Published online 13 October 2009

Nucleic Acids Research, 2009, Vol. 37, No. 22 ***7483–7497***
doi:10.1093/nar/gkp813

Local false discovery rate facilitates comparison of different microarray experiments

Wan-Jen Hong[1,2], Robert Tibshirani[3,4] and Gilbert Chu[1,*]

[1]Department of Medicine, [2]Department of Biochemistry, [3]Department of Statistics and [4]Health Research and Policy, Stanford University Medical Center, Stanford, CA 94305, USA

Received August 15, 2009; Accepted September 14, 2009

ABSTRACT

The local false discovery rate (LFDR) estimates the probability of falsely identifying specific genes with changes in expression. In computer simulations, LFDR <10% successfully identified genes with changes in expression, while LFDR >90% identified genes without changes. We used LFDR to compare different microarray experiments quantitatively: (i) Venn diagrams of genes with and without changes in expression, (ii) scatter plots of the genes, (iii) correlation coefficients in the scatter plots and (iv) distributions of gene function. To illustrate, we compared three methods for pre-processing microarray data. Correlations between methods were high ($r = 0.84$–0.92). However, responses were often different in magnitude, and sometimes discordant, even though the methods used the same raw data. LFDR complements functional assessments like gene set enrichment analysis. To illustrate, we compared responses to ultraviolet radiation (UV), ionizing radiation (IR) and tobacco smoke. Compared to unresponsive genes, genes responsive to both UV and IR were enriched for cell cycle, mitosis, and DNA repair functions. Genes responsive to UV but not IR were depleted for cell adhesion functions. Genes responsive to tobacco smoke were enriched for detoxification functions. Thus, LFDR reveals differences and similarities among experiments.

INTRODUCTION

To understand complex biological systems, methods are needed for comparing different experiments on a genomic or proteomic scale. For example, ultraviolet (UV) and ionizing radiation (IR) generate DNA damage in different ways, producing thymine dimers and double strand breaks, respectively. Another DNA damaging agent, tobacco smoke, produces benzo[a]pyrene adducts on guanine bases in DNA. Methods are needed to compare microarray experiments in order to identify genes that respond to different agents, as well as genes that respond to one agent but fail to respond to others.

Significance analysis of microarrays (SAM) identifies genes that respond to a perturbation (1). SAM assigns each gene a *d*-score *d(i)*, which can be used to estimate the probability that gene *(i)* has changed expression. SAM estimates the false discovery rate (FDR) by randomly permuting the sample labels to estimate the number of genes that by chance would have a *d(i)* score greater than an adjustable threshold. A 5% FDR means that 5% of the genes ranked higher than a threshold value were falsely identified as significant. The *q*-value for each gene *(i)* is the FDR for the set consisting of gene *(i)* and all higher ranked genes (2). Investigators find the q-value to be useful, but one must remember that the *q*-value for gene *(i)* is lower than the probability that gene *(i)* itself was falsely identified.

Others have proposed using the local false discovery rate (LFDR) to estimate the probability that gene *(i)* was falsely identified (3–9). Unlike *q*-value, estimates of LFDR calculate the false discovery rate for genes ranked near gene *(i)*. Thus, LFDR is based on genes with *d*-scores similar to gene *(i)*, and not on genes with more extreme *d*-scores. Previous methods for estimating LFDR have yielded comparable results. Here, we chose to estimate the LFDR for gene *(i)* by counting the number of falsely discovered genes with scores in the local neighborhood of *d(i)* after random permutation of sample labels.

Others have used LFDR to identify genes with changes in expression. Here, we show that LFDR can also identify genes without changes in expression. A gene without a change in expression in one experiment may be of particular interest, if the same gene changes expression in a second experiment.

While FDR characterizes a set of genes in a particular experiment, LFDR characterizes each individual gene. We hypothesized that LFDR could thus identify genes that either change or fail to change in expression, and thus facilitate comparisons between different experiments.

*To whom correspondence should be addressed. Tel: +1 650 725 6442; Fax: +1 650 736 2282; Email: chu@stanford.edu

We confirmed that LFDR could successfully identify genes that change or fail to change for computer-simulated data. To compare microarray experiments graphically and quantitatively, we exploited several tools: Venn diagrams, scatter plots, Pearson correlation coefficients and distributions of gene function. To illustrate the utility of these tools, we compared responses to UV, IR and tobacco smoke. We also compared results generated by three methods of pre-processing a single set of raw microarray data.

MATERIALS AND METHODS

Cell lines and treatment with UV and IR

Fifteen healthy individuals were enrolled as described previously (10). Lymphoblastoid cells were established by immortalizing peripheral blood B-lymphocytes with Epstein-Barr virus. Cells were irradiated with UV using a germicidal lamp (254 nm) to a dose of 10 J/m^2 and harvested for RNA 24 h later. For IR treatment, cells were exposed to 5 Gy of ^{137}Cs γ-rays and harvested for RNA 4 h later.

Microarray analysis

Total RNA was labeled with biotin and hybridized to an U95A_v2 GeneChip® microarray according to the manufacturer's protocols (Affymetrix, Santa Clara, CA). The expression level for each probe set was computed by Affymetrix Microarray Suite (MAS) version 5.0 software. Data were scaled to the average of all datasets, as described in ref. (1). Two other pre-processing methods were used to compute gene expression levels: gene-chip robust multi-array average (GCRMA) (available at the BioConductor website, http://www.bioconductor.org/) and DNA-Chip analyzer (dChipv1.3) based on the model-based expression index using PM only (available at the dChip website, http://www.dchip.org/). The complete dataset is available on the Gene Expression Omnibus (GEO) database, http://www.ncbi.nlm.nih.gov/geo/.

We used the paired option in SAM version 1.21 to compute the *d*-scores for genes in the simulated data set and for probe sets in the UV and IR datasets. The Excel plug-in software is available at http://www-stat.stanford.edu/~tibs/SAM/.

Estimation of LFDR

We estimated the LFDR for gene *(i)* by the following procedure:

(1) Compute *d*-score *d(i)* for each gene *(i)* in the array using SAM (1).
(2) Assign rank *r(i)* to gene *(i)*, where *r(i)* is the number of genes with *d*-scores $\geq d(i)$.
(3) For a window of *n* genes, compute *d*-scores, $d_1(i)$ and $d_2(i)$, for the genes ranked $r(i) - n/2$ and $r(i) + n/2$, respectively.
(4) Permute the sample labels and count the resulting number of genes $n_p(i)$ with *d*-scores in the neighborhood of *d(i)* given by the interval $[d_1(i), d_2(i)]$.
(5) Estimate the LFDR as

$$LFDR(i) = \frac{\pi_0 \times n_p(i)}{n} \quad \mathbf{1}$$

Here, π_0 is the fraction of genes that did not change expression, as estimated by SAM (11). The interval $[d_1(i), d_2(i)]$ is defined as the 'window' for the LFDR. For example, a 1% window corresponds to *n* equal to 1% of the total number of genes.

We generated a smooth curve of the LFDR values with a smoothing function, which is available in the *R* Functional Data Analysis (FDA) package (12). Biological functions were assigned from published literature and from the Gene Ontology (GO) database through the Affymetrix NetAFFX™ Analysis Center.

Smoking induced epithelial gene expression data

Affymetrix HG-U133A GeneChip expression data from epithelial cells of 40 current smokers, 25 non-smokers and 13 former smokers was downloaded from the Smoking Induced Epithelial Gene Expression (SIEGE) database (13) on 4 January 2005. Expression values were estimated using MAS 5.0 software. The multi-class option in SAM was used to compute the *d*-scores for each probe set using 1000 permutations. Hierarchical clustering of the genes for each group of 250 genes was performed using all 81 samples using uncentered Pearson correlation and complete linkage and displayed using Treeview (14). Hierarchical clustering of the samples was then performed separately within each class. The values used for clustering were the logarithm of the ratio of the expression value of the gene to the median of the expression values of each gene across all samples.

RESULTS

LFDR identifies genes from simulated data with and without changes in expression

To assess the accuracy of LFDR, we generated computer-simulated data containing genes with and without changes in expression. The data represented a simulated experiment with 15 measurements for the expression of 10 000 genes (Figure 1A). We introduced experimental noise by generating data with a normal distribution centered on an average change in expression of $\Delta x = 0$, ± 0.5, ± 1.0, ± 1.5, ± 2.0 or ± 2.5 U, with standard deviations of 1 U. A total of 5000 genes had data equally distributed among $\Delta x = \pm 0.5$, ± 1.0, ± 1.5, ± 2.0 and ± 2.5 U. Since the largest number of genes in a biological experiment undergo insignificant changes in expression, 5000 genes had data corresponding to $\Delta x = 0$.

To validate the computer simulation, we computed the SAM *d*-score for each simulated gene (Figure 1B). Approximately 35% of the unpermuted *d*-scores had outlying values of < -1 or $> +1$, compared to 2% of randomly permuted *d*-scores. The number of genes with outlying *d*-scores reflected the number of genes with changes in expression. In particular, 30% (3000) of the simulated genes had been assigned changes exceeding the

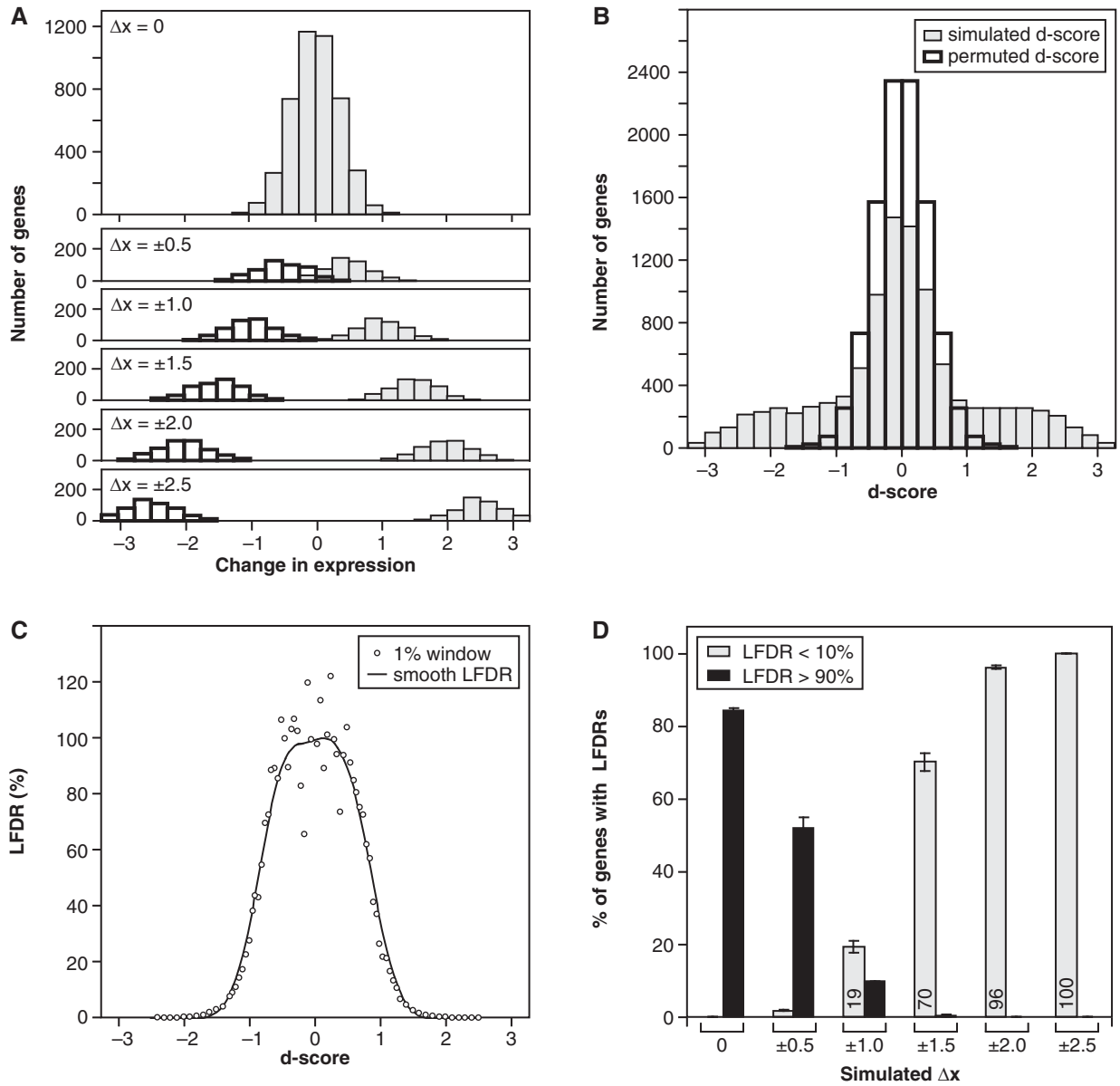


Figure 1. LFDR identified genes with and without simulated changes in expression. (**A**) Computer simulation generated changes in expression. We simulated changes in expression before and after a perturbation for 10 000 genes. Experimental noise among 15 replicates generated a normal distribution centered on each pre-specified change in expression, Δx, with a standard deviation of 1 U. The change in expression was $\Delta x = 0$ for 5000 genes, and $\Delta x = \pm 0.5$, ± 1.0, ± 1.5, ± 2.0, and ± 2.5 U equally distributed among the remaining 5000 genes. Grey bars represent the distribution of genes with positive changes in expression, and white bars outlined in black represent the distribution of genes with negative changes. (**B**) The distribution of *d*-scores for simulated changes in expression. We used paired SAM to estimate the *d*-score for each gene in (A). Grey bars indicate the distribution of genes as a function of *d*-score. White bars outlined in black indicate the distribution of genes as a function of 'permuted' *d*-scores generated by 500 permutations, in which data before and after the perturbation were switched for randomly chosen replicates. (**C**) LFDR estimates contain fluctuations that can be controlled by a smoothing function. We estimated the LFDR for each gene using a 1% window. Each point represents a selected gene. Because of fluctuations in the calculated LFDR values, we estimated LFDR for each gene from a smoothing function represented by the solid line. (**D**) LFDR accurately identifies genes with or without changes in expression. For each change in expression in the simulated data of (A), the histogram shows the percentage of genes identified as changing (LFDR $<10\%$, gray bars) or not changing (LFDR $>90\%$, black bars). For example, LFDR $>90\%$ identified as not changing 84% of the genes with pre-specified changes of zero ($\Delta x = 0$), and LFDR $<10\%$ identified as changing 96% of the genes with pre-specified changes of two standard deviations ($\Delta x = 2.0$). The error bars indicate the standard deviation of three simulated experiments.

noise level of 1 U ($\Delta x = \pm 1.5$, ± 2.0, or ± 2.5 units), and another 10% of the genes had changes equal to the noise level ($\Delta x = \pm 1.0$ unit). Conversely, 65% of the 10 000 genes had *d*-scores between -1 and $+1$, which reflected the number of genes with changes in expression smaller than the noise level. In particular, 60% of the genes had been assigned changes smaller than 1 U ($\Delta x = 0$, ± 0.5), and another 10% had changes equal to the noise level.

To further validate the computer simulation, we estimated π_0, the fraction of genes without changes in

expression. SAM divides the randomly permuted data into quartiles by *d*-score. Genes with *d*-scores in the middle two quartiles of the permuted data are assumed to be unchanged in expression. Then, π_0 equals the number of genes from the unpermuted data with *d*-scores that would fall in the middle two quartiles divided by the number of genes with the same range of *d*-scores from the permuted data (i.e. 50% of all the genes in the experiment (11).

For the simulated data, SAM estimated $\pi_0 = 0.58$. This was consistent with the simulated input, in which 50% of the genes had no change in expression, and another 10% had a change smaller than noise in the data, $\Delta x = \pm 0.5$. Thus, SAM generated *d*-scores consistent with the input for the simulated data.

To compute the LFDR for each gene, we employed a 1% window. In other words, we created a 'local' window containing 1% of all the genes, chosen for having d-scores closest to the d-score of the index gene. The LFDR was calculated as the number of the genes from the permuted data that were contained in the window, divided by the number of genes from the unpermuted data (i.e. 1% of all the genes in the experiment). We then fitted the data with a smoothing function (Figure 1C). As expected, LFDR >90% identified genes with low *d*-scores ($-0.5 < d < +0.5$), and LFDR <10% identified genes with extreme *d*-scores. To evaluate the utility of LFDR, we chose LFDR >90% as the criterion that a gene has failed to change expression, and LFDR <10% as the criterion that a gene has changed expression. Of course, one could adjust the stringency of these criteria, depending on the particular application.

We assessed the accuracy of LFDR <10% in identifying the genes with simulated changes in expression (Figure 1D). For the 2000 genes with $\Delta x = \pm 2.0$ and ± 2.5, almost all (96 and 99.9%, respectively) had values for LFDR <10%, and thus identified as changing. When the change in expression was smaller than the noise, $\Delta x = \pm 0.5$, only 1.7% of the genes were identified as changing. Thus, LFDR <10% accurately identified genes with changes in expression, provided that the changes in expression were greater than the level of noise in the simulated data.

We also assessed the accuracy of LFDR >90% in identifying genes that were not changing. For the 5000 genes with $\Delta x = 0$, a total of 84% were identified as not changing. Also, 52% of genes with $\Delta x = \pm 0.5$ were identified as not changing. For genes with a modest change in expression, $\Delta x = \pm 1.5$, only 0.3% of the genes were identified as not changing. For larger changes in expression, $\Delta x = \pm 2.0$ and ± 2.5, genes were never identified as not changing. Thus, LFDR >90% accurately identified genes without significant changes in expression.

LFDR can be estimated from experimental data

Next, we applied LFDR to actual microarray data. We previously collected data for UV responses in lymphoblastoid cell lines from 15 healthy individuals, using oligonucleotide microarrays for 12 625 probe sets (10). Figure 2A plots the distribution of genes as a function of *d*-scores for unpermuted and permuted data. SAM estimated $\pi_0 = 0.70$, indicating that 70% of the probe sets did not undergo changes in expression.

To assess the effect of window size, we estimated LFDR with 1 and 10% windows (Figure 2B). The 1% window generated LFDR estimates with smooth behavior for genes with high *d*-scores, but large fluctuations for low *d*-scores, which we fitted with a smoothing function (Figure 2C) (12). The 10% window generated well-behaved estimates for LFDR. However, the 10% window cannot estimate the LFDR for genes with the most extreme *d*-scores, since the window around each gene must include 10% of all probe sets. We concluded that a combination of windows would be the best way to obtain accurate estimates of LFDR for the largest number genes. For example, a 10% window could be used to estimate LFDR for probe sets with *d*-scores in the middle two quartiles. Then, a 1% window could estimate LFDR for probe sets with more extreme *d*-values.

The relationship between LFDR and *q*-value differs among experiments

A *q*-value and a LFDR can be assigned to each gene in a microarray experiment (2). The *q*-value for gene *(i)* equals the FDR for the set of top-ranked genes up to rank *r(i)*. In fact, the *q*-value equals the average LFDR for the set of top-ranked genes. LFDR is usually greater than the *q*-value, since the LFDR characterizes genes with ranks near gene *(i)*, while *q*-value reflects gene *(i)* plus higher-ranked genes.

To validate our estimates for LFDR, we confirmed that estimated *q*-values equaled the average LFDR for the set of top-ranked genes associated with each *q*-value (Figure 3A). To examine the relationship between LFDR and *q*-value, we used paired SAM to analyze responses to UV and IR in lymphoblastoid cells (Figure 3B and C, respectively), and multi-class SAM to analyze gene expression in lung epithelial cells from current, former and never smokers (Figure 3D). As expected, LFDR assumed values up to 100%, which indicates certainty that a gene remained unchanged in expression. By contrast, *q*-value remained less than 80%, since *q*-value is limited by π_0, the fraction of genes with unchanged expression.

Differences between LFDR and *q*-value were substantial and varied among experiments. Genes with *q*-values of 10% had LFDR values of 37, 34 and 28% in the data for responses to UV, IR and tobacco smoke, respectively (Figure 3). Despite a one-to-one relationship between *q*-value and LFDR in a given experiment, the relationship changes from one experiment to the next.

These results illustrate the utility of LFDR for estimating the likelihood that a specific gene has been falsely identified. A user might harbor a false sense of confidence in a gene with a *q*-value of 10%. However, in the UV response experiment, the same gene had a LFDR of 37%, which indicates the true likelihood for falsely identifying that particular gene. The user might account for the difference between *q*-value and LFDR by multiplying the *q*-value for each gene by a 3.7-fold correction factor. But the correction factor would fail, since the

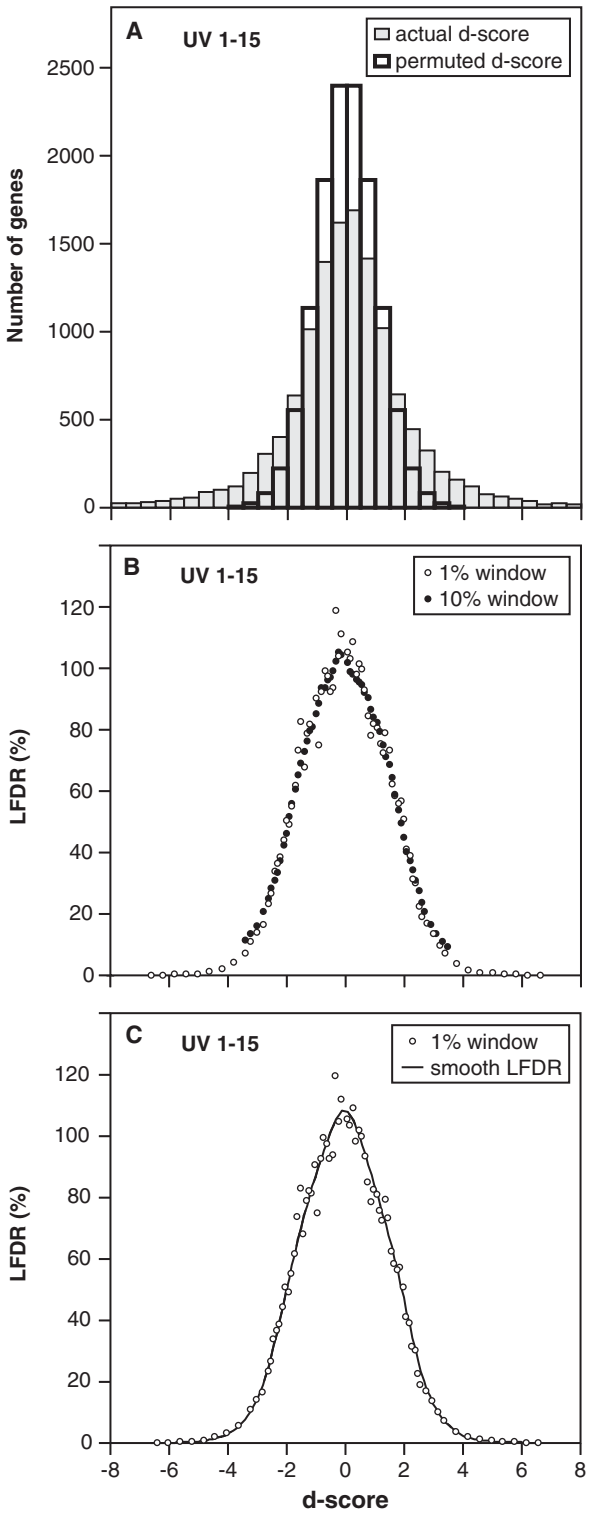


Figure 2. LFDR identified genes responding and genes not responding to UV radiation. (**A**) Distribution of *d*-scores for UV responses. To determine transcriptional responses to UV, paired SAM computed *d*-scores from microarray data for 12 625 probe sets (representing ~10 000 genes) in 15 lymphoblastoid cell lines. Grey bars show the distribution of actual d-scores for all probe sets, except for 76 probe sets with *d*-score $> +8.0$, and 76 probe sets with *d*-score < -8.0. White bars outlined in black show the distribution of permuted *d*-scores, generated by 500 random permutations. (**B**) Estimating LFDR for probe sets with low *d*-scores or extreme *d*-scores required different window sizes. We estimated the LFDR as a function of *d*-score for selected probe sets using window sizes of 1% (open circles) and 10% (solid circles). Since the microarray contained 12 625 probe sets, window sizes of 1 and 10% corresponded to 126 and 1262 probe sets, respectively. (**C**) A smoothing function estimated the LFDR for all probe sets. We generated a continuous curve (solid line) by fitting a smoothing function to the LFDR for selected probe sets, as estimated with a 1% window (open circles).

relationship between *q*-value and LFDR is non-linear (Figure 3). Furthermore, a *q*-value of 10% could be associated with a different LFDR in a different experiment, since the *q*-value depends on the top-ranked genes in each experiment. In the tobacco smoke response, a gene with a *q*-value of 10% had a LFDR of 28%, significantly different from the values of 37 and 34% in the UV and IR response experiments, respectively. Thus, if the user wishes to focus on a specific gene, the LFDR provides the best estimate for the likelihood that the gene was falsely identified.

LFDR identifies genes without changes more accurately if they change expression in another experiment

Biological experiments produce changes in expression for some, but not all genes. Genes that fail to respond to one perturbation may be of great interest, especially if they respond to a different perturbation. When we compared the UV and IR responses, a total of 3490 probe sets had LFDR $>90\%$ after both UV and IR. As expected, most of the probe sets showed unchanged expression, with similar distributions for fold-change in expression (Figure 4A and B). Changes >1.2-fold occurred in only 12% of the UV responses and 9.5% of the IR responses. These results confirm the validity of LFDR $>90\%$ as a useful criterion for identifying genes without changes in expression.

Next, we examined genes that failed to change after IR, but changed after UV. We plotted the distribution of probe sets as a function of fold-change for the 437 probe sets with LFDR $>90\%$ after IR, and LFDR $<10\%$ after UV. Strikingly, only 1 of 437 probe sets changed expression more than 1.2-fold after IR (Figure 4C). As expected, almost all of the probe sets showed changes in expression after UV (Figure 4D). Thus, virtually all genes with LFDR $>90\%$ after IR failed to change expression, if they changed in another experiment (i.e. LFDR $<10\%$ after UV).

The number of samples influences the comparison of UV and IR responses

We previously compared the UV and IR responses in human cells using the FDR estimated by SAM (10). FDR identified a large number of genes induced by both agents, but also identified many genes with a clear response to one agent but an ambiguous response to the other. LFDR provided an opportunity to compare the UV and IR responses quantitatively.

To establish a baseline for variations in the UV response, we compared the UV responses in cells from individuals 1–7 and 8–14 (Figure 5A). The table shows

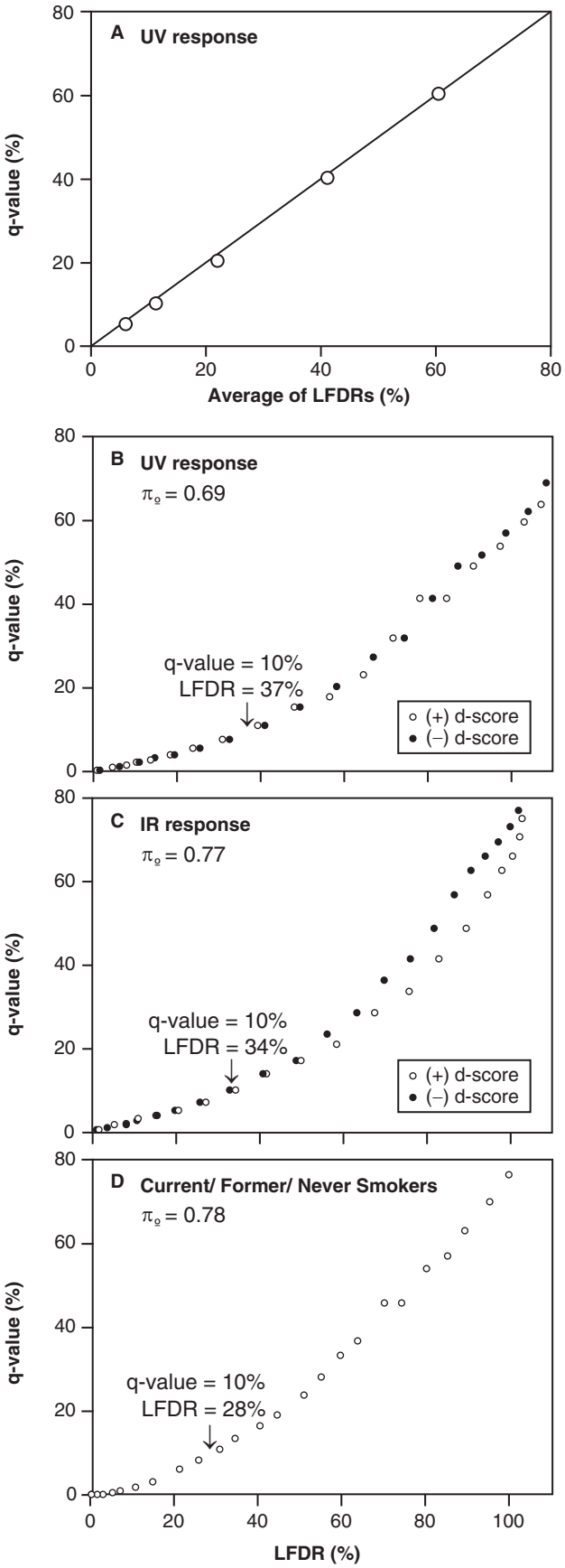


Figure 3. Relationship between LFDR and *q*-value differed among different experiments. (**A**) The *q*-value of the probe set ranked *r(i)* equaled the average LFDR for probe sets ranked *r(i)* or higher. To confirm our protocol for estimating LFDR, we calculated the average LFDR for probe sets ranked *r(i)* or higher, and plotted it against the *q*-value for the probe set ranked *r(i)*. There was a close correspondence between *q*-value and average LFDR, confirming the protocol for estimating LFDR. (**B**) LFDR versus *q*-value for selected UV-responsive probe sets. Open circles represent probe sets with positive *d*-scores, and solid circles represent probe sets with negative *d*-scores. The arrow indicates a probe set with *q*-value = 10% and LFDR = 37%. (**C**) LFDR versus *q*-value for IR-responsive probe sets. The arrow indicates a probe set with *q*-value = 10% and LFDR = 34%, which differed from the result in (B). (**D**) LFDR versus *q*-value for responses in lung epithelial cells from current, former, and never smokers. The arrow indicates a probe set with *q*-value = 10% and LFDR = 28%, which differed markedly from the results in (B) and (C).

responses in both groups for 163 probe sets (LFDR $<10\%$) and no change in both groups for 3295 probe sets (LFDR $>90\%$). The Venn diagram and table show that LFDR identified 71 discordant responses: i.e. a change in expression (LFDR $<10\%$) in one group of individuals, and unchanged expression in the other group (LFDR $>90\%$). For example, 479 probe sets responded to UV in individuals 1–7, but 39 of the 479 probe sets remained unchanged in individuals 8–14. In addition, 277 of the 479 UV-responsive probe sets for individuals 1–7 had poorly defined responses ($10\% <$ LFDR $<90\%$) in individuals 8–14. These discordant and poorly defined responses could be due to anomalies in pre-processing the raw data (as discussed below), changes obscured by noise, or heterogeneous responses in the human population.

We also compared the IR responses for individuals 1–7 and 8–14 (Figure 5B). LFDR detected changes in both groups for 33 probe sets, absence of changes in both groups for 5419 probe sets, and discordant responses for 16 probe sets. The IR response included a significantly smaller fraction of the genome than the UV response.

When we compared the UV response to the IR response for both sets of individuals 1–7 and 8–14, the UV response was more extensive than the IR response. LFDR identified 32 concordant and 153 discordant responses for individuals 1–7 (Figure 5C). LFDR identified 46 concordant and 218 discordant responses for individuals 8–14 (Figure 5D). Thus, relatively small sample sizes of seven individuals produced somewhat different results from different sets of individuals, and identified relatively few concordant responses.

When the number of individuals increased from 7 to 15, LFDR identified many more responses (Figure 5E). LFDR identified 366 probe sets with responses to both UV and IR, about 10-fold more than the number identified from seven individuals. This striking 10-fold increase occurred because estimates for the *d*-score became much more accurate, even though the number of samples increased by only 2.1-fold. LFDR also identified many more discordant responses, particularly responses to UV but not IR. For each of the comparisons of the UV and IR responses (Figure 5C–E), the number of genes responding to UV was significantly larger than the number of genes responding to IR.

The scatter plot provided a quantitative comparison of the UV and IR responses (Figure 6A). The scatter plot

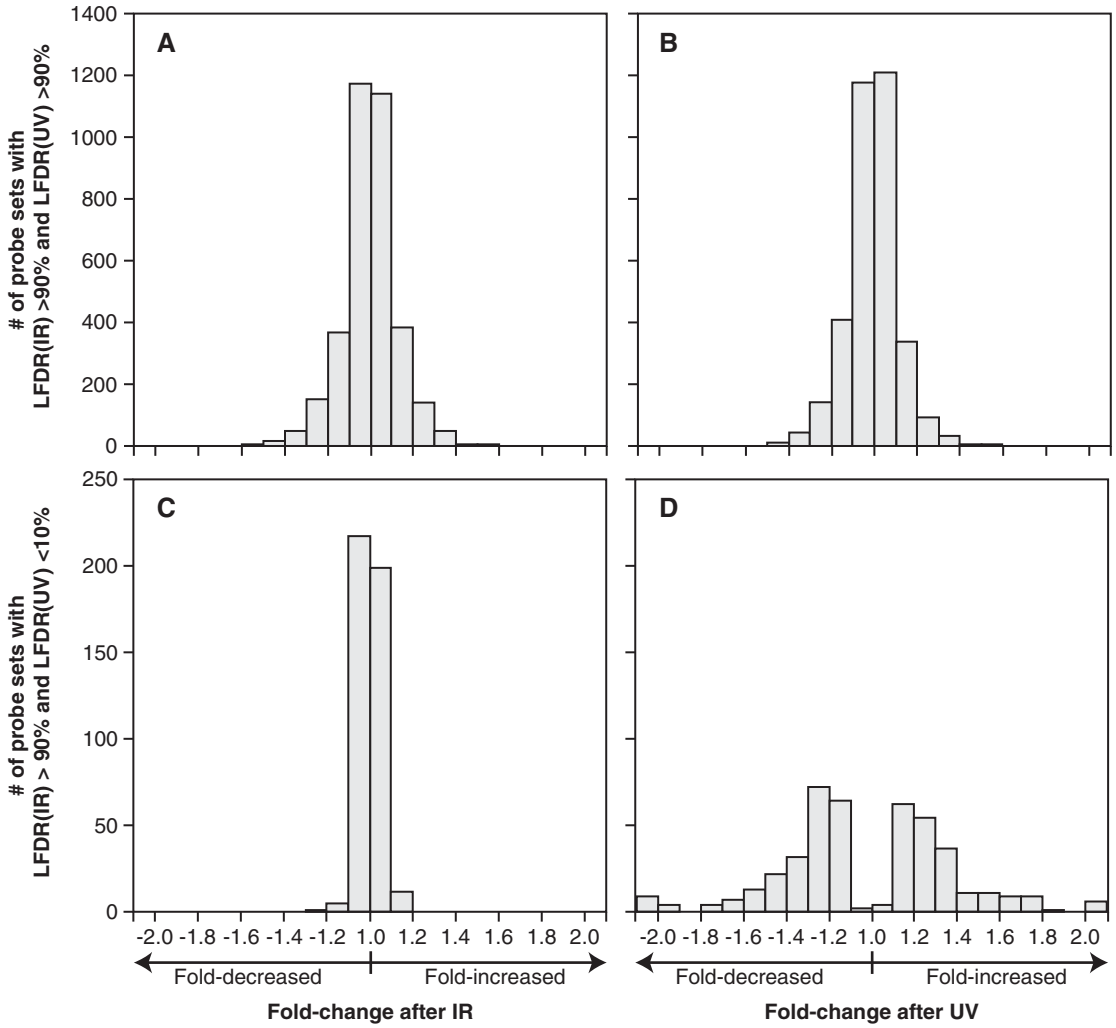


Figure 4. LFDR identified genes that did not change expression. To examine the utility of LFDR in identifying genes without changes in expression, we plotted fold-change distributions for the 3490 probe sets not responding to either UV or IR (LFDR >90%), and for the 437 probe sets responding to UV (LFDR <10%), but not IR (LFDR >90%). (**A**) IR-induced changes for probe sets predicted to not respond to UV or IR (LFDR >90%). IR-induced fold changes >1.2-fold occurred in only 9.5% of the probe sets. (**B**) UV-induced changes in probe sets predicted to not respond to UV or IR (LFDR >90%). UV-induced fold changes >1.2-fold occurred in only 12% of the probe sets. (**C**) IR-induced changes in probe sets predicted to respond to UV (LFDR <10%), but not IR (LFDR >90%). IR-induced fold changes >1.2-fold occurred in only 1 or 437 probe sets, demonstrating improved accuracy compared to (A) in identifying probe sets without changes for those probe sets that detect changes in a different experiment. (**D**) UV-induced changes in probe sets predicted to respond to UV (LFDR <10%), but not IR (LFDR >90%). UV-induced fold changes >1.1-fold occurred in more than 99% of the 437 probe sets, as expected.

included the 911 probe sets from the Venn diagram in Figure 5E. By restricting the scatter plot to this subset of genes with well-defined responses, we eliminated the confounding influence of genes with poorly defined responses (10% < LFDR < 90%). The scatter plot shows graphically that the doses of UV and IR (10 J/m^2 and 5 Gy, respectively) produced roughly equivalent fold-changes in the 366 genes responding to both UV and IR. Despite biological equivalence for the UV/IR-responsive genes, a large number of genes (437) responded to UV but not IR. Thus, UV affects a much larger fraction of the genome than IR.

For individual genes, fold-changes in the UV and IR responses were often different. In addition, some probe sets were induced by one agent but repressed by the other. These included probe sets for genes such as insulin-like growth factor 1 (*IGF1*), absent in melanoma 2 (*AIM2*), and RNA binding motif protein 14 (*RBM14*).

To compare responses in different experiments with a single parameter, we used the Pearson correlation coefficient (Figure 6B). We performed the calculation using the logarithm of fold-change in expression so that the small number of genes with large changes would not dominate the correlation coefficient. We also restricted the calculation to the well-defined responses in order to eliminate the confounding effects of noise in the data.

To calibrate the correlation coefficients generated by microarray data, we calculated the correlation between responses to the same agent. Using the responses in the Venn diagrams of Figure 5A and B, the correlation

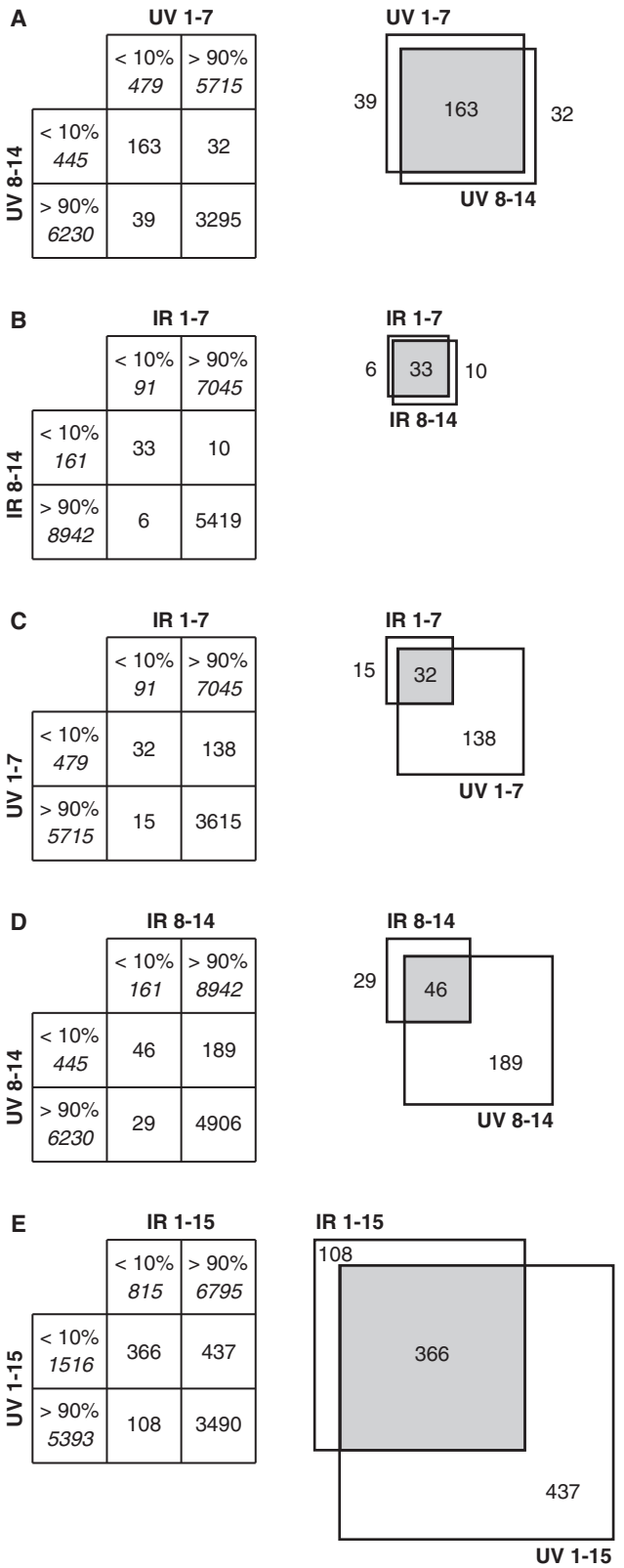


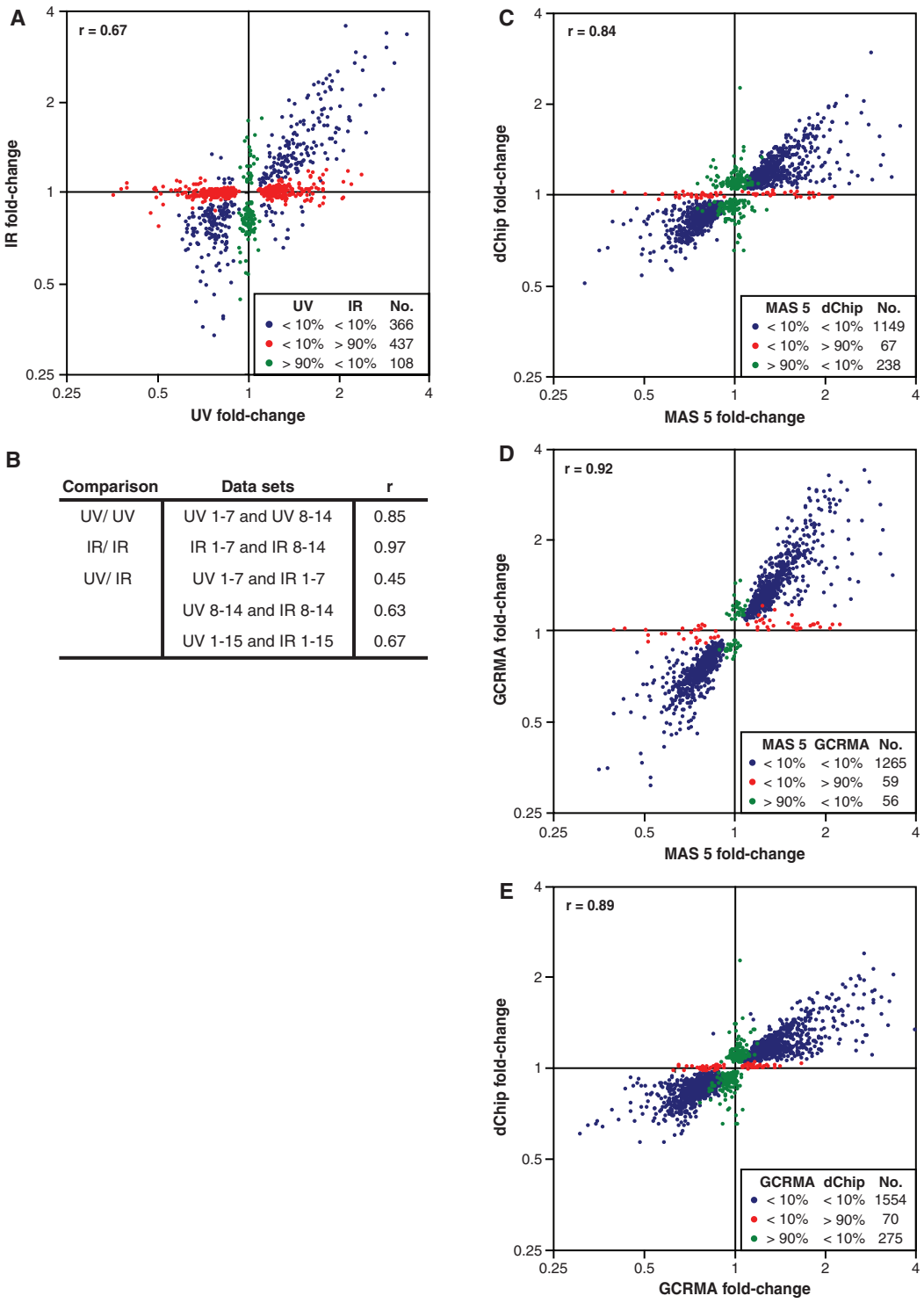
Figure 5. LFDR facilitated comparison of different experiments. Bold type headings above and to the left of the tables indicate experiments undergoing comparison. Row and column headings show the number of probe sets with the indicated LFDR (< 10% or > 90%) for the corresponding experiment. Table entries show the number of probe sets with the indicated LFDR in each experiment undergoing comparison. Venn diagrams display the number of probe sets with LFDR < 10% in both experiments (gray), and the number of probe sets with LFDR < 10% in one experiment and LFDR > 90% in the other (white). (**A**) UV responses in two sets of individuals showed general concordance. We used LFDR to compare the UV responses in cell lines 1–7 and cell lines 8–14. (**B**) IR responses in two sets of individuals showed general concordance. We used LFDR to compare the IR responses in cell lines 1–7 and cell lines 8–14. The scope of the IR response was much smaller than the UV response. (**C**) UV and IR responses in cell lines 1–7 showed significant discordance. The data generated only 32 concordant responses, but revealed 153 discordant responses. (**D**) UV and IR responses in cell lines 8–14 showed significant discordance. The concordance and discordance were similar to that seen for cell lines 1–7, but the Venn diagram included a larger number of genes. The difference in Venn diagrams indicates the level of uncertainty in using LFDR to compare different experiments using LFDR. (**E**) UV and IR responses in cell lines 1 = 15 showed significant concordance and discordance. When the sample size increased from 7 to 15 cell lines, the number of discordant responses increased from 2.3- to 3.7-fold, while the number of concordant responses increased 8-fold.

between UV 1–7 and UV 8–14 was $r = 0.85$, and the correlation between IR 1–7 and IR 8–14 was $r = 0.97$ (Figure 6B). High correlation coefficients were expected, since responses to the same agent should be highly correlated between two groups of individuals.

Next, we calculated the correlation coefficients between responses to different agents, UV and IR. The correlation between UV 1–7 and IR 1–7 was only $r = 0.45$ (Figure 6B). To determine the reproducibility of this result in a separate experiment, we calculated the correlation between UV and IR in a different group of individuals. The correlation between UV 8–14 and IR 8–14 was $r = 0.63$ (Figure 6B). The discrepancy between correlation coefficients illustrated the uncertainties arising from sample sizes of only seven individuals. Small sample sizes are sensitive to both experimental noise in the data and biological variations among individuals.

To compare the UV and IR responses more precisely, we included data for all 15 individuals. When we restricted the calculation to the well-defined responses in the Venn diagrams of Figure 5E, the correlation coefficient between the UV and IR responses increased to $r = 0.67$ (Figure 6B). Nevertheless, the correlation coefficient remained significantly less than unity, since a sizable number of genes responded to UV but not IR, and many genes responded to UV and IR with different fold-changes. Thus, the correlation coefficient provided a useful parameter that reflected the moderate degree of similarity between the UV and IR responses.

Finally, we tested the effect of expanding our calculations to data with poorly defined responses. Instead of focusing on the 911 well-defined responses in the Venn diagram of Figure 5E, we expanded the calculation to all 12 625 probe sets on the microarray. The correlation coefficient between the UV and IR responses in all 15 individuals was $r = 0.50$. Thus, the correlation coefficient decreased significantly when the calculation expanded to include poorly defined responses. Much of this decrease in correlation may be due to experimental noise in the data. Therefore, the calculated correlation coefficient appears to better reflect the true biological correlation when it includes only the well-defined responses, and excludes



B

Comparison	Data sets	r
UV/ UV	UV 1-7 and UV 8-14	0.85
IR/ IR	IR 1-7 and IR 8-14	0.97
UV/ IR	UV 1-7 and IR 1-7	0.45
	UV 8-14 and IR 8-14	0.63
	UV 1-15 and IR 1-15	0.67

Figure 6. Scatter plots facilitated graphic comparison of different experiments. (**A**) Scatter plot displayed differences between the UV and IR responses. The scatter plot displays the fold-changes for the probe sets from the Venn diagram in Figure 5E, on a logarithmic scale. The legend shows the numbers of probe sets (No.) that responded to both UV and IR (blue), to UV but not IR (red) and to IR but not UV (green). The correlation coefficient (r) for the probe sets appears in the upper left corner. (**B**) Correlation coefficients from scatter plots quantified the concordance of different data sets. We generated scatter plots for different pairs of UV and IR data sets (left and middle columns), and calculated the corresponding correlation coefficients (r) (right column). (**C**) Scatter plot compared the dChip and MAS 5.0 methods for pre-processing microarray data. Genes induced by UV according to one method were never repressed according to the other method. However, some genes responding to UV according to one method (LFDR $<10\%$) failed to respond according to the other method (LFDR $>90\%$). Furthermore, MAS 5.0 tended to generate larger fold-changes than dChip. (**D**) Scatter plot compared the GCRMA and MAS 5.0 methods for pre-processing microarray data. The concordance between these two methods ($r = 0.92$) was higher than the concordance between dChip and MAS 5.0 ($r = 0.84$). Furthermore, the two methods tended to generate roughly equal fold-changes. (**E**) Scatter plot compared the dChip and GCRMA methods for pre-processing microarray data. A single gene induced by UV according to dChip was repressed according to GCRMA. Furthermore, GCRMA tended to generate larger fold-changes than dChip. On the other hand, the concordance between these two methods ($r = 0.89$) was almost as high as the concordance between GCRMA and MAS 5.0.

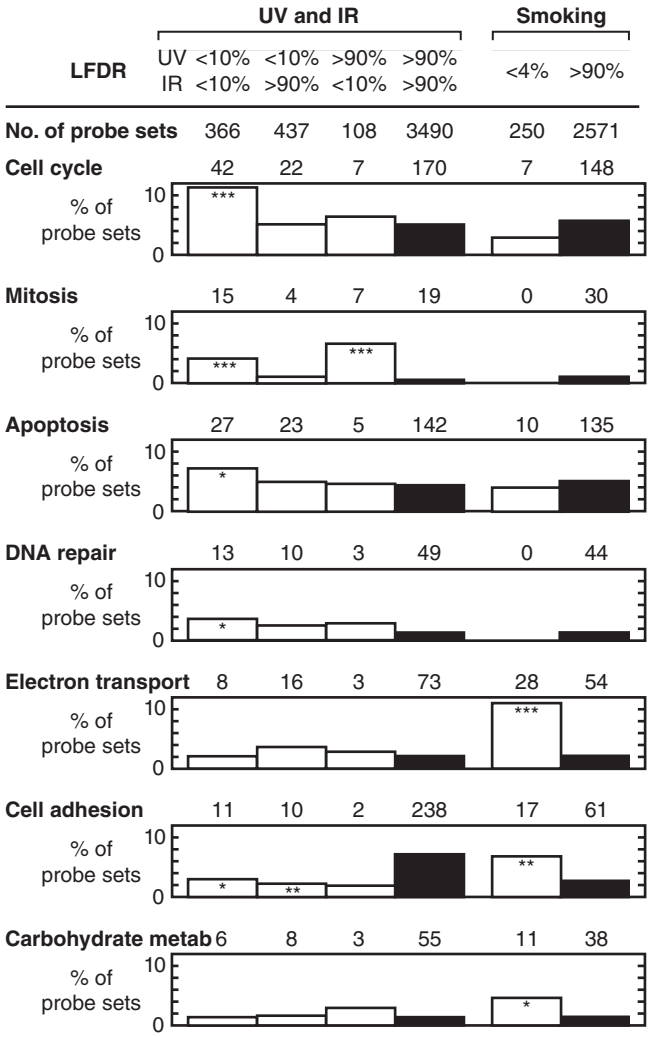


Figure 7. LFDR facilitated comparison of different responses according to gene function. Probe sets were assigned to six categories according to: whether they responded (LFDR $<10\%$) or failed to respond (LFDR $>90\%$) to UV and IR; or whether they differed (LFDR $<4\%$) or failed to differ (LFDR $>90\%$) among smokers, former smokers and never smokers. The number of probe sets in each LFDR category appears below the category. Some GO (Gene Ontology) functions were enriched or depleted (white bars) compared to functions in the non-responding probe sets (black bars). The percentage of probe sets with a specific function appears as the height of each bar, and the number of probe sets with that function appears above each bar. Asterisks denote the significance of enrichment or depletion of gene functions by Fisher's exact test ($*P < 10^{-2}$; $**P < 10^{-3}$; $***P < 10^{-4}$).

the poorly defined responses such as those defined by $10\% < \text{LFDR} < 90\%$.

LFDR permits comparison of the UV and IR responses by gene function

LFDR provided a tool for determining whether damage-responsive genes might be enriched or depleted for specific functions within the cell. To compare the UV and IR responses by gene function, we classified the UV and IR responses by GO (gene ontology) terms (Figure 7). The bar graphs depict the distribution of probe sets according to whether or not they responded to UV or IR. To account for the distribution of gene functions represented in the microarray experiment, we used a control consisting of the 3490 probe sets that failed to respond to both UV and IR (LFDR $>90\%$).

Cell cycle genes respond transcriptionally to DNA damaging agents. Among the 3490 non-responsive probe sets serving as the control, 170 (4.9%) had functions in the cell cycle. Among the 437 probe sets responding to UV, but not IR, 22 probe sets (5.0%) had functions in the cell cycle, a distribution nearly identical to the control. Among the 108 probe sets responding to IR, but not UV, seven probe sets (6.5%) involved the cell cycle, a distribution not statistically different from the control ($P = 0.37$). Thus, we failed to detect any enrichment in cell cycle genes responding exclusively to UV or IR. In contrast, among the 366 probe sets responding to both UV and IR, 42 probe sets (11.5%) had functions in the cell cycle. Thus, the probe sets responding to both forms of DNA damage were significantly enriched ($P = 2.6 \times 10^{-6}$) for cell cycle genes.

Genes responding to both UV and IR were also enriched for functions in mitosis ($P < 10^{-5}$), apoptosis ($P < 10^{-2}$), and DNA repair ($P < 10^{-2}$), but depleted for functions in cell adhesion ($P < 10^{-2}$) (Figure 7). Genes responding to UV but not IR were depleted for functions in cell adhesion, and genes responding to IR but not UV were enriched for functions in mitosis. Thus, LFDR facilitated a quantitative analysis of the distribution of cellular functions for genes responding to DNA damage.

LFDR reveals differences in methods for pre-processing oligonucleotide microarray data

Oligonucleotide microarrays measure the abundance of a gene transcript by hybridization of mRNA to multiple oligonucleotide probe pairs. Each probe pair detects the signal from a perfect matched probe (PM) and a mismatched probe (MM). The use of multiple probe pairs mitigates the effect of aberrant hybridizations that might otherwise generate incorrect estimates for some mRNA levels.

Several methods pre-process the data from multiple probe pairs to estimate the mRNA level for each gene. We examined three methods: MAS 5.0, GCRMA and dChip. MAS 5.0 (Microarray Suite version 5.0) estimates the expression of each gene by computing differences (PM–MM) for each probe pair. For a given gene, probe pairs often generate PM–MM differences that diverge markedly from each other. These differences may be due to cross-hybridization from other genes, or to manufacturing biases. MAS 5.0 addresses this problem by discarding differences that diverge beyond a cut-off, and averaging the remaining differences (15).

DNA-Chip analyzer (dChip) exploits the observation that a specific probe generates highly reproducible data, even when multiple probes for a single mRNA generate data that diverge from each other. The dChip algorithm generates a model for the behavior of each probe from all microarrays in a given experiment, and then generates a model to estimate the expression of each gene (16).

The model includes methods for handling cross-hybridizing probes and contaminated regions on the microarray.

GCRMA (GC-robust multi-array analysis) uses a normalization algorithm that includes compensation for the GC content of the oligonucleotides. Gene expression is estimated using the One-Step Tukey's Biweight Estimate, which yields a weighted mean that is relatively insensitive to outliers (17,18).

To compare the three pre-processing methods, we applied LFDR to the UV and IR responses. The methods identified UV-responsive genes with different efficiencies (Figure 8A). GCRMA and dChip identified the largest number of genes with changes in expression (LFDR $<10\%$): 2063 and 2259 probe sets, respectively. In contrast, MAS 5.0 identified only 1516 probe sets. The methods also identified IR-responsive genes with different efficiencies. GCRMA and dChip identified 1146 and 1337 probe sets, respectively, while MAS 5.0 was again less effective, identifying only 815 probe sets. Thus, MAS 5.0 consistently identified fewer responsive genes than either GCRMA or dChip.

We compared the genes identified by the different pre-processing methods. MAS 5.0, GCRMA and dChip showed significant concordance in identifying UV-responsive genes (Figure 8B–D). However, some genes yielded discordant results when analyzed by the three methods, as revealed by scatter plots of fold-change in expression (Figure 6C–E). The differences were biased: dChip tended to report smaller fold-changes than GCRMA or MAS 5.0 (Figure 6C and E). On the other hand, the methods rarely reported opposite changes in expression: a single gene was induced according to dChip, but repressed according to GCRMA (Figure 6E).

Despite overall concordance among pre-processing methods, significant numbers of genes reported as changing by one method (LFDR $<10\%$) were reported as not changing by another method (LFDR $>90\%$). For example, of the 2259 genes identified responsive to UV according to dChip, 11% (238 genes) were not responsive according to MAS 5.0, and 12% (275 genes) were not responsive according to GCRMA (Figure 8B and D). Discordant identification of genes as changing versus not changing also occurred between GCRMA and MAS 5.0 (Figure 8C). Thus, caution must be exercised in interpreting the results for a specific gene.

The correlation coefficients for the different methods were:

$r = 0.84$ for dChip versus MAS 5.0;
$r = 0.92$ for GCRMA versus MAS 5.0;
$r = 0.89$ for dChip versus GCRMA.

The correlations among the three methods were high enough to provide confidence that any one of the methods provided reasonable data for comparing the UV and IR responses. However, note that these correlations were no higher than the IR/IR and UV/UV correlations (Figure 6B). This was unexpected, since the correlations among the different pre-processing methods were based on a single set of raw UV response data, while the IR/IR and UV/UV correlations were based on two distinct sets of data from different sets of individuals. Therefore, our analysis provides further evidence that pre-processing methods may affect the apparent outcomes of microarray experiments.

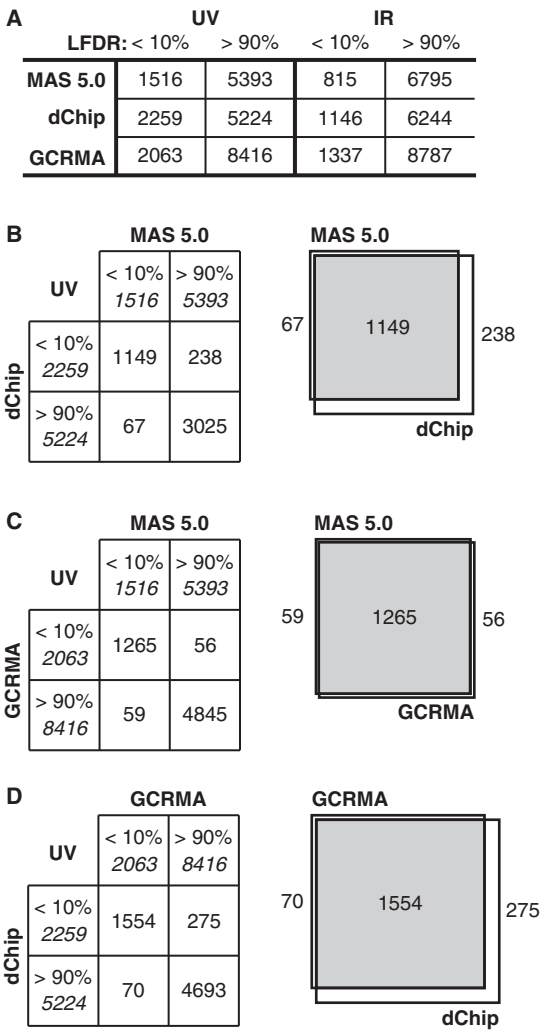


Figure 8. Different methods for pre-processing microarray data generated different results. (**A**) Different methods identified different numbers of genes that changed or failed to change after UV or IR. We estimated expression levels from raw oligonucleotide microarray data using three pre-processing methods: MAS 5.0, GCRMA and dChip. SAM identified genes that changed (LFDR $<10\%$) or remained unchanged (LFDR $>90\%$) after cells were exposed to UV or IR. (**B**) Comparison of dChip versus MAS 5.0. The table and Venn diagram show the concordant and discordant UV responses identified from data pre-processed by dChip and MAS 5.0. (**C**) Comparison of GCRMA versus MAS 5.0. The table and Venn diagram show fewer discordant UV responses between these two methods, when compared to dChip versus MAS 5.0. (**D**) Comparison of dChip versus GCRMA. The table and Venn diagram show the largest number of concordant UV responses (1554 probe sets) between these two methods, when compared to dChip versus MAS 5.0, or to GCRMA versus MAS 5.0.

LFDR reveals changes in gene expression from smoking exposure

To illustrate the simultaneous application of LFDR to data from a three-armed experiment, we analyzed

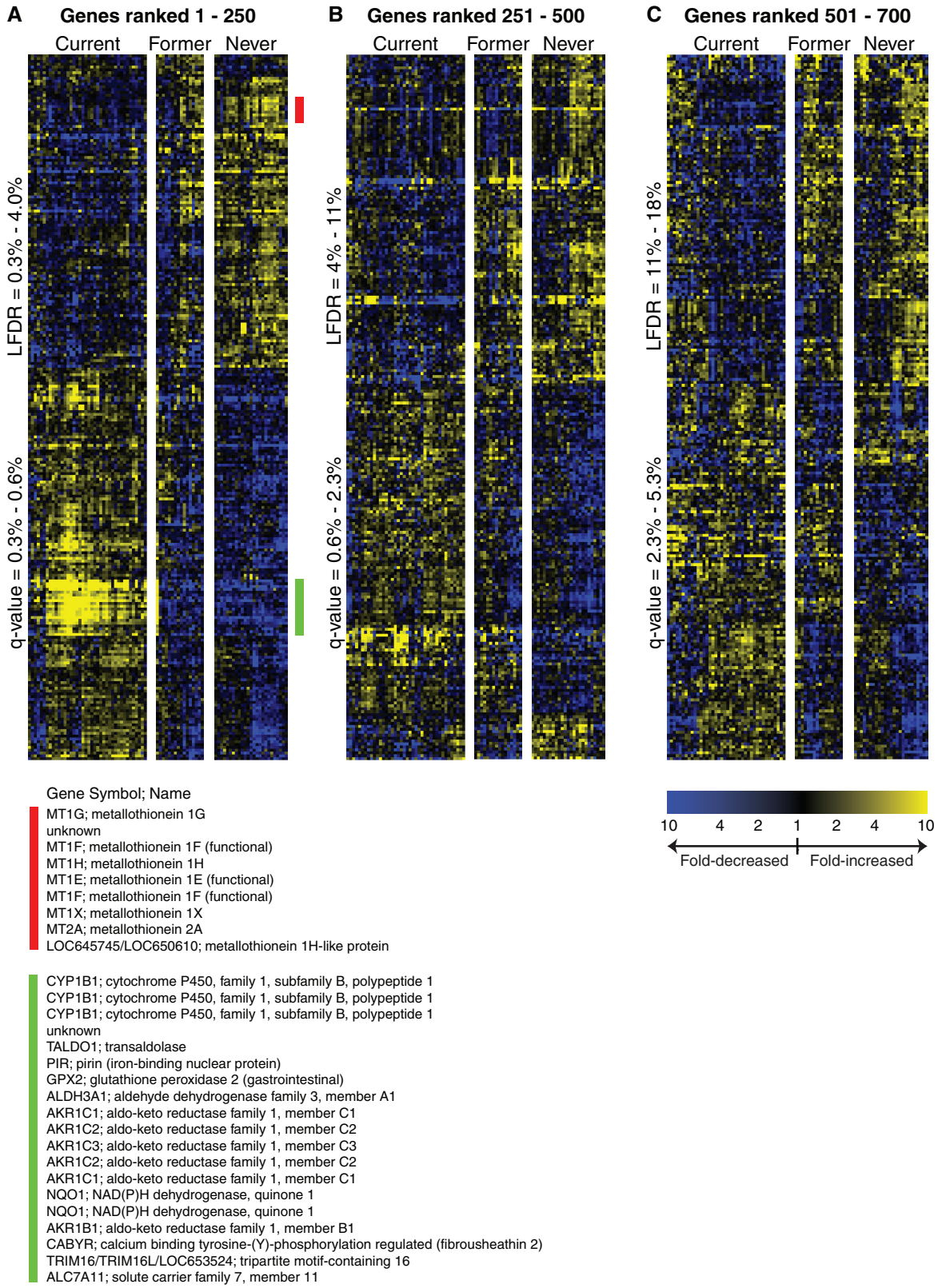


Figure 9. LFDR facilitated comparison of three sample classes defined by smoking history. Probe sets measured gene expression in epithelial cells obtained by bronchoscopy from three sample classes: 40 current smokers; 16 former smokers; and 25 individuals who had never smoked. We used multi-class SAM to identify the top-ranked 750 probe sets associated with the three classes (current, former and never smokers), and analyzed the probe sets by hierarchical clustering of the probe sets ranked: (**A**) 1–250, (**B**) 251–500 and (**C**) 501–750. Hierarchical clustering was performed separately for samples in each class. Yellow represents increased expression and blue represents decreased expression relative to the median for each probe set, as quantified in the color scale (lower right). The ranges for *q*-value and LFDR are shown to the left of each heat map. Current and never smokers showed large differences in clusters for metallothionein genes (red bar) and for genes involved in detoxification of xenobiotics (green bar).

microarray data from the Smoking Induced Epithelial Gene Expression (SIEGE) database (13). At the time of this analysis, the SIEGE database contained data from 40 current smokers, 13 former smokers and 25 never smokers.

We used multi-class SAM, which generates the *d*-score from Fisher's discriminant. For gene (*i*), the d-score is given by $d(i) = r(i)/[s(i) + s_0]$, where

$$r(i) = \left\{ \left[\frac{\sum_k n_k}{\prod_k n_k} \right] \sum_k n_k [\overline{x}_k(i) - \overline{x}(i)]^2 \right\}^{1/2} \quad \mathbf{2}$$

Using these multi-class *d*-scores, we estimated the LFDR and q-value for each probe set. A total of 750 probe sets had a *q*-value $\leq$5%. However, the LFDR for these probe sets rose to values as high as 18%, confirming the observation that a given *q*-value may correspond to a much higher LFDR.

To examine the utility of the top-ranked probe sets relative to less highly-ranked probe sets, we applied hierarchical clustering to the probe sets ranked 1–250, ranked 251–500, and ranked 501–750 (Figure 8). For the probe sets ranked 1–250, the *q*-value ranged from 0.3 to 0.6% and the LFDR ranged from 0.3 to 4.0% (Figure 9A). The heat map successfully distinguished current from never smokers. Former smokers did not appear as a distinct class, but instead resembled either current or never smokers. When we used two-class SAM to search specifically for genes that distinguished former smokers from current and never smokers, we identified only 18 probe sets with a *q*-value of 5%, and failed to find any probe sets with a reliable LFDR.

For the probe sets ranked 251–500, the distinction between current and never smokers was still apparent, but not as clear (Figure 9B). For the probe sets ranked 501–750, the distinction among the classes was barely discernable, even though the *q*-values were $<5\%$ (Figure 9C). Therefore, LFDR was more useful than *q*-value for identifying genes associated with smoking exposure.

We examined functions of the genes that correlated most strongly with smoking status. The genes associated with the 250 top-ranked probe sets (LFDR $<4\%$) included a cluster of genes strongly expressed in current smokers, but not in former or never smokers (Figure 9, green bar). This cluster included 18 probe sets with known functions in the detoxification of xenobiotics: *CYP1B1* (a cytochrome P450 family member), *GPX2* (glutathione peroxidase 2), and several aldo-keto reductase family members. Many of the genes in this cluster have previously been reported to have functions in metabolizing the toxins in tobacco smoke (19–23).

A second cluster was strongly repressed in current smokers compared to never smokers, and showed variable responses in former smokers (Figure 9, red bar). This cluster included seven probe sets for metallothionein genes and one probe set for an unknown gene. In previous studies, acute exposure to tobacco smoke increased the expression of metallothionein genes (24). This finding raises the possibility that smoking might first induce pathways for metabolizing heavy metals, and then suppress the pathways after chronic exposure.

Genes associated with smoking were distinct from genes that responded to UV and IR. LFDR facilitated such comparisons, even though different laboratories collected the data on different microarray platforms (U95A versus U133A, Affymetrix). Of the 21 unique probe sets in the detoxification and metallothionein clusters, only six probe sets responded to either UV or IR, while 12 probe sets failed to respond to both UV and IR. Smoking-related genes were enriched for functions in electron transport, cell adhesion, and carbohydrate metabolism, but not for functions characteristic of the IR and UV responses, such as cell cycle, mitosis, apoptosis, or DNA repair (Figure 7). Although cell adhesion genes were enriched among smoking responses, they were depleted among UV and IR responses. Thus, genes associated with smoking differed significantly from genes responsive to UV or IR.

The distinct responses to smoking may have been due to several factors. Responses were measured in different cell types: lung epithelial cells for smoking exposure and lymphoblastoid cells for UV and IR. Responses were measured at different time points: during chronic exposure to tobacco smoke and only hours after a single UV or IR dose. Nevertheless, many of the distinct responses to smoking are likely due to the different effects of tobacco, UV and IR.

DISCUSSION

LFDR can be used to identify genes that change, or fail to change, in response to a biological perturbation. Computer simulation demonstrated that LFDR accurately identifies such genes. LFDR $<10\%$ identified 96% of genes with changes equal to twice the standard deviation of noise in the data. Conversely, LFDR $>90\%$ identified 84% of genes with no change.

An important limitation is that LFDR requires data from enough probe sets to permit a reliable estimate. Here, we estimated the LFDR from 12 625 probe sets. A 10% window used 1260 probe sets, which produced stable estimates for LFDR, but excluded the 10% of probe sets with the most extreme *d*-scores (Figure 1). On the other hand, a 1% window generated estimates for 99% of the probe sets, but the estimates fluctuated for genes with low *d*-scores. To ameliorate these problems, one can apply different size windows, depending on *d*-score.

A second limitation is that it is difficult to estimate LFDR for probe sets with the most extreme *d*-scores. For example, in a data set with 10 000 probes, even a 1% window fails to generate a LFDR for the 100 probe sets with the most extreme *d*-scores. To address this problem, one can use the LFDR from genes with less extreme *d*-scores to provide an upper limit for the LFDR.

The LFDR for a given gene depends on genes with similar *d*-scores. In contrast, the *q*-value for a given gene depends on the set of all genes with more extreme *d*-scores in each specific experiment. Although there is a one-to-one relationship between *q*-value and LFDR, the relationship varied among different experiments (Figure 3B–D): genes

with a *q*-value of 10% could be interpreted with an unwarranted level of confidence, since LFDR for the gene was 28–37%, depending on the experiment. Thus, LFDR provides a more direct estimate for the likelihood that a gene has changed expression.

Here, we report that LFDR >90% can identify genes that fail to change expression. However, LFDR >90% may falsely identify a gene as not changing due to a high level of noise. A probe set may generate noise due to inconsistent manufacturing or cross-hybridization to other genes. Such noise should be absent if the same probe set has LFDR <10% after a different perturbation. Indeed, we were able to identify genes without changes with particularly high accuracy if those genes changed expression in another experiment (Figure 4).

Other authors have addressed the problem of analyzing microarray experiments performed on different platforms. Bayesian approaches generated meta-signatures shared by multiple datasets for chronic lymphocytic leukemia and for breast cancer (25,26). However, approaches are needed to compare as well as combine different experiments.

Here, we have shown that LFDR can identify genes with and without changes in expression, and then used several tools to compare different experiments. First, Venn diagrams displayed the *number* of probe sets with concordant and discordant responses. Second, scatter plots displayed the *magnitudes* and *directions* of responses. Third, the Pearson correlation coefficient provided a single parameter for the similarity between experiments. Finally, gene ontology compared different experiments in terms of the functions of responsive genes, and determined whether the responses to different perturbations were enriched or depleted for specific functions.

We used these tools to analyze three methods for pre-processing the UV response data: MAS 5.0, dChip and GCRMA. SAM identified more responsive genes with data from dChip and GCRMA. Since SAM identifies genes based on changes in expression relative to the standard deviation among samples, our results suggest that MAS 5.0 may introduce more uncertainty than dChip or GCRMA. MAS 5.0 eliminates outlier data from individual probe pairs by employing a fixed cut-off. Thus, a probe pair may influence the estimate for gene expression in one hybridization but not another, depending on whether the probe pair data exceeds the cut-off for outliers. Thus, the cut-off could generate noise by accentuating otherwise modest fluctuations in the raw data. Such phenomena may occur frequently enough to explain why MAS 5.0 identified fewer responsive genes than dChip and GCRMA. Nevertheless, data from the three pre-processing methods showed a high level of correlation with each other.

Armed with tools for comparing microarray experiments, we defined similarities and differences among responses to DNA the damaging agents, UV, IR and tobacco smoke. The UV response included significantly more genes than the IR response. Some genes even showed discordant responses between UV and IR.

Tools for comparing experiments provided insights in terms of gene function. For example, cell adhesion genes were depleted among the UV and IR responses, but enriched among tobacco responses. Electron transport genes were neither enriched nor depleted among the UV and IR responses, but strongly enriched among tobacco responses.

These results for gene function illustrate LFDR can complement methods such as gene set enrichment analysis (GSEA) (27). GSEA focuses on gene sets associated with a specific biological function, and determines whether a specific gene set undergoes a coordinated change in gene expression. For example, GSEA may report that a particular gene set is enriched for genes that increase expression, compared to genes that decrease expression in response to a perturbation. GSEA does not account for the genes that fail to change expression. LFDR analyzes gene sets from the perspective of the responsive versus unresponsive genes. Thus, LFDR provides data on whether responsive genes are enriched for a specific function, using the number of unresponsive genes as a control.

In conclusion, LFDR facilitated comparisons among a broad range of different experiments, including responses to different biological perturbations. LFDR can also be used to compare proteomic and genomic responses to the same perturbation. For example, a perturbation might induce changes in the level of a phosphorylated protein without affecting the transcription of the corresponding gene. Software for calculating LFDR is available at the SAM website http://www-stat.stanford.edu/~tibs/SAM/.

FUNDING

National Institutes of Health (grants N01-HV-28183 to G.C. and R.T.); (5RCA-111487 to Patrick O. Brown). Funding for open access charge: National Institutes of Health grants (N01-HV-28183).

Conflict of interest statement. None declared.

REFERENCES

1. Tusher,V., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
2. Storey,J.D. and Tibshirani,R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
3. Aubert,J., Bar-Hen,A., Daudin,J.J. and Robin,S. (2004) Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics*, **5**, 125.
4. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B*, **57**, 289–300.
5. Efron,B. (2004) Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Stat. Assoc.*, **99**, 96–103.
6. Liao,J.G., Lin,Y., Selvanayagam,Z.E. and Shih,W.J. (2004) A mixture model for estimating the local false discovery rate in DNA microarray analysis. *Bioinformatics*, **20**, 2694–2701.
7. Ploner,A., Calza,S., Gusnanto,A. and Pawitan,Y. (2006) Multidimensional local false discovery rate for microarray studies. *Bioinformatics*, **22**, 556–565.
8. Reiner,A., Yekutieli,D. and Benjamini,Y. (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, **19**, 368–375.

9. Scheid,S. and Spang,R. (2005) Twilight; a BioConductor package for estimating the local false discovery rate. *Bioinformatics*, **21**, 2921–2922.
10. Rieger,K.E. and Chu,G. (2004) Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res.*, **32**, 4786–4803.
11. Storey,J.D. (2002) A direct approach to false discovery rates. *J. Royal Stat. Soc. B*, **64**, 479–498.
12. Ramsay,J.O. and Silverman,B.W. (2002) *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, New York.
13. Shah,V., Sridhar,S., Beane,J., Brody,J.S. and Spira,A. (2005) SIEGE: smoking induced epithelial gene expression database. *Nucleic Acids Res.*, **33**, D573–D579.
14. Eisen,M., Spellman,P., Brown,P. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
15. Affymetrix (2001) *Statistical Algorithms Reference Guide*. Santa Clara, CA.
16. Li,C. and Wong,W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA*, **98**, 31–36.
17. Wu,Z. and Irizarry,R.A. (2005) Stochastic models inspired by hybridization theory for short oligonucleotide arrays. *J. Comput. Biol.*, **12**, 882–893.
18. Wu,Z., Irizarry,R.A., Gentleman,R., Murillo,F.M. and Spenccer,F. (2004) *A Model-based Background Adjustment for Oligonucleotide Expression Arrays*. Department of Biostatistics Working Papers, Johns Hopkins University, Baltimore, MD.
19. Gilks,C.B., Price,K., Wright,J.L. and Churg,A. (1998) Antioxidant gene expression in rat lung after exposure to cigarette smoke. *Am. J. Pathol.*, **152**, 269–278.
20. Kim,J.H., Sherman,M.E., Curriero,F.C., Guengerich,F.P., Strickland,P.T. and Sutter,T.R. (2004) Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. *Toxicol. Appl. Pharmacol.*, **199**, 210–219.
21. Nagaraj,N.S., Beckers,S., Mensah,J.K., Waigel,S., Vigneswaran,N. and Zacharias,W. (2006) Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. *Toxicol. Lett.*, **165**, 182–194.
22. Piipari,R., Savela,K., Nurminen,T., Hukkanen,J., Raunio,H., Hakkola,J., Mantyla,T., Beaune,P., Edwards,R.J., Boobis,A.R. *et al.* (2000) Expression of CYP1A1, CYP1B1 and CYP3A, and polycyclic aromatic hydrocarbon-DNA adduct formation in bronchoalveolar macrophages of smokers and non-smokers. *Int. J. Cancer*, **86**, 610–616.
23. Woenckhaus,M., Klein-Hitpass,L., Grepmeier,U., Merk,J., Pfeifer,M., Wild,P., Bettstetter,M., Wuensch,P., Blaszyk,H., Hartmann,A. *et al.* (2006) Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J. Pathol.*, **210**, 192–204.
24. Bosio,A., Knorr,C., Janssen,U., Gebel,S., Haussmann,H.J. and Muller,T. (2002) Kinetics of gene expression profiling in Swiss 3T3 cells exposed to aqueous extracts of cigarette smoke. *Carcinogenesis*, **23**, 741–748.
25. Shen,R., Ghosh,D. and Chinnaiyan,A.M. (2004) Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics*, **5**, 94.
26. Wang,J., Coombes,K.R., Highsmith,W.E., Keating,M.J. and Abruzzo,L.V. (2004) Differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells: a meta-analysis of three microarray studies. *Bioinformatics*, **20**, 3166–3178.
27. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.