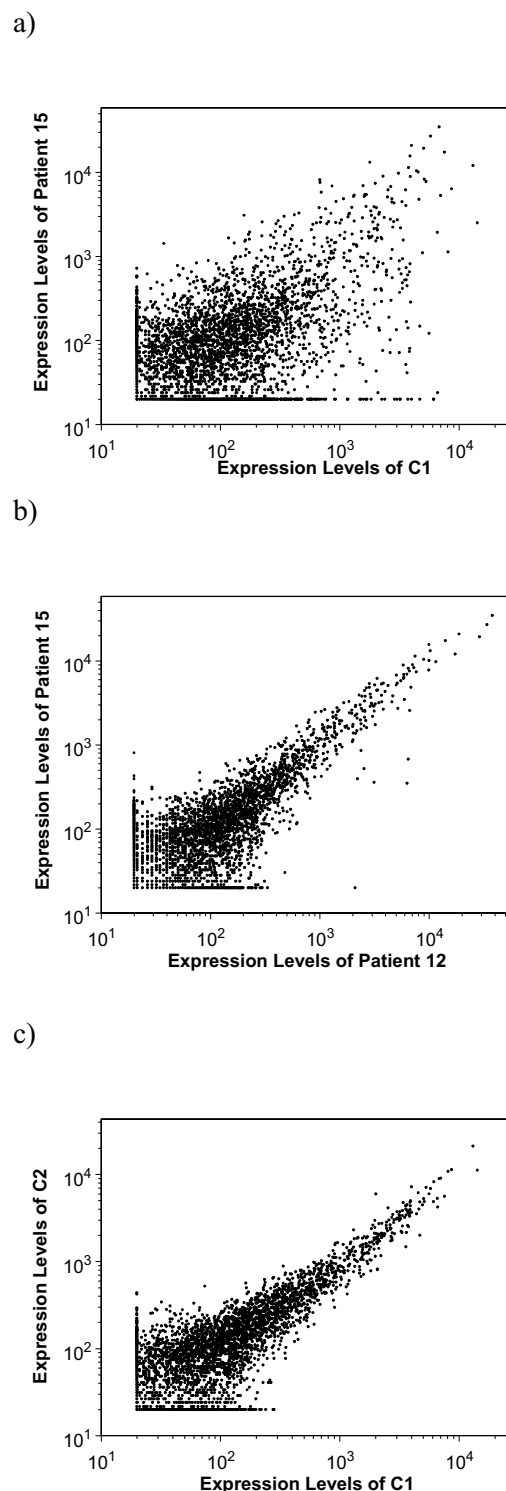
a)

b)

c)

Figure 1
Global comparison of the differences in gene expression levels between a) a patient (P15) and a normal subject (C1); b) a pair of patients (P15 and P12); c) a pair of normal subjects (C1 and C2).
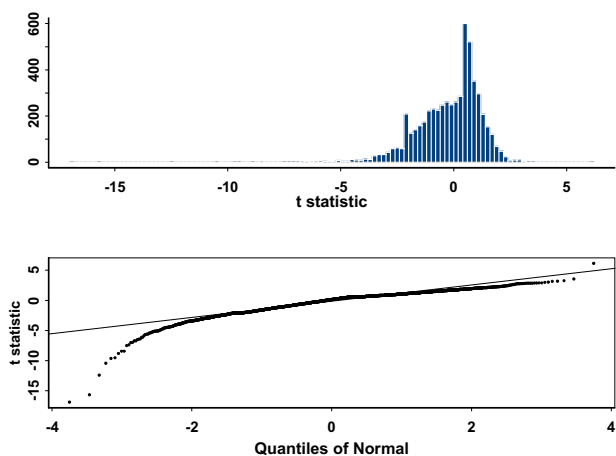
Figure 2
Histogram and QQ plot of *t*-statistics calculated from the normalized gene expression levels. Large *t*-statistics are labeled in red.

minimum of the pooled standard deviations over all the genes, which fails to reduce FDRs. We therefore elected to set s_0 at the 5th percentile of the pooled standard deviations (s_0 = 2.8); this choice does reduce FDR.

As Δ increases, the number of significant genes decreases, but at the same time the percentage of falsely called genes also decreases. The curve of FDR vs. Δ in Figure 4 shows that FDR generally decreases monotonically as Δ increases, but from some Δ on the decrease flattens markedly (see also Table 1). The choice of Δ is somewhat arbitrary, but is driven by choosing an acceptable FDR and an appropriate number of significant genes. We choose Δ = 12, corresponding to an FDR at 18%. This yields a set of 73 significant genes that are described next. Not surprisingly, the concordance between genes identified via graphical methods and those selected by SAM is high.

Genes identified by t-SAM
The 73 genes that are most highly ranked by t-SAM may represent more than one interacting molecular network and could be divided into different functional categories. Interestingly, only one gene (Autoantigen pericentriol material) has increased expression while the expression of the other 72 genes is decreased in the patient group. Unfortunately, in the present instance samples are no longer available and experimental verification of identified genes awaits future studies.

Cytokines / Neurotrophines
Mild tissue damage, nerve damage and infection produce inflammation and associated pain. This results in the migration of immune cells to the site of injury, which is fol-

Table 1: Comparison of t-statistics and Welch statistics in identifying significant genes under different values of Δs

Δ	Significant genes	False positives	FDR
	t-statistics		
2	622	312	50%
4	223	77.3	35%
6	131	35.1	27%
8	88	21.2	24%
12	73	13.3	18%
16	62	9.8	16%
20	57	7.7	14%
25	46	5.8	13%
40	30	3.4	11%
60	27	2.4	9%
90	19	1.5	8%
	Welch statistics		
2	1348	648.3	48%
4	525	155.9	30%
6	264	64.5	24%
8	97	24.0	25%
12	50	10.4	21%
16	20	4.0	20%
20	12	2.5	21%
25	7	1.0	14%
40	7	1.0	14%
60	6	0.4	7%
90	6	0.4	7%

lowed by release of cytokines [13]. Five genes identified as significant by *t*-SAM have been reported in the literature as relevant to the inflammation and pain pathways. Interleukine 13 inhibits inflammatory cytokine production. Midkine is associated with neurite growth and was recently discovered to have potent neuroprotective activity in vivo [14]. IGFBP6 plays a significant role in the differentiation, maintenance and regeneration of the central cholinergic neurons[15]. Cysteine-rich fibroblast growth factor receptor is also implicated in neuronal growth and differentiation. TrkA plays a crucial role in the development and function of the nociceptive reception system.

Myelin function
Three genes were identified that are involved in myelin structure and myelination: MPZ (myelin protein zero), PLP1 (myelin proteolipid protein 1) and MBP (myelin basic protein). Defects in these genes are the cause of demyelinating neuropathies.
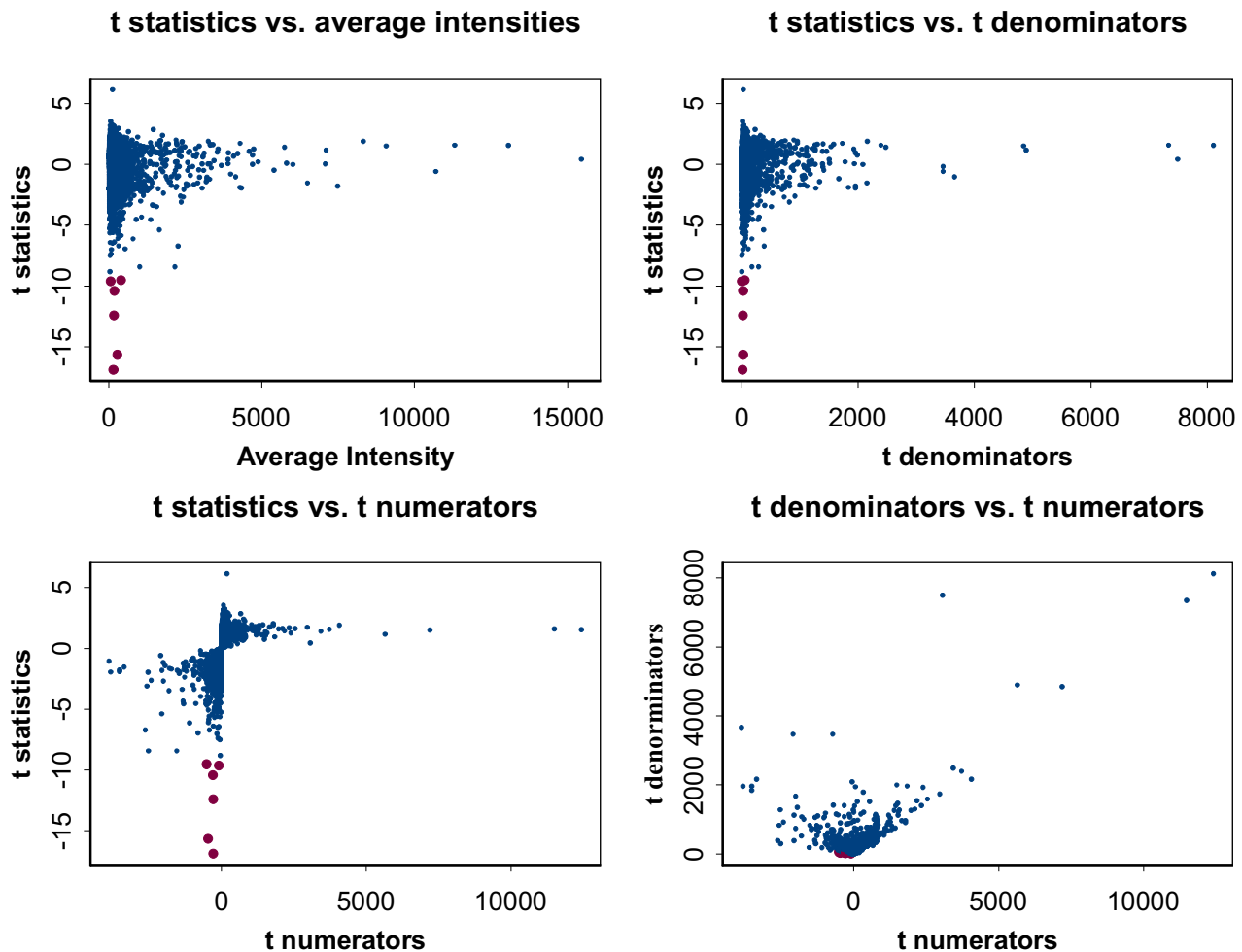
Figure 3
Plots of *t*-statistics, *t*-numerators, *t*-denominators and average intensities against each other. Large *t*-statistics are labeled in red.

Signal transduction
PKC and Frizzled play roles in apoptosis. The down regulation of both might represent a cellular response to prevent neurons from death.

Cytoskeleton
Three genes: Myosin light chain, Myosin heavy chain and MLCK (Myosin light-chain kinase) are notable in this category. The altered expression of them together suggests a probable link between calcium signaling and rearrangement of the cytoskeleton, possibly leading to long-term changes in neuronal morphology.

Others
MT-III has been recently implicated as being involved in sensory and nociceptive transmission [16]. Its down regu-

lation might also lead to promotion of neurite extension to promote recovery in the patients.

Welch-SAM
Thomas, *et al.*[10] showed that in the two-group microarray study setting the *t*-statistic assumption of equal variances between the two groups could result in poor performance. We therefore extended the SAM framework to include Welch statistics. Under the assumption of unequal variances, the associated degrees of freedom are variable, and the strict monotonic decreasing relationship between *p* values and test statistic values no longer holds. Accordingly, it is misplaced to use the values of the statistics under permutation. Rather, for each data permutation, we standardized the Welch statistic using an appropriate *t* referent distribution and used the resultant "*p* values" as described in the Method section; see also
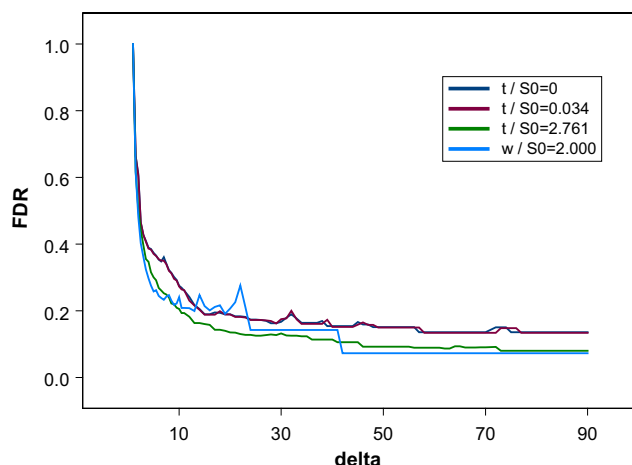
Figure 4
The relationship between FDR and in SAM. *t*-SAM is represented with three different lines with different s_0 values: $s_0 = 0$, $S_0 = 0.034$ (the minimum of the pooled standard deviations) and $S_0 = 2.761$ (5th percentile of the pooled standard deviations). Welch-SAM is graphed with $S_0 = 2.000$ (5th percentile of the pooled standard deviations).

Westfall and Young [17]. By so taking into account between-group heterogeneity, Welch-SAM identified a set of up regulated genes that were not detected by *t*-SAM. Several of these up regulated genes are engaged in transcription and translation regulation: transcription factor AP-4, transcriptional coactivator Pc4, Nrf2, EIF5 (Eukaryotic translation initiation factor 5), and PABPL1 (polyadenylate-binding protein 1). In addition, two genes are involved in cell cycle regulation: Autoantigen pericentriol material and CDC-like kinase. Their identification suggests a secondary mechanism to repair the damaged cells. The entire list of genes identified by *t*-SAM and Welch-SAM can be found at [2].

t-SAM vs. Welch-SAM
The relationship between FDR and Δ for Welch is also included in Figure 4. The un-pooled variance in the denominator of Welch statistic is highly unstable under permutation in this small sample setting. Consequently, FDR for Welch-SAM is not a smooth function of Δ, as is the case of *t*-SAM. FDR values for selected Δs for both *t*-SAM and Welch-SAM when s_0 was chosen at the 5th percentile of standard deviations, are listed in Table 1 along with corresponding numbers of significant genes and false positives.

Application of Westfall and Young step-down adjusted p values
Westfall and Young's step-down adjusted p values were calculated using both t-statistics and Welch statistics. As

before, the Welch statistics and p values are not monotonely related because of the variation in degrees of freedom. So, we again calibrated the Welch statistic against the appropriate t referent distribution and determined (unadjusted) "p values" for each permutation. The smallest Westfall and Young adjusted p values for t and Welch statistics were 0.053, 0.042 respectively. Figure 5 illustrates the relationship between adjusted p values and the number of genes being tested. A small simulation shows that this relationship is governed by the number of true null hypotheses in the tests (unpublished data). Profiles for nine selected genes are displayed; t-statistics are presented in Figure 5A and Welch-statistics in Figure 5B. For each gene, the adjusted p value increased with the increase of the number of null genes being tested. Genes with smaller statistics had greater adjusted p values and their adjusted p values increased at a greater rate when the number of null genes being tested was larger.

Discussion
FWER vs. FDR
In a single hypothesis test, a Type I error occurs when a true hypothesis is rejected. In multiple testing, FWER refers to the probability of rejection of any true hypothesis. When many hypotheses are tested, such as expression data of thousands of genes being compared between different groups, FWER increases dramatically, and is typically much larger than the significance level at which the individual hypotheses are tested.

As defined by Shaffer, given any test procedure, an adjusted *p* value corresponding to the test of a single hypothesis H_j is the level of the entire test procedure at which H_j would just be rejected, given the values of all test statistics involved [18]. Several approaches have been proposed to calculate adjusted *p* values for the purpose of providing control of FWER. Among these, the method developed by Westfall and Young is best suited for microarray data analysis. It is superior to the single-step methods proposed by Bonferroni and Šidák by allowing different *p* values to be adjusted differently; therefore the power of the procedure is improved. It is also able to take into account the dependence structure between variables. In a biological system, expression levels of groups of genes may correlate with each other for various reasons such as co-regulation.

However, when the Westfall and Young's permutation algorithm was applied to calculate step-down adjusted *p* values in our dataset, only one of the adjusted *p* values from Welch statistics reached a conventional significance level of 0.05. A few factors might contribute to the absence of adjusted *p* values at that significance level. First and foremost, the small sample sizes are limiting. Furthermore, the unbalanced allocation of (10) patients and (3) controls limits the precision of the adjusted *p* values. The
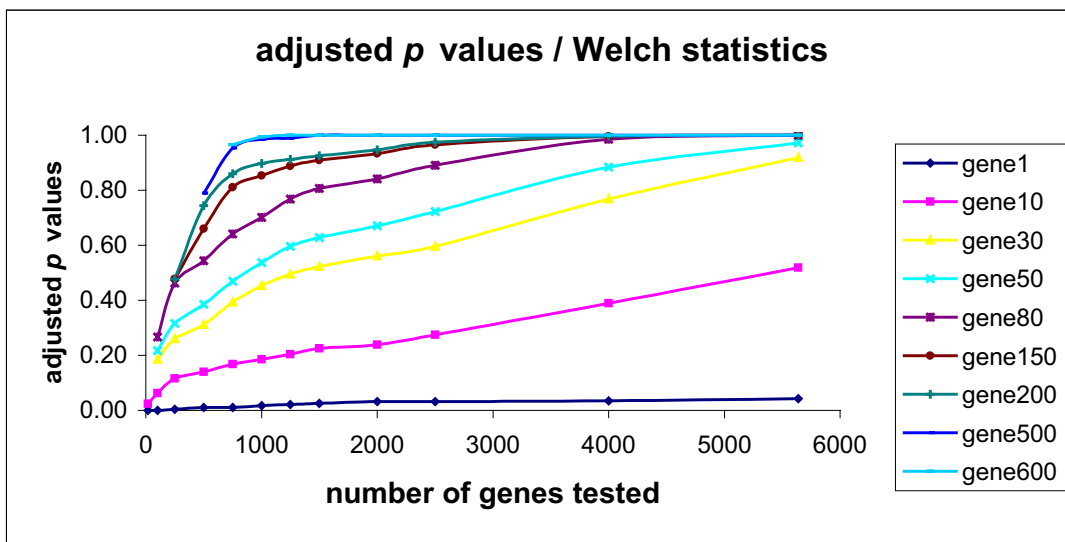
a)
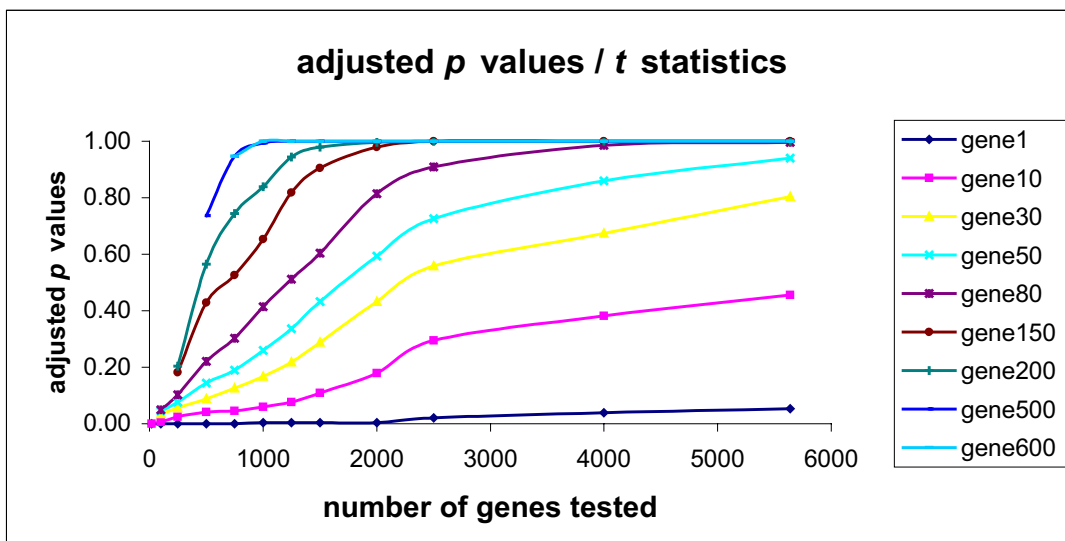
b)

Figure 5
The relationship between adjusted *p* values and number of genes tested. a) *t*-statistics; b) Welch statistics. Colored lines are genes ordered by their unadjusted *p* values, e.g. gene1 has the smallest unadjusted *p* value and gene600 has the 600th smallest unadjusted p value. c) A comparison of panel a) and b). Red lines represent the results of Welch statistics and black lines represent the results of *t*-statistics.

number of permutations we can do is only 285. A more balanced experimental design will increase the number to 1715 (13 choose 6, then minus 1). Second, the vast majority of genes are expressed at low levels, as depicted in Figure 3, where biological signals may become indistinguishable from noise. For those cases where the difference in mean expression levels is small, a pooled standard deviation that is also small can give rise to a significant statistic, where in fact no difference exists. Third, Figure 5 shows that adjusted *p* values are sensitive to the number of null genes being tested. Given that many genes are likely to be irrelevant to the biological processes underlying the brachial plexus injuries and hence are expected to show little difference in their expression levels between injured patients and control subjects, it is not surprising that no significant result was obtained for *t*-SAM when applying the complete data set. In Figure 5 we also try to illustrate the potential abuse of using filtered gene subsets (based on statistics measuring differential expression) to achieve significant results, given that such filtering is commonplace in microarray literature. Last, although Westfall and Young's algorithm is more powerful than Bonferroni, Šidák, and Holm adjustments by taking into account the joint distribution of the test statistics, it may still be too stringent. The same result also occurred for Tusher, *et al.*[3] when applying this algorithm to study the transcriptional response of lymphoblastoid cells to ionizing radiation. A less stringent method may therefore be needed to analyze microarray data.

The FDR provides a different point of view on how the errors in multiple testing should be controlled. Having many hypotheses rejected signals clearly that many hypotheses are not true. It is therefore more crucial to control the rate of false positives among all rejected hypotheses than the probability of one single erroneous rejection. It might thus be argued that the control of FDR is a more appropriate approach to dealing with the multiplicity concern. That argument is especially true in the setting of microarray data analysis, where one is more interested in studying as many relevant genes as possible, and less interested in sacrificing possible targets for the concern of making one mistake [6]. The algorithm of SAM proposed by Tusher, *et al.*[3] provides an estimate of the FDR for each value of the parameter Δ in the application of microarray settings. The estimated FDR is computed from permutation of the data and allows for the possibility of dependent tests [3]. We derived a set of 73 significant genes using SAM, out of which 13 were false positives, giving rise to a FDR of 18%. Although SAM is not able to disclose the identities of these false positives, it does provide biologists a reasonable set of potential target genes and a sense of the trustworthiness of the outcome. By estimating FDR by permutations of the data, SAM assumes that all null hypotheses are true; hence the

estimated FDR is biased upward. Storey and Tibshirani tackled this problem by multiplying FDR by an estimate of the proportion of true null hypotheses, π_0. A detailed description of the algorithm can be found in Storey and Tibshirani [6]. Following their approach, the 18% false discovery rate is lowered to about 13% by assuming that not all hypotheses are null. Further recent comparisons of multiple testing procedures are provided by Holland and Cheung [19].

t-statistics vs. Welch statistics
We applied both *t*-statistics and Welch statistics. The Welch statistics identified a set of genes having increased expression in the patient group that were not detected by *t*-statistics under the same stringency. One can argue that the active genes have greater variability in gene expression than inactive ones have [20]. Such a difference will be reflected in larger variances in the patient group relative to the control group, a scenario more appropriate for Welch statistics than *t*-statistics. In addition, subjects (patients) in our analysis might be in different conditions. Some patients might have more tissue damage and/or more severe pain than others. Further, the patient group is likely to be inherently more heterogeneous with respect to the extent of tissue damage and other relevant yet unmeasured attributes. These considerations also argue for the appropriateness of Welch statistics. Thomas, *et al.* also found that Welch statistics conformed better than *t*-statistics to their regression modeling approach used to discover differentially expressed genes in a microarray setting where between-group heterogeneity was evident [10]. There are drawbacks, however, to using the Welch statistics. Firstly, they need to be calibrated, introducing distributional assumptions. Secondly, and more importantly, there are stability and power concerns that pertain in small sample settings [17].

Conclusion
The use of high-density microarray technology to identify specific genes that are expressed differently under one or more biological conditions is particularly relevant to drug discovery and development. However, in the context of microarray data analysis, multiple testing concerns are forefront. In this study, we attacked the very important problem of application of appropriate multiple testing methods in identifying differential gene expression, and in addition, we have provided a framework for future experimental research in brachial plexus injuries and pain pathways. Simple statistical summaries, such as fold change, *t*-statistics and Welch statistics, have been applied. To further investigate some recently proposed multiple testing schemes, we compared the performance of two approaches, SAM and Westfall and Young step down adjusted *p* values, as applied to a microarray study of gene expression in a brachial plexus injury. Our results show

that using SAM and controlling FDR leads to identification of a set of potentially interesting genes consistent with prior knowledge of their function. In addition, SAM readily quantitates the trade-offs between false discovery rates and numbers of selected genes. On the other hand, the Westfall and Young algorithm, by controlling FWER, is highly conservative in this small sample setting. To the extent that FWER provides appreciably more stringent selection, attendant findings of significant differential gene expression are more likely to be validated (by follow-up experiment) than those deriving from FDR-controlling methods. However, there has been recent recognition [5,6,21] and our study also shows that FWER-control can be unnecessarily stringent since falsely selecting a few genes will not be a serious problem if the majority of differentially expressed genes are chosen correctly. In addition to comparing multiple testing algorithms, we also compared the performance of *t*-statistics and Welch statistics. By taking into account the heterogeneous nature of the patient group, Welch statistics identified a set of genes having increased expression in the patient group, whereas these genes were not selected by the *t*-statistics under the same criteria.

Methods
Materials
Briefly, cervical avulsed human DRG tissues (n = 10) were removed in the setting of a dorsal rhizotomy of lower cervical (and upper thoracic in one case) nerve roots in patients with persistent pain syndromes after accidental avulsion injuries. Details on obtaining, processing and measuring the expression of total RNA in the human DRG and control tissue samples are located elsewhere [http://itsa.ucsf.edu/~yxiao/Avulsion]. The control human DRG sample was pooled from total RNA (n = 8; prepared by Clontech) obtained from post-mortem tissues. The pool of n=8 was then run in triplicate (n=3). Total RNA was extracted from the tissue by Trizol (LifeTechnologies) extraction and polyA+ RNA was recovered by Oligotex mRNA Spin Columns (Qiagen). Biotinylated cRNA targets were prepared and hybridized to Affymetrix HU6800 oligonucleotide arrays according to the manufacturer's protocol (Affymetrix, Santa Clara, California).

Data analysis methods
Normalization
All data from patient and control samples were preprocessed (background subtraction, differential intensities) using Affymetrix GeneChip software, and average differential intensity (ADI) was used for analysis. The average of all ADI values of each chip was calculated and the mean of average values of all chips was used as the target intensity. The ADI value of each chip was normalized to be equal to the target intensity. Since negative, zero, and very small positive values (<20) of ADI most likely represent

noise instead of actual gene expression, all such values were truncated at 20 resulting in 5,638 probe sets for analysis.

Graphical display
A Quantile-Quantile plot (Q-Q plot) was generated by plotting the quantiles of *t* statistics of all genes against the quantiles of a standard normal distribution. In addition to examining the distribution of *t* statistics, we used this plot to observe points that deviate markedly from the bulk of the observations since they may represent difference in means of the patient and control groups. Four scatter plots were generated to study the features of large *t* statistics: *t* statistics vs. average intensities, *t* statistics vs. *t* denominators, *t* statistics vs. *t* numerators, and *t* numerators vs. *t* denominators.

Significance Analysis of Microarrays (SAM)
The SAM procedure described by Tusher *et al.* [3] based on the *t*-statistic (under the assumption of equal variances between the two groups) and the Welch statistic (under the assumption of unequal variances between the two groups) was modified as follows:

1. The relative difference in the expression *d(i)* for the *i*th of *k* = 5638 probe sets was defined as,

$$d(i) = \frac{\bar{x}_{1i} - \bar{x}_{2i}}{S(i) + S_0} \; ; i = 1, 2, ..., k$$

where $\bar{x}_{1i}$ and $\bar{x}_{2i}$ are means of expression levels of probe set *i* in the patient and control groups, respectively.

For statistics,

$$S(i) = \sqrt{\frac{(n_1 - 1)S_{1i}^2 + (n_2 - 1)S_{2i}^2}{n_1 + n_2 - 2}}$$

and for Welch statistics,

$$S(i) = \sqrt{\frac{S_{1i}^2}{n_1} + \frac{S_{2i}^2}{n_2}} \; ;$$

where S_{1i}^2 and S_{2i}^2 are the variances of expression levels for probe set *i* in the patient and control groups. The determination of s_0 is discussed in the Results section. For all *d(i)* values, corresponding *p* values were calculated with *d(i)* referenced to an appropriate t distribution.

2. All *p* values were ordered: $p_1 \leq p_2 \leq ... \leq p_k$.

3. An exhaustive set of 285 (286 minus the original configuration, 286 being 13 choose 3) permutations were conducted. For each permutation *b*, *p* values were calculated and ordered in the way of

$$p_1^{(b)} \le p_2^{(b)} \le ... \le p_k^{(b)}.$$

4. Expected *p* values for all probe sets were calculated as

$$\bar{p}_i = \frac{\sum_b p_i^{(b)}}{N} \text{, where } i = 1, 2, ..., k.$$

5. For a fixed threshold Δ, a probe set was called "significant" if its *p* value satisfied the criterion

$$\frac{\bar{p}_i}{p_i} \ge \Delta .$$

The total number of "significant" probe sets was counted. The $p_i$largest among the "significant" probe set was defined as the cutoff value.

6. For each permutation, all probe sets whose *p* values were smaller than the cutoff value were found and were called "falsely significant". The total number of these probe sets was counted. The estimated number of "falsely significant" probe sets was defined as the average of the number of "falsely significant" probe sets from all permutations. The false discovery rate (FDR) was computed as the ratio of the estimated number of "falsely significant" probe sets to the total number of "significant" probe sets.

7. FDR was computed with different values of threshold Δ.

Westfall and Young step-down adjusted p values
Both *t*-statistics and Welch statistics were used to compare gene expression between the patient and control groups. Our data were organized in a matrix in which each row represents the expression of one gene for all subjects and each column represents the expression of all genes for one subject. The permutation algorithm developed by Westfall and Young to obtain step-down adjusted *p* values was then applied as follows. Note that the procedures were modified when using Welch statistics. Instead of using the Welch statistic to compute adjusted *p* values, we used unadjusted *p* values derived from Welch statistics and the appropriate degrees of freedom. An exhaustive set of 285 permutations were conducted.

1. Compute the *t* statistic for each gene in the original dataset.

2. Order them: $|t_{r1}| \ge |t_{r2}| \ge |t_{r3}| \ge ... \ge |t_{rk}|$.

3. Permute the 13 columns of the data matrix. The first 10 columns now represent the pseudo-patient group and the other 3 columns represent the pseudo-control group.

4. Compute *t* statistics for all probe sets for the permuted dataset:

$$t_1^{(b)}, t_2^{(b)}, ..., t_k^{(b)}.$$

5. Compute $u_k^{(b)} = \left| t_{rk}^{(b)} \right|$

and

$$u_j^{(b)} = \max \left(u_{j+1}^{(b)}, \left| t_{rj}^{(b)} \right| \right),$$

$1 \le j \le k - 1$, where r_j is such that $|t_{r1}| \ge |t_{r2}| \ge |t_{r3}| \ge ... \ge |t_{rk}|$ for the original dataset.

6. Repeat 1–5 *N* (*N* = 286) times and calculate the adjusted *p* values:

$$\tilde{p}_{rj}^* = \frac{\sum_{b=1}^{N} I\left(u_j^{(b)} \ge \left| t_{rj} \right| \right)}{N},$$

where I(•) is the indicator function setting to 1 if the condition in parentheses is true and 0 if false. The monotonicity was enforced as

$$\tilde{p}_{r1}^* = \tilde{p}_{r1}^*, \tilde{p}_{rj}^* = \max \left(\tilde{p}_{rj}^*, \tilde{p}_{rj-1}^* \right) \text{, for } 2 \le j \le k.$$

Supplementary material
Supplementary material is located at [2]. This web page includes our data analysis results and also further tissue and patient information.

Authors' contributions
YX performed the statistical studies, participated in its co-ordination, and drafted the manuscript. MRS initiated and designed the statistical analyses and drafted the manuscript. DR carried out the tissue preparation and microarray hybridization. AHA participated in the drafting of the manuscript regarding the biology of avulsion injuries. PA initiated and organized the study, and participated in the collection and tissue dissection of the DRG. LS initiated the study and participated in its design and coordination. DH participated in the statistical analyses and the initial manuscript drafting. CAH initiated the different methods of statistical analysis and participated in its design, coordination and manuscript drafting. YX and MRS contributed equally to this work.

dustry.berkeley.edu/sectorprograms/lsi.htm] of the industry-University Cooperative Research Program.

Note

Revisions to the text were added in proof

References

1. Berman JS, Birch R, Anand P: **Pain following human brachial plexus injury with spinal cord root avulsion and the effect of surgery.** *Pain* 1998, **75:**199-207
2. **Avulsion Dataset Home Page** [http://itsa.ucsf.edu/~yxiao/Avulsion]
3. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98:**5116-5121
4. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Stat Soc B* 1995, **57:**289-300
5. Storey JD: **A direct approach to false discovery rates.** *J Roy Stat Soc B* 2001
6. Storey JD, Tibshirani R: **Estimating false discovery rates under dependence, with applications to DNA microarrays.** *Stanford: Stanford University, Department of Statistics; Report No.: Technical Report 2001–28* 2001
7. Dudoit S, Yang Y, Callow M, Speed T: **Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments,** *Report No.: Technical Report #578.* 2000
8. Callow MJ, Dudoit S, Gong EL, Speed TP, Rubin EM: **Microarray expression profiling identifies genes with altered expression in HDL-deficient mice.** *Genome Res* 2000, **10:**2022-2029
9. **Working with Multiple Microarrays** [http://biosun1.harvard.edu/~rgentlem/Wshop/lect2.pdf]
10. Thomas JG, Olson JM, Tapscott SJ, Zhao LP: **An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles.** *Genome Res* 2001, **11:**1227-1236
11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al*: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286:**531-537
12. Lonnstedt I, Speed TP: **Replicated Microarray Data.** *Statistical Sinica* 2002, **12:** [http://stat-www.berkeley.edu/users/terry/zarray/Html/ papersindex.html]
13. Facer P, Mann D, Mathur R, Pandya S, Ladiwala U, Singhal B, Hongo J-A, Sinicropi DV, Terenghi G, Anand P: **Do nerve growth factor-related mechanisms contribute to loss of cutaneous nociception in leprosy?** *Pain* 2000, **85:**231-238
14. Yoshida Y, Ikematsu S, Moritoyo T, Goto M, Tsutsui J, Sakuma S, Osame M, Muramatsu T: **Intraventricular administration of the neurotrophic factor midkine ameliorates hippocampal delayed neuronal death following transient forebrain ischemia in gerbils.** *Brain Res* 2001, **894:**46-55
15. Konishi Y, Takahashi K, Chui DH, Rosenfeld RG, Himeno M, Tabira T: **Insulin-like growth factor II promotes in vitro cholinergic development of mouse septal neurons: comparison with the effects of insulin-like growth factor I.** *Brain Res* 1994, **649:**53-61
16. Yu WH, Lukiw WJ, Bergeron C, Niznik HB, Fraser PE: **Metallothionein III is reduced in Alzheimer's disease.** *Brain Res* 2001, **894:**37-45
17. Westfall PH, Young SS: *Resampling-based Multiple testing: examples and methods for p-value adjustment. New York: Wiley;* 1993
18. Shaffer JP: **Multiple Hypothesis Testing.** *Annu Rev Psychol* 1995, **46:**561-584
19. Holland BB, Cheung SH: **Familywise robustness criteria for multiple-comparison procedures.** *J Roy Stat Soc B* 2002, **64:**63-77
20. Wu TD: **Analysing gene expression data from DNA microarrays to identify candidate genes.** *J Pathol* 2001, **195:**53-65
21. Smyt GK, Yang YH, Speed TP: **Statistical issues in cDNA microarray data analysis.** 2002 [http://www.stat.Berkeley.EDU/users/terry/zarray/Html/papersindex.html]