

# Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets

Adaikalavan Ramasamy\*, Adrian Mondry, Chris C. Holmes, Douglas G. Altman

Microarray technology measures the mRNA levels of tens of thousands of genes in tissue samples simultaneously in a high-throughput and cost-effective manner. Since its introduction over a decade ago [1], it has found widespread use in the fields of molecular genetics and functional genomics. It has been applied in order to understand underlying biological mechanisms [2], to discover novel subgroups of diseases [3–5], to examine drug response [6,7], to classify patients into disease groups [3], and to predict disease outcomes [8–10]. Some molecular signatures discovered with microarray technology are now being evaluated in prospective randomized clinical trials [11,12].

Despite their great promise, microarray-based studies may report findings that are not reproducible [13] or not robust to the mildest of data perturbations [14,15]. Common causes include improper analysis or validation, insufficient control of false positives, and inadequate reporting of methods [16,17]. The situation is exacerbated by the small sample sizes relative to large numbers of potential predictors; typically tens of thousands of probes are investigated in only tens or hundreds of biological samples.

Generalizability across studies [18] also needs to be assessed before considering widespread practical application. For example, the findings of a study using historical controls from a particular geographical region may not be applicable to newer cohorts of patients or different regions.

Combining information from multiple existing studies can increase the reliability and generalizability of results. The use of statistical techniques to combine results from independent but related studies is called “meta-analysis.” However, the term meta-analysis is also widely used to describe the whole study process (as we do here), not just the statistical techniques, for which an alternative term is a “systematic review.” Through meta-analysis, we can increase the statistical power to obtain a more precise estimate of gene expression differentials, and assess the heterogeneity of the overall estimate. Meta-analysis is relatively inexpensive, since it makes comprehensive use of already available data.

Indeed, the advantages of meta-analysis of gene expression microarray datasets have not gone unnoticed by researchers in various fields [19–28]. Several meta-analysis techniques have been proposed in the context of microarrays [19,22,29–40]. However, no comprehensive framework exists on how to carry out a meta-analysis of microarray datasets.

There is a considerable literature to guide the whole review process, including statistical methods for clinical trials and epidemiological studies [41–43]. As yet, however, there is little guidance for conducting a meta-analysis of microarray

## Summary Points

- Improvements in microarray technology and its increasing use have led to the generation of many highly complex datasets that often try to address similar biological questions.
- Meta-analysis, a statistical approach that combines results from independent but related studies, is a relatively inexpensive option that has the potential to increase both the statistical power and generalizability of single-study analysis.
- Meta-analysis of microarray datasets, and genomic data in general, is desirable, and is much enhanced when raw data are available.
- We identify seven key issues and suggest a stepwise approach in conducting meta-analysis of microarray datasets: (1) Identify suitable microarray studies; (2) Extract the data from studies; (3) Prepare the individual datasets; (4) Annotate the individual datasets; (5) Resolve the many-to-many relationship between probes and genes; (6) Combine the study-specific estimates; (7) Analyze, present, and interpret results.
- We give practical guidance to assist those conducting or reviewing such a meta-analysis.
- The approaches presented here can be adapted to other areas of high-throughput biological data analysis.

datasets. Therefore, in this paper, we disentangle this complex topic and identify seven distinct key issues specific to meta-analysis of microarray datasets, each comprising several steps. The first five issues are related to data acquisition and

**Funding:** AR and DGA are funded by Cancer Research UK. AM is supported by Imperial College Healthcare NHS Trust. CCH is partly supported by the UK Medical Research Council and the University of Oxford. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Citation:** Ramasamy A, Mondry A, Holmes CC, Altman DG (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* 5(9): e184. doi:10.1371/journal.pmed.0050184

**Copyright:** © 2008 Ramasamy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** FDR, false discovery rate; FLEO, feature-level extraction output; GEDM, gene expression data matrix; IPD, individual patient-level data; MIAME, Minimum Information About a Microarray Experiment; PGL, published gene list

Adaikalavan Ramasamy and Douglas G. Altman are with the Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom. Adaikalavan Ramasamy and Chris C. Holmes are with the Department of Statistics, University of Oxford, Oxford, United Kingdom. Adrian Mondry is with Imperial College Healthcare NHS Trust, London, United Kingdom.

\* To whom correspondence should be addressed. E-mail: ramasamy@stats.ox.ac.uk

**Provenance:** Not commissioned; externally peer reviewed

The Guidelines and Guidance section contains advice on conducting and reporting medical research.

curation. We discuss the sixth issue—choosing a meta-analysis technique—using the two-class comparison as an example. The seventh issue of analyzing, presenting, and interpreting data is discussed briefly using an illustrative meta-analysis of 25 datasets. We provide a practical checklist, shown in Table 1, that should enable the reader to make informed decisions on how to conduct a meta-analysis, and to understand better the underlying concepts that make this approach so attractive for analysis of microarray data.

### Issue 1: Identify Suitable Microarray Datasets

The first step in any research project is to clearly define the objectives (Step 1). Meta-analysis could be used to identify genes expressed differentially between two groups [19,22,29,30,32,33,35,37,38,40], to robustify cross-platform classification [34], to identify overlaps between samples from heterologous datasets [30], to identify co-expressed genes, or to reconstruct gene networks [31,36,39].

Having a detailed review protocol can further help to clarify the research objectives and methods and to minimize

bias from unplanned data-driven analysis. We suggest developing the review protocol by outlining the solutions to the steps in the checklist shown in Table 1. For example, Step 7 (Check the selected study against inclusion-exclusion criteria) might be expanded in the review protocol as follows: “Two reviewers will check the eligibility of the identified studies, with disagreements resolved by a third reviewer. A log of excluded studies, with reasons for exclusions, will be maintained.” The protocol can be turned into a useful project management tool by incorporating timelines and division of labor.

The inclusion-exclusion criteria (Step 2) are eligibility criteria for studies that will help achieve the stated objectives. These criteria could be biological (e.g., specific disease, type of outcome, type of tissues) or technical (e.g., density of array, minimum number of arrays). The retrieved articles must be evaluated as to whether they met the inclusion criteria.

Once the inclusion-exclusion criteria have been defined, one needs to perform a comprehensive literature search (Step 3) to identify suitable studies, usually based on

**Table 1.** A Checklist for Conducting Meta-Analysis of Microarray Datasets

Step	Action
<b>Identify suitable microarray studies (Issue 1)</b>	
1	Formulate objectives and a review protocol.
2	Define inclusion-exclusion criteria and suitable keywords.
3	Perform literature search using the keywords on the Web sites listed in Table 2.
4	Search public microarray repositories listed in Table 2.
5	Contact collaborators and experts in the field to help find published and unpublished data.
6	Search the reference section of retrieved studies for other relevant studies.
7	Check the selected study against inclusion-exclusion criteria.
<b>Extract the data from studies (Issue 2)</b>	
8	Scan the literature to identify FLEO data (e.g., CEL, GPR files).
9	If the main text does not contain a link to FLEO data, search the repositories and group/lab's Web pages. If unsuccessful, write to the authors.
10	If multiple publications use overlapping data, identify the most comprehensive one. Combine any training and validation dataset together.
<b>Prepare the individual datasets (Issue 3)</b>	
11	Identify and remove any arrays with poor quality.
12	Preprocess the FLEO data into a GEDM.
13	Check for batch effects among arrays, especially in large studies.
14	Filter out any probes with poor spot quality in the arrays (optional).
15	Aggregate any technical replicates.
16	Check that the processed expression values from multiple platforms are compatible.
<b>Annotate the individual datasets (Issue 4)</b>	
17	Identify either (a) the probe sequence or (b) the most sequence-specific probe annotation information.
18	Either (a) cluster the probe sequences or (b) map the most sequence-specific probe annotation to a gene-level identifier. Use the same mapping build for all datasets.
<b>Resolve the many-to-many relationship between probes and genes (Issue 5)</b>	
19	Discard any probe that does not map to any GeneID.
20	For every GeneID within a study, calculate the study-specific estimate(s).
21	If a probe maps to multiple GeneIDs within a study, “expand” it by replacing it with a new record for each GeneID with the same study-specific estimate(s) or expression profile.
22	For GeneIDs with multiple records within a study, “summarize” them by either selecting one of the records or by aggregating them.
<b>Combine the study-specific estimates (Issue 6)</b>	
23	For every GeneID, identify the studies that provide usable information. Optionally, discard any GeneID that is not found in at least a prespecified number of studies.
24	For every GeneID, combine the study-specific estimates across the studies using a meta-analytic technique. Record the resulting summary statistic(s).
25	Calculate the nominal p-value of the summary statistic(s) for every GeneID and adjust for multiple testing.
<b>Analyze, present, and interpret results (Issue 7)</b>	
26	Examine the sensitivity of results to individual studies with a leave-one-out analysis and by varying the selections made (e.g., type of data available).
27	Present the summary statistics graphically (e.g., forest plot) for genes of interest.
28	Analyze findings using computational tools (e.g., gene set enrichment analysis).
29	If possible, validate using an alternative technology and/or different samples.
30	Consider strength of evidence, limitations, and generalizability of current findings.

GeneID refers to either the sequence cluster or gene-level identifier used in Step 18. See text for further details. Where possible, we have indicated the step number near the relevant text. doi:10.1371/journal.pmed.0050184.t001

**Table 2.** Useful Internet Resources to Identify Studies for Meta-Analysis of Microarray Studies

Database	Web site	Reference
<b>Online repositories of abstracts</b>		
PubMed	<a href="http://www.pubmed.gov/">http://www.pubmed.gov/</a>	
Google Scholar	<a href="http://scholar.google.com/">http://scholar.google.com/</a>	
Web of Science (requires subscription)	<a href="http://wos.mimas.ac.uk/">http://wos.mimas.ac.uk/</a>	
SCOPUS (requires subscription)	<a href="http://www.scopus.com/">http://www.scopus.com/</a>	
<b>Microarray repositories recommended by MIAME for mandatory data deposition</b>		
Array Express	<a href="http://www.ebi.ac.uk/arrayexpress/">http://www.ebi.ac.uk/arrayexpress/</a>	[44]
CIBEX	<a href="http://cibex.nig.ac.jp/">http://cibex.nig.ac.jp/</a>	[45]
Gene Expression Omnibus (GEO)	<a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a>	[46]
<b>Other useful sites for data identification</b>		
ONCOMINE	<a href="http://www.oncomine.org/">http://www.oncomine.org/</a>	[49]
Stanford Microarray Database (SMD)	<a href="http://smd.stanford.edu/">http://smd.stanford.edu/</a>	[50]

doi:10.1371/journal.pmed.0050184.t002

appropriate keywords for automated queries. We recommend searching all the major online repositories of abstracts listed in Table 2 to maximize data acquisition. Reading the latest review articles and directly contacting researchers in relevant fields (Step 5) may help to identify both work potentially missed by automated search, and ongoing research efforts with possibly unpublished data.

In the case of microarrays, one should also search public microarray data repositories [44–46] recommended by the Minimum Information About a Microarray Experiment (MIAME) requirements [47,48], as well as a few more specialized repositories [49,50], listed in Table 2 (Step 4).

Having identified potentially eligible studies from abstracts, one needs to retrieve the articles, where available, and confirm eligibility (Step 7). This process may best be done by at least two people.

## Issue 2: Extract Data from Studies

Before we consider how to extract the data, we need to first decide what type of data to extract. This partially depends on the choice of meta-analysis technique (Issue 6), but the underlying principles will be discussed here. Figure 1 shows the four types of data arising from microarray analysis.

A published gene list (PGL) represents the genes that are declared as differently expressed in a given study. PGLs are often presented in the main or supplementary text of microarray-based studies and are thus easy to obtain. Unfortunately, such PGLs are of limited use for meta-analysis since they represent only a subset of the genes actually studied, and information from many genes will be completely absent. Furthermore, PGLs depend heavily on the preprocessing algorithm, the analysis method, the significance threshold, and the annotation builds used in the original study, all of which usually differ between studies [51]. Thus individual patient-level data (IPD), which for microarrays represents the measurement for every probe in every hybridization, are far more useful. Ioannidis et al. [52] discuss further the advantages of a meta-analysis using IPD versus PGLs.

The gene expression data matrix (GEDM) represents the gene expression summary for every probe and sample and is thus ideally suited as input for meta-analysis. Published GEDMs, however, are unsuitable for meta-analysis because they depend on the choice of the preprocessing algorithms used, which may produce non-combinable results. At present, image files are neither routinely deposited in public

microarray repositories nor technologically uniform enough to be used as input for meta-analysis.

In order to eliminate bias due to specific algorithms used in the original studies, and to allow consistent handling of all datasets, we recommend obtaining the feature-level extraction output (FLEO) files (Step 8), such as CEL and GPR files, and converting them to GEDMs in a consistent manner (see Issue 3). FLEO files are likely to be available, especially for newer studies, because the widely supported MIAME requirements [48] now ask authors to make the FLEO data available in public microarray repositories.

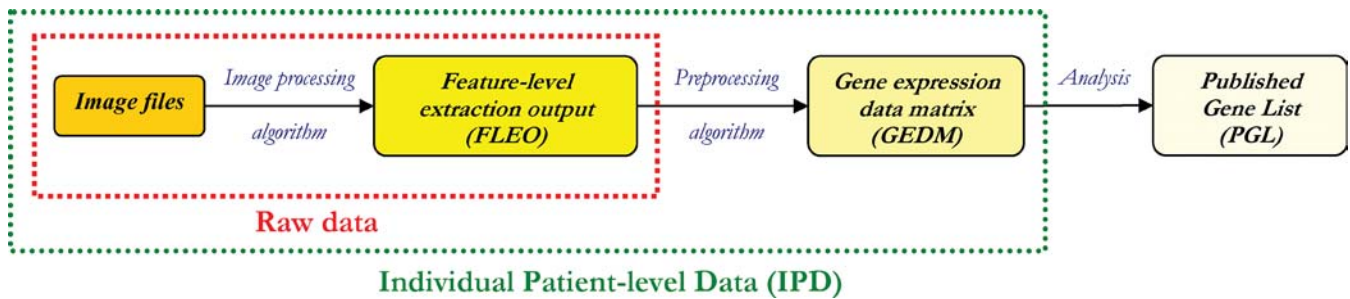
If the main text and supplementary information do not state the location of the FLEO data, then one should try searching public microarray repositories or the research group's Web page before contacting the authors (Step 9). If multiple publications use overlapping sets of data, one should identify and use the most comprehensive dataset available (Step 10), and combine any datasets that were split for algorithm training and validation purposes.

## Issue 3: Prepare Datasets from Different Platforms

FLEO data have to be converted into GEDMs, which can then be used as input for the meta-analysis. The same preprocessing algorithm should be used for multiple studies conducted on the same platform. To combine studies from different platforms, which may have different designs and thus have different options of preprocessing algorithms, it is desirable to try to identify comparable preprocessing algorithms. There are many microarray platforms, but we focus on the most popular: the Affymetrix platform and a set of platforms that could be generically classified as “two-color technology” platforms.

Before the preprocessing step, one may wish to first identify and remove any arrays that are of poor quality (Step 11). There are many comprehensive, free, and open-source packages in BioConductor [53] for quality assessment including arrayMagic [54] for the two-color technology platform and Simpleaffy [55], and affyPLM [56] for the Affymetrix platform.

Next, all good quality arrays should be preprocessed consistently to remove any systematic differences (Step 12). This is an important stage, since preprocessing directly affects the gene expression measurements, and thus all subsequent steps. In practice, researchers are likely to combine datasets from multiple platforms and there are very few preprocessing algorithms that can be applied universally, such as the



doi:10.1371/journal.pmed.0050184.g001

**Figure 1.** The Flow from Data to Information to Biological Knowledge in Gene Expression Microarray Research  
The image files are obtained from optical scanning of hybridized samples.

variance stabilizing normalization [57], which accounts for the dependence between variance and mean of the output expression measure. By contrast, it is more common to use different preprocessing algorithms for each platform [58–61]. Unfortunately, there is currently no consensus on which preprocessing algorithm(s) produce comparable expression measurements across different platforms.

Third, one may also want to check and correct for any batch effects (Step 13), especially in large studies. Unsupervised visualization [62] can help to identify any grouping caused by experimental factors.

Fourth, one needs to decide whether to use all available probes on the array, or a filtered set of probes (Step 14). It is common to filter out probes that have visible defects (e.g., using quality flags), probe-set calls (e.g., absent/present calls from MAS 5.0 preprocessing algorithm), or probes that show little variation (e.g., using minimum coefficient of variation) in single-study analysis. However, it is unclear if such filtering is beneficial from a meta-analysis perspective.

Fifth, one needs to deal with multiple technical replicates (i.e., multiple measurements from the same biological subject) if relevant (Step 15). These should not be treated as independent observations. One approach is to select one of the replicates at random. Alternatively, one can average the replicates. If we assume that all technical replicates have similar array quality, then a simple average or median can be used.

Finally, one could check that the processed expression values from multiple platforms are comparable (Step 16). Microarray platform manufacturers typically include housekeeping genes or negative controls, which are genes expected to be transcribed at a constant level, and may be used for this purpose. Additionally, one may use a visualization technique such as multidimensional scaling [63,64] to inspect for any clustering of arrays by studies.

#### Issue 4: Annotate the Individual Datasets

Microarray probe designers use short, highly specific regions in genes of interest because using the full-length gene sequence can lead to non-specific binding or noise. Different design criteria lead to the creation of multiple probes for the same gene. Therefore, one needs to identify which probes represent a given gene within and across the datasets.

One option is to cluster the probes based on the sequence data (Step 17a) using the BLAST algorithm [65], for example, by using the Ensembl browser [66] (Step 18a). It has been shown that sequence-matched datasets can increase cross-platform concordance [67]. Such methods can also

accommodate Affymetrix probe-set redefinitions [68], which better addresses the problem of alternative splicing. However, the probe sequence may not be available for all platforms and the clustering of probe sequences could be computer intensive for very large numbers of probes.

Alternatively, one can map probe-level identifiers such as I.M.A.G.E. CloneID, Affymetrix ID, or GenBank accession numbers to a gene-level identifier such as UniGene, RefSeq, or Entrez Gene ID. UniGene [69], which is an experimental system for automatically partitioning sequences into non-redundant gene-oriented clusters, is a popular choice to unify the different datasets. For example, UniGene Build #211 (released March 12, 2008) reduces nearly 7 million human sequences to 124,181 clusters. To translate probe-level identifiers to gene-level identifiers, one can use either the annotation packages in BioConductor [53] or Web tools such as SOURCE [70] and RESOURCERER [71] (Step 18b). We suggest using I.M.A.G.E. CloneID [72] or Affymetrix ID first, if available, as they are more sequence-specific (Step 17b). The same mapping build, ideally the most recent, should be used for all datasets to avoid inconsistencies between releases [73,74].

#### Issue 5: Resolve the Many-to-Many Relationships between Probes and Genes

In this section, we will refer to either the sequence cluster ID or the gene-level identifier (such as UniGene ID or RefSeq ID) used to annotate the datasets, simply as the GeneID.

Many probes can map to the same GeneID because of the clustering nature of the UniGene, RefSeq, and BLAST systems involved, or because the microarray chips used contain duplicate spotted probes. On the other hand, a probe may map to more than one GeneID if the probe sequence is not specific enough. Sometimes, a probe has insufficient information to be mapped to any GeneID, and we recommend omitting these from further analysis (Step 19). Inconsistencies between annotation databases or releases and software [73–75] complicate the matter further. The illustrative example of a meta-analysis of 25 datasets presented later in this paper contains 537,686 probes. Of these probes, 47,154 (or 8.7%) could not be mapped to any UniGene ID, while 29,774 (or 6.1%) of the remaining probes mapped to more than one UniGene ID.

This “many-to-many” relationship can fragment the available information for meta-analysis. For example, a probe could map to GeneID X in half of the datasets but to both GeneIDs X and Y in the remaining datasets. Software that performs automated meta-analysis on several thousand

genes will treat such probes as two separate gene entities, failing to fully combine the information for GeneID X from all studies.

A simple approach is to use only the probes with one-to-one mapping for further analysis, but this means losing information, and so is not recommended. In the example above, potentially half of the information for GeneID X (i.e., from probes mapping to both X and Y) will be ignored. Therefore, when relevant, we recommend replacing probes with multiple GeneIDs by a new record for each GeneID (Step 21). This greedy approach of “expanding” the probes with multiple GeneIDs ensures the software uses all possible information.

On the other hand, how should one deal with multiple probes that map to the same GeneID within a given study? Grützmann et al. [24] treated these as independent observations in the meta-analysis, but we recommend summarizing them (Step 22) into a single representative value per key within a study.

Several options are available to summarize information in this situation. First, one could select a probe at random, but this means losing information. Simply averaging the expression profiles before proceeding is not desirable either, as different probe sequences have different binding affinity, giving rise to the problem of different measurement scales. Thus, it is preferable to work with standardized measures such as the  $p$ -value or effect size. When working with standardized measures, one could select the most extreme value, since it is least likely to occur by chance. For example, Rhodes et al. [19] used the smallest  $p$ -value of the probes that corresponded to each GeneID. A more sophisticated approach, when working with effect size, is to meta-analyze the probes.

Recently, the MicroArray Quality Control (MAQC) project [61] described another alternative to resolve the many-to-many mapping. For a probe that mapped to multiple RefSeq IDs, the authors selected the RefSeq ID that was annotated by TaqMan assays and, secondarily, one that was present in the majority of platforms. Next, if many probes mapped to a given RefSeq ID, they chose the one closest to the 3' end of the gene.

After resolving for the many-to-many relationship by expanding and summarizing probes, we are left with one summary statistic per GeneID per study. In the next step, we proceed with meta-analyzing the summary statistic for each GeneID in turn across the studies.

## Issue 6: Choosing a Meta-Analysis Technique

The choice of meta-analysis technique depends on the type of response (e.g., binary, continuous, survival) and objective. In this article, we focus on a fundamental application of microarrays: the two-class comparison where the objective is to identify genes expressed differentially between two well-known conditions. There are four generic ways of combining information in such a situation. (For clarity of presentation, we indicate the steps only for the inverse-variance technique.)

**Vote counting.** Here, one counts the number of studies in which a gene was declared significant [76]. For very small numbers of studies, the results can be visualized using a Venn diagram [77]. Vote counting in the context of microarrays is perhaps best described by Rhodes et al. [22], who also suggest calculating the null distribution of votes using permutation testing. Alternatively, one could calculate the significance of the overlaps using the normal approximation to binomial as

described in Smid et al. [30]. Yang et al. [35] extend both of these techniques into the concept of meta-analysis pattern matches.

**Combining ranks.** Unlike vote counting, this technique accounts for the order of genes declared significant. DeConde et al. [37] use three different approaches to aggregate the rankings of, say, the top 100 lists (the 100 most significantly up-regulated or down-regulated genes) from different studies. Two of the algorithms use Markov chains to convert the pair-wise preference between the gene lists to a stationary distribution; the third algorithm is based on an order-statistics model. Zintzaras and Ioannidis [40] proposed METa-analysis of RAnked DIScovery datasets (METRADISC), which is based on the average of the standardized rank and has the advantage of incorporating the between-study heterogeneity (sum of squared deviations from the average). The null distributions for the average rank and heterogeneity are then estimated using non-parametric Monte Carlo permutation testing and matched for pattern of occurrence in studies. Hong et al. [38] proposed the RankProd [78], which calculates the product of the rank of pair-wise differences between every biological sample in one group versus another group across the studies.

**Combining  $p$ -values.** Rhodes et al. [19] use Fisher's sum of logs method [79], which sums the logarithm of the (one-sided hypothesis testing)  $p$ -values across  $k$  studies for a given gene. The test statistic can be compared against a chi-square distribution with  $2k$  degrees of freedom.

**Combining effect sizes.** Choi et al. [29] and others [24,32,80] used the inverse-variance technique [81,82] in the context of microarrays. The first step is to calculate the effect size and the variance associated with the effect size for every gene in every study (Step 20). Effect size can be calculated as the Cohen's  $d$  [83], which is the difference in two group means standardized by its pooled standard deviation [84]. Hedges and Olkin (1985) showed that this standardized difference overestimates the effect size for studies with small sample sizes. They proposed a small correction factor to calculate the unbiased estimate of the effect size, which is known as the Hedges' adjusted  $g$ . The study-specific effect sizes for every gene are then combined across studies into a weighted average (Step 24). As the name suggests, the study weights are inversely proportional to the variance of the study-specific estimates.

Additionally, the integrative correlation technique proposed by Parmigiani et al. [33] could be first used to select only the “reproducible” genes for meta-analysis. First, the correlation profile of gene  $G$  is calculated as the correlation between gene  $G$  and every other gene in a study. Next, the correlation of correlation profiles of gene  $G$  in every pair of studies is computed, and if the average exceeds a certain threshold, the gene is called reproducible.

Given the various statistical options for meta-analysis, how should one choose the most suitable technique? We present a series of questions that could help a meta-analyst make an informed choice.

First, what are the minimum data required for each technique? Fisher's method, the inverse-variance technique, METRADISC, and the RankProd all require IPD, which are less readily available than PGLs. Vote counting, DeConde and colleagues' algorithms, and combining  $p$ -values are techniques that in theory could use the PGLs, but may not be able to

do so in practice. For example, most publications report the significant genes or their rankings based on two-sided  $p$ -values, while vote counting and rank aggregation techniques require a one-sided  $p$ -value. Using  $p$ -values from two-sided testing means ignoring the directionality of the significance and may lead one to select genes that are discordant in direction of gene regulation between the studies. As noted earlier in Issue 2, we strongly prefer to use the IPD to minimize the influence of differing methods across datasets.

Second, which set of genes does each technique use? Vote counting and rank aggregation techniques (using PGLs) only consider the genes declared significant in the original studies. Thus, these techniques depend on an arbitrary threshold, and completely ignore genes that fall below this selected threshold. By contrast, the rank aggregation technique (using IPD), Fisher's method, and the inverse-variance technique consider information from all available genes. However, it is also important to note that the ranking of genes in an individual study depends on which other genes are included in the chip, and thus can influence the rank aggregation techniques. Since microarrays are often used as a hypothesis generating tool, we would prefer a technique that captures information from as many genes as possible.

The third question, related to the previous question, is how does each technique treat frequently studied and rarely studied genes? Newer microarrays chips have more comprehensive sets of genes compared to older chips. Thus some genes will be studied more frequently across the studies than others. For example, Affymetrix version HGU-133 plus 2.0 (released in 2003) contains almost all of 6,065 UniGene IDs available in Affymetrix version HU-6800 (released in 1998), plus a further additional 13,624 UniGene IDs. Ideally, we would prefer a technique that treats a frequently studied and a rarely studied gene equally.

Since vote counting and rank aggregation use the genes declared significant in the original studies, they do not account for the frequency of the genes. For example, a gene found significant in four studies and not significant in 16 studies will be favored over a gene found significant in three studies but absent in the other 17 studies. METRADISC accounts for this by matching each gene to the null distribution of genes that have the same absent/present patterns. Although the test statistic for Fisher's method is based on an unstandardized sum, it can address this problem by comparing it to a chi-square distribution where the degree of freedom is determined by the number of studies or by permutation. The inverse-variance technique addresses this problem directly as it calculates a weighted average of the effect sizes.

Fourth, what is the ability of each technique to rank the genes, especially if only a small number of studies, say three to five, are available? A ranked list can help researchers to prioritize genes for further testing and validation. The vote counting technique produces very granular results, while other techniques produce results on a much finer scale.

Fifth, what is the computational complexity involved for each technique once the datasets have been prepared and annotated? The computing time for meta-analyzing the prepared and annotated GEDM for the 25 datasets in the illustrative example that follows, using vote counting, Fisher's method and inverse-variance technique are approximately two minutes, two minutes, and eight minutes respectively. We used R version 2.5.1 [85] on a Windows-based personal

computer with a 1.86 GHz Intel Pentium M processor and 1 GB of RAM memory. Further, any technique that uses PGLs has to extract the information and annotation in a standardized format. The question of computational complexity becomes important, especially when one wants to estimate the null distribution using permutation techniques.

We believe that combining the effect sizes using an inverse-variance model is the most comprehensive approach for meta-analysis of two-class gene expression microarrays. In addition to the characteristics discussed above, this method has several other decisive advantages. First, it yields a biologically interpretable discrimination measure—the pooled effect size of differential expression and its standard error. Second, it is the only technique that weights the contribution of each study by its precision, which is related to the study sample size. Third, one is able to use a forest plot [86] to visually investigate the contributions of individual studies and the amount of heterogeneity across datasets. The use of effect size, a unitless measure not dependent on sample size, facilitates the combining of signals from one-color and expression ratios from two-color technology platforms.

### Illustrative Example: Differential Gene Expression in Cancer Tissues

We demonstrate one exemplary meta-analysis using a subset of an ongoing meta-analysis where we look at the differences between cancerous tissues relative to normal tissues across various cancer types. This example stops short of discussing the biological significance of the findings, which is beyond the scope of this article.

We concisely describe the meta-analysis protocol in Table 3, using the same ordering as in Table 1. Figure 2 shows the data acquisition process, and Table 4 lists the characteristics of the 21 studies included [87–107]. Arrays from the Affymetrix-based studies were preprocessed using the robust multichip average [108], and arrays from two-color technology were LOESS (local regression) normalized [109, 110]. All analysis (unless stated otherwise) was carried out in R version 2.5.1 [85] and BioConductor release 2.0 [53]. The R codes are available upon request.

We chose to combine the effect sizes using the inverse-variance model for the reasons described previously. Note that there are two variants of the inverse-variance technique. The random effects model used differs from the fixed effect model in that it incorporates the between-study heterogeneity into study weights. We use the random effects model in Step 24, where we can expect significant between-study heterogeneity since the studies combined are both biologically (e.g., different tumors) and technically diverse (e.g., different platforms, laboratories). We used the fixed effects used in Step 22 to summarize probes within a study as we can expect a reasonable level of homogeneity within a study.

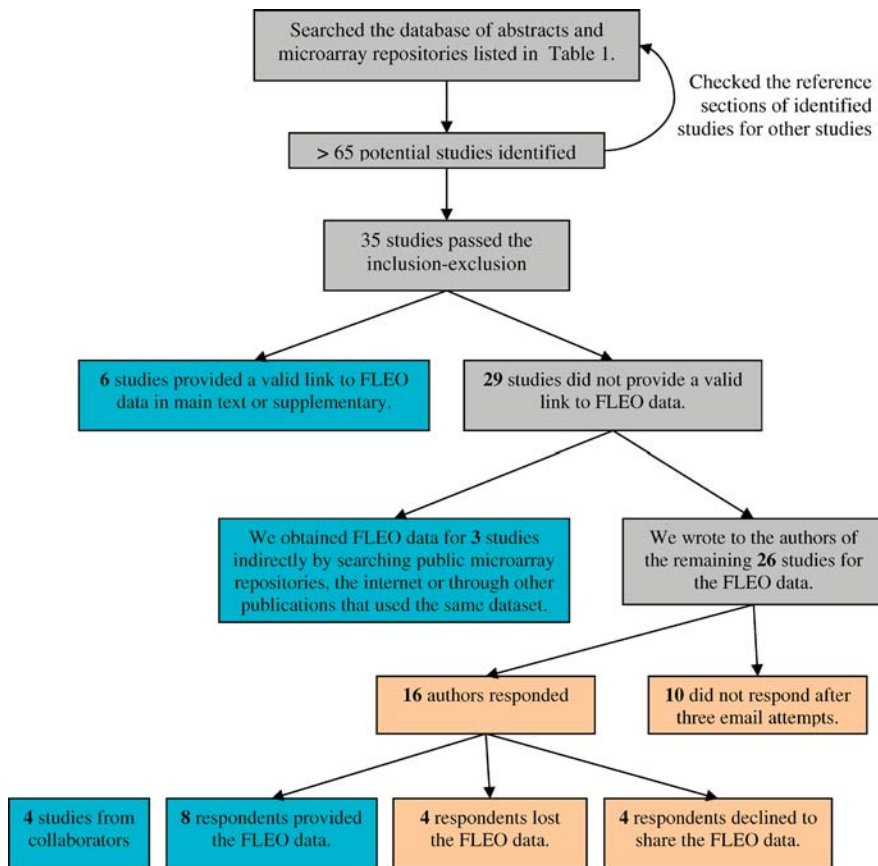
The pooled effect size and its 95% confidence interval for all 16,803 genes can be visualized simultaneously as in Figure 3.

The  $z$ -statistic (ratio of the pooled effect size to its standard error) for every UniGene ID was compared to a standard normal distribution to obtain the  $p$ -value and adjusted for false discovery rate (FDR) [111] (Step 25). Table 5 shows the output from the inverse-variance technique for the top five statistically significant up-regulated and down-regulated genes.

**Table 3.** Outline of the Illustrative Example of Meta-Analysis

Step	Action
1	<b>Objective:</b> To identify genes that are consistently up- or down-regulated in cancers globally.
2	<b>Inclusion-exclusion criteria:</b> Any human studies investigating at least 7 patients with primary cancer and at least 7 patients with corresponding normal samples using high-density arrays. Any patients with metastatic tumors, cell lines, or benign tumors, or studies using specialized arrays, were excluded.
3–10	<b>Data identification and acquisition:</b> See Figure 2 for data retrieval of 21 studies. Ramaswamy et al. [104] had 5 sets of cancer-normal tissues that satisfied our criteria, and thus we have 25 datasets (see Table 4).
11	<b>Array quality check:</b> Not performed as this information was not available for all retrieved studies.
12	<b>Preprocessing FLEO files:</b> Arrays from the Affymetrix platform were RMA-preprocessed [108], and arrays from two-color technology were LOESS-normalized [88,89] and then standardized to have the same variability across arrays within studies. We omitted any arrays with more than 25% missing values. Note that the expression values are on log base 2 scale.
13 & 14	<b>Batch effect and spot quality check:</b> Not performed as this information was not available for all studies.
15	<b>Aggregate technical replicates:</b> Only Bhattacharjee et al. (2001) [90] had technical replicates, which we averaged using a simple mean.
16	<b>Compatibility check:</b> Not performed.
17 & 18	<b>Annotation matching:</b> For the Affymetrix studies, we mapped the probe-sets to UniGene using the annotation packages in BioConductor 1.8.0 (built on March 26, 2006). For the two-color technology arrays, we mapped the clone IDs, and if not available the GenBank Accession number, to UniGene using the web tool SOURCE [70] in March 2006.
19	<b>Discard non-identifiable probes:</b> About 8.7% of the cumulative 537,686 probes could not be mapped to any UniGene and were discarded.
20	<b>Calculate study-specific estimates:</b> For every probe and for every study, we calculated effect size as the Hedges' adjusted <i>g</i> .
21	<b>Expanding probes with multiple UniGene IDs:</b> As described in text of Issue 5.
22	<b>Summarizing multiple probes per UniGene ID within a study:</b> We used the fixed effect inverse-variance model.
23	<b>Discard poorly represented probes:</b> The probes map to 28,365 unique UniGene IDs, but we restricted the analysis to the 16,803 that were identified in at least 5 of the 25 sets.
24	<b>Combine study-specific estimates:</b> We use the random effects inverse-variance technique for each UniGene ID in turn to obtain the pooled effect size and its standard error.
25	The <i>z</i> -statistics (ratio of the pooled effect size to its standard error) for every UniGene ID was compared to a standard normal distribution to obtain the nominal <i>p</i> -value. We adjusted for multiple testing using the FDR [111].

LOESS, local regression; RMA, robust multichip average.  
doi:10.1371/journal.pmed.0050184.t003



doi:10.1371/journal.pmed.0050184.g002

**Figure 2.** Data Acquisition to Summarize Steps 3–10 in Table 3

In total, 21 studies (6 + 3 + 8 + 4) are included in the meta-analysis. The characteristics of the included studies are given in Table 4.

**Table 4.** Datasets Used in the Illustrative Meta-Analysis

Study Information			# Samples		Array and Annotation Information				
Citation	Tissue	PubMed ID	Normal	Tumor	Platform	Total Probes	Probes with UniGenes	After UniGene Expanding	After UniGene Summarizing
Aldred et al. (2003) [87]	Thyroid	12776192	7	9	HGU95Av2	12,625	12,104	13,134	9,364
Basso et al. (2005) [88]	B cells	15778709	10	66	HGU95Av2	12,625	12,104	13,134	9,364
Beer et al. (2002) [89]	Lung	12118244	10	86	HU6800	7,129	6,919	7,556	6,069
Bhattacharjee et al. (2001) [90]	Lung	11707567	17	139	HGU95Av2	12,625	12,104	13,134	9,364
Chen et al. (2002) [91]	Liver	12058060	75	104	cDNA	22,607	21,032	22,880	14,772
Chen et al. (2003) [92]	Gastric	12925757	22	90	cDNA	42,037	38,276	40,471	22,150
Couvelard et al. (2005) [93]	Pancreas	15910598	7	12	cDNA	9,910	8,446	8,446	5,097
Dyrskjot et al. (2004) [94]	Bladder	15173019	9	41	HGU133A	22,283	21,288	23,134	13,223
Hippo et al. (2002) [95]	Stomach	11782383	8	22	HU6800	7,129	6,919	7,556	6,069
Hu et al. (2005) [96]	Lung	15592519	14	47	cDNA	9,927	8,443	8,443	5,096
Huang et al. (2001) [97]	Thyroid	11752453	8	8	HGU95A	12,626	12,110	13,145	9,391
Jones et al. (2004) [98]	Lung	15016488	19	24	cDNA	40,368	35,370	35,370	20,993
Klein et al. (2001) [99]	B cells	11733577	20	32	HGU95A	12,626	12,110	13,145	9,391
Kuriakose et al. (2004) [100]	Head & neck	15170515	17	17	HGU95Av2	12,625	12,104	13,134	9,364
Lapointe et al. (2004) [101]	Prostate	14711987	41	62	cDNA	40,699	37,460	39,673	21,228
Lenburg et al. (2003) [102]	Renal	14641932	8	9	HGU133 set	44,928	39,884	43,194	18,249
Pellagatti et al. (2006)[103]	Bone marrow	16527891	11	58	HGU133A plus 2.0	54,675	47,147	50,987	19,738
Ramaswamy et al. (2001) [104]	Bladder	11742071	7	11	HU8000 + HU35kSubA	16,063	15,069	16,369	10,881
	Brain		8	20					
	Colon		11	9					
	Kidney		11	8					
	Pancreas		10	11					
Singh et al. (2002) [105]	Prostate	12086878	50	52	HGU95Av2	12,625	12,104	13,134	9,364
Welsh et al. (2001) [106]	Prostate	11507037	8	25	HGU95A	12,626	12,110	13,145	9,391
Winter et al. 2007 [107]	Head & neck	17409455	11	11	HGU133A plus 2.0	54,675	47,147	50,987	19,738
<b>Total</b>			<b>419</b>	<b>973</b>		<b>537,686</b>	<b>490,532</b>	<b>525,658</b>	<b>28,365</b>

doi:10.1371/journal.pmed.0050184.t004

At the FDR rate of 1%, we found 168 significantly down-regulated and up-regulated genes. At this rate, we should expect 1% of the significant genes list, and in this case 1.68 and 3.25 in each list respectively, to be false positives.

After having identified the genes of most interest, we can proceed as in a traditional meta-analysis and visualize the contribution of individual studies using forest plots (Step 27). Figure 4 shows the forest plot for the most significantly up-regulated (Hs.478481) and down-regulated (Hs.117835) genes.

We can also proceed as in a typical single-study analysis. For example, using significant genes identified from the meta-analysis, we can use computational tools such as pathway enrichment (Step 28), conduct a literature search, and/or validate them on an alternative technology or on different patient sets (Step 29).

In this illustrative example of a meta-analysis, we have shown how the inverse-variance technique can identify consistently up- or down-regulated genes, information that suggests further lines of investigation.

## Discussion

Meta-analysis of microarray datasets shares many features with meta-analysis in other areas of health care research. Perhaps the main differences are the large numbers of variables involved and technical complexities of integrating data across multiple platforms. Furthermore, most microarray studies are not prospectively planned and often do not have detailed

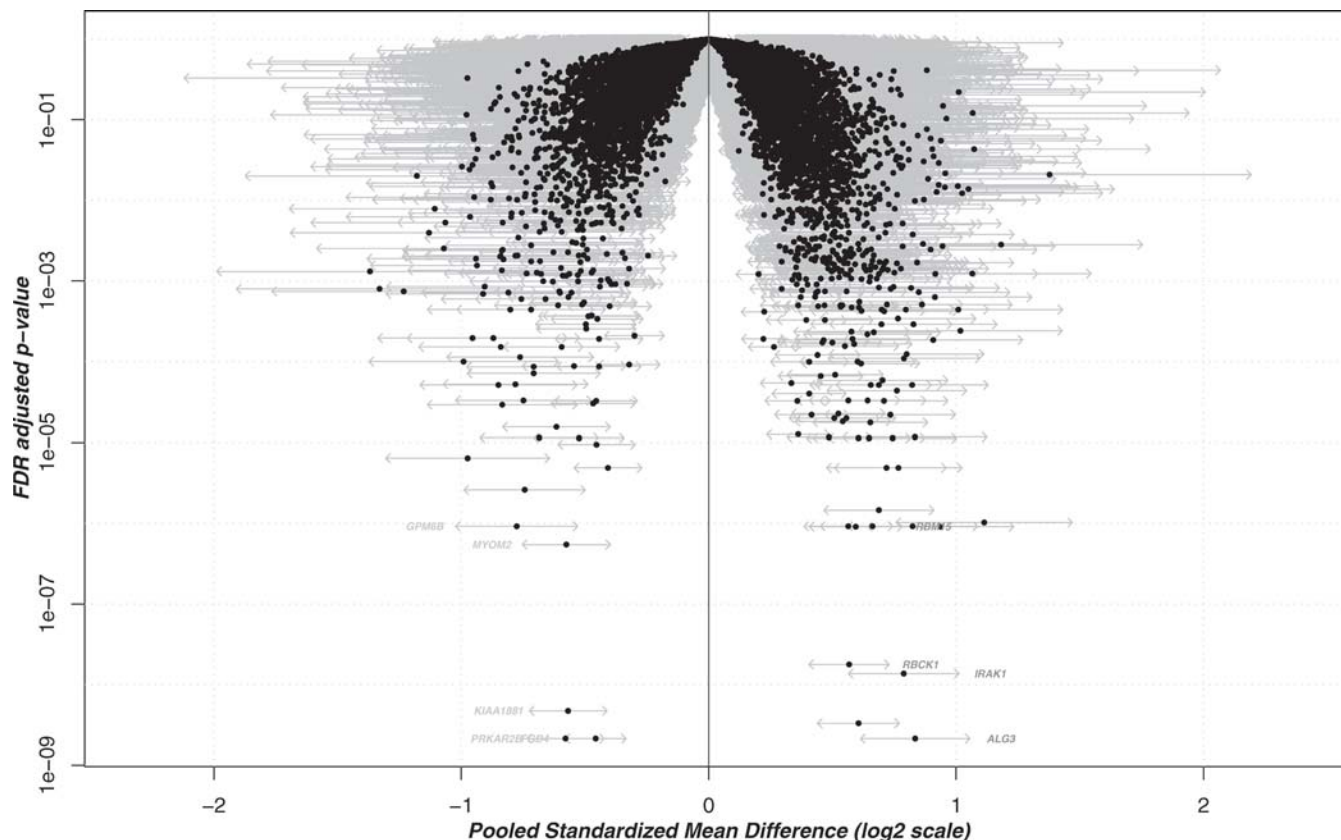
protocols, but rather tend to make use of existing samples. Table 6 gives an overview of the advantages and disadvantages of various aspects of meta-analysis of microarray datasets. We discuss some of these points below.

Working with FLEO files allows for better standardization of information and the incorporation of data from unpublished studies, but it also requires significant effort to acquire and manage the datasets due to increased data complexity. This is further hampered by data sharing issues ([112–115] and Ramasamy et al., unpublished data).

Sample matching between “cases” and “controls” may be a problem in meta-analysis as much as in single studies. Leaving aside the choice of biological equivalency of cases and controls, the numerical problem is highlighted by the imbalance of samples between the two groups in the illustrative example (see Table 4). For example, while the proportion of normal to total biological samples in prostate and lung cancer (the two tissues with the greatest number of biological samples in the illustrative example) is far less than half, the proportions do vary (105 out of 452 or 23.2% in prostate cancer versus 60 out of 356 or 16.9% in lung cancer).

Another major concern associated with meta-analysis in many clinical and epidemiological studies is the problem of publication bias, which is a consequence of selectively publishing statistically significant and favorable results [116,117]. On the surface, we do not expect to find a publication bias at a gene level in a given study because of the discovery-driven and high-density nature of microarrays.





doi:10.1371/journal.pmed.0050184.g003

**Figure 3.** A Summary Plot of the Pooled Effect Size (Black Dots) and Its 95% Confidence Interval (Gray Bars) Sorted by the FDR. The GenBank identifier (if available) for the top five most statistically significant up-regulated and down-regulated genes is shown.

However, anecdotal evidence based on sales figures (J. P. Ioannidis, personal communication) suggests that data from only 10% of all the Affymetrix chips sold are published. The possibility of publication bias in microarray research needs further investigation.

Furthermore, within a single-study microarray analysis, the particular choice of down-stream analysis may lead to different results depending on the objective of the study [118,119]. It is unclear to what extent this problem affects meta-analysis of microarrays, even with coherently preprocessed datasets.

Finally, the sensitivity of the results from meta-analysis, as with any other research study, should be tested before a final conclusion is reached (Step 26). We did not present any sensitivity analysis for the illustrative example presented here, but there are several possibilities. First, we could investigate sensitivity of the results to the choices we made here (e.g., using probes present in at least five studies). Secondly, we can test if any particular study is particularly influential, by repeating the meta-analysis without each study in turn and comparing the change. Finally, we could test if the inclusion of studies that provide only the GEDM into the

**Table 5.** The Output from the Inverse-Variance Technique for the Top Five Statistically Significant Up-Regulated and Down-Regulated Genes

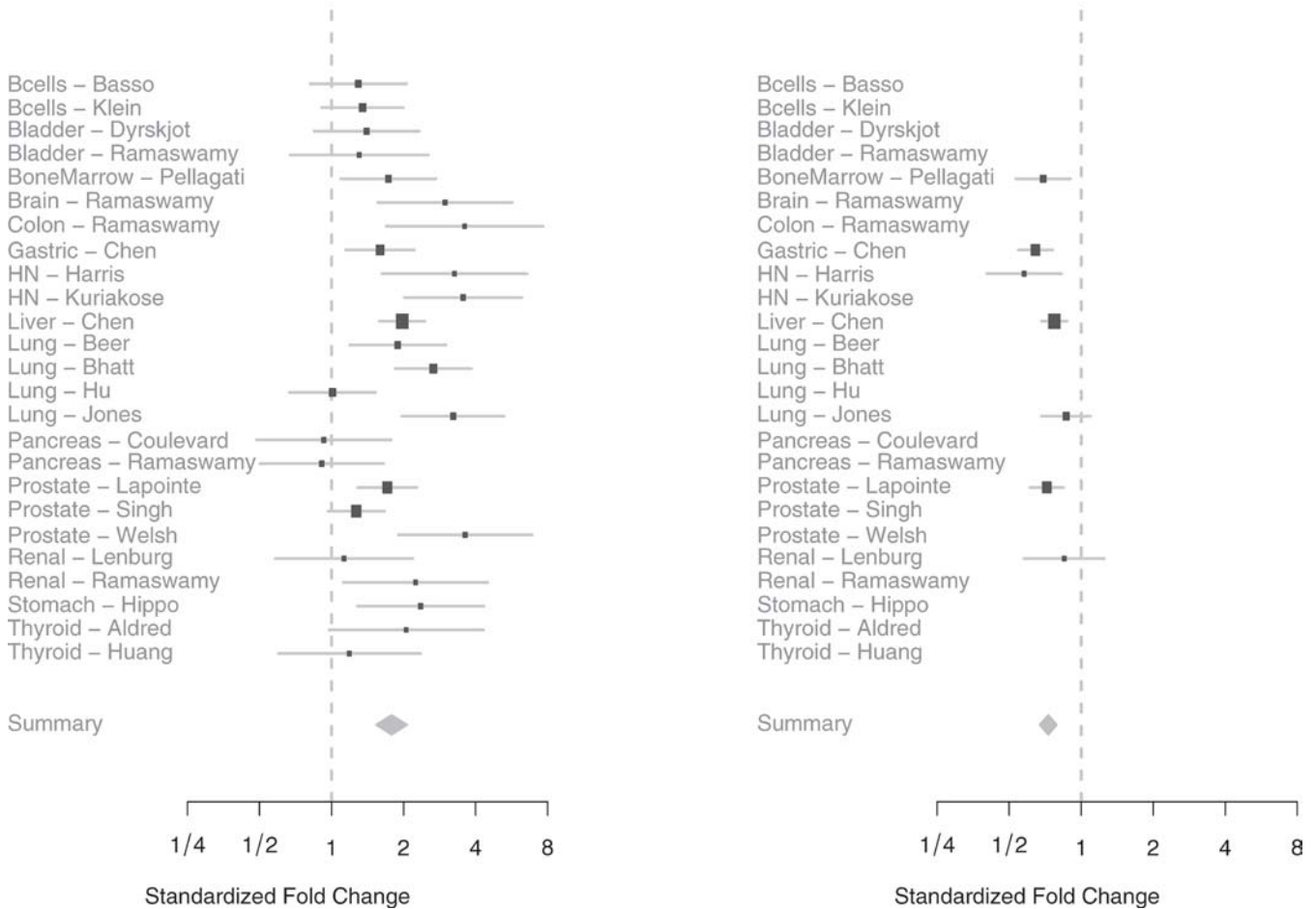
UniGene ID	<i>n</i>	Effect	Size	Standard Error	$\delta^2$	<i>z</i>	pFDR	Symbol
<b>Up-regulated</b>								
Hs.478481	25	0.834	0.113	0.175	7.4	1.39E-13	2.14E-09	ALG3
Hs.643516	7	0.605	0.084	0	7.16	7.89E-13	3.31E-09	
Hs.522819	23	0.788	0.114	0.177	6.91	4.89E-12	1.37E-08	IRAK1
Hs.247280	21	0.567	0.083	0.082	6.85	7.41E-12	1.78E-08	RBCK1
Hs.435947	7	0.594	0.096	0.029	6.22	4.93E-10	9.13E-07	RBM15
<b>Down-regulated</b>								
Hs.117835	7	-0.457	0.063	0.005	-7.3	2.95E-13	2.14E-09	FGD4
Hs.433068	25	-0.578	0.08	0.052	-7.26	3.82E-13	2.14E-09	PRKAR2B
Hs.591387	7	-0.568	0.08	0.004	-7.08	1.41E-12	4.74E-09	KIAA1881
Hs.443683	24	-0.575	0.091	0.074	-6.32	2.60E-10	5.47E-07	MYOM2
Hs.495710	25	-0.777	0.126	0.305	-6.16	7.49E-10	9.21E-07	GPM6B

For each UniGene ID, the table shows the number of studies the gene was present in (*n*), the pooled effect size, standard error of the pooled effect size, between-study heterogeneity measure ( $\delta^2$ ), *z*-score, *p*-value of the *z*-score, the FDR adjusted value, and GenBank symbol (if available).

doi:10.1371/journal.pmed.0050184.t005

Hs.478481

Hs.117835



doi:10.1371/journal.pmed.0050184.g004

**Figure 4.** Forest Plot of the Most Statistically Significant Up-Regulated and Down-Regulated Genes Identified from the Meta-Analysis

**Table 6.** Advantages and Disadvantages of Various Aspects of Meta-Analysis of Microarray Datasets

Advantages	Disadvantages
<b>Combining independent but related studies</b>	
Increases statistical power	Assumes there are sufficient numbers of suitable studies available
Increases generalizability of results	Study selection issues such as study diversity and data quality need to be addressed
Financially inexpensive as it uses existing studies	Time and effort to acquire and manage data can be significant
	Potential publication bias might limit meta-analysis
<b>IPD versus PGLs</b>	
IPD avoids selective reporting of genes	Harder to acquire IPD compared to gene lists
IPD permits re-analysis of individual studies for reproducibility or to carrying out other analysis	
<b>FLEO data versus GEDM</b>	
FLEO data allows us to standardize preprocessing algorithms and analysis methods	More time and effort required to acquire and manage FLEO relative to GEDM
	In practice, some researchers may withhold access to FLEO and thereby introduce a possible bias
<b>Inverse-variance technique versus other techniques for combining data</b>	
Interpretable results with standard error to construct confidence intervals	A large study in the collection may influence the overall results of meta-analysis
Treats a rarely studied and frequently studied genes equally	Ignores correlation between genes (as do most techniques)
Weights the contribution from different studies by its precision	IPD may not be available for all studies.
Good ability to rank results when applied on small number of studies	
<b>Use of forest plots</b>	
Visualize contributions from individual studies	Only possible to look at the forest plots for a small number of genes
Assess heterogeneity of result across datasets	Descriptive in nature

doi:10.1371/journal.pmed.0050184.t006

meta-analysis along with the studies that provide FLEO data changes the results.

In this paper, we have formulated and explored key issues encountered in conducting a meta-analysis of microarray datasets. We considered the available solutions and made some practical recommendations. First, we showed how to obtain suitable datasets by searching the published literature and public microarray repositories. Second, we proposed that using FLEO files allows for better standardization of information. Third, we outlined the issues involved in preparing datasets from multiple platforms. Fourth, we discussed how to match the different datasets using gene-level identifiers. Fifth, we explained how to resolve the problems caused by the many-to-many relationship between the probes and genes by “expanding” probes with multiple GeneIDs and then “summarizing” the multiple probes that correspond to a GeneID within a study. Sixth, we argued that the inverse-variance technique, initially proposed in the microarray context by Choi et al. [29], has many desirable properties over other techniques used for two-class comparison of gene expression microarray studies. Finally, we presented an illustrative meta-analysis of 25 datasets to briefly demonstrate the issue of how to present, analyze, and interpret a meta-analysis of microarray datasets. All of this information is neatly captured in a practical checklist, shown in Table 1. ■

## Glossary

**Feature-level extraction output file (FLEO):** A file representing the quantification of optical image scans of a microarray chip. Every row in this file gives the pixel-level summaries of foreground and background signals for a probe as well as any quality measure. Examples of FLEO files generated include those with .CEL and .GPR file extensions.

**Gene expression data matrix (GEDM):** A file that contains the summary gene expression from all the FLEO files in a given study. The format is typically a matrix where every row represents a probe and every column represents a hybridization.

**Individual patient-level data (IPD):** In microarray studies, a dataset that provides the gene expression summary for every hybridized sample.

**Minimum Information About a Microarray Experiment (MIAME):** Data-reporting requirements that have been widely adopted by many journals.

**Preprocessing algorithm:** An important step in microarray analysis that tries to minimize systematic variation. It typically consists of background noise correction within an array, normalization between arrays, and a probe-set summary.

**Probe:** The DNA sequence spotted on the microarray surface to represent a gene. For a given gene, many probes can be designed. A probe can ambiguously map to more than one gene if its sequence is not specific enough.

**Published gene list (PGL):** A published list of genes that are declared differently expressed in a given study. It depends on the preprocessing algorithm, analysis method, chosen significance threshold, and annotation build used.

**Sample:** Biological material from a research participant or subject (e.g., a patient or animal) that can be hybridized onto a microarray chip.

## Acknowledgments

We would like to thank Francesco Pezzella, Jianting Hu, Lance D. Miller, and Philip M. Long for initiating the projects that motivated this paper, and Francesca Buffa and Jennifer Taylor for helpful comments on the manuscript. Special thanks to all the authors of the studies that provided the FLEO data for the illustrative case study.

## References

1. Schena M, Shalon D, Davis R, Brown P (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.
2. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, et al. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14: 457-460. doi:10.1038/ng1296-457
3. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537.
4. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, et al. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 96: 9212-9217.
5. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511. doi:10.1038/35000501
6. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, et al. (2001) Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* 98: 10787-10792. doi:10.1073/pnas.191368598
7. Dan S, Tsunoda T, Kitahara O, Yanagawa R, Zembutsu H, et al. (2002) An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res* 62: 1139-1147.
8. van't Veer LJV, Dai H, van de Vijver MJ, He YD, Hart AAM, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536. doi:10.1038/415530a
9. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999-2009. doi:10.1056/NEJMoa021967
10. Chang HY, Nuyten DSA, Sneddon JB, Hastie T, Tibshirani R, et al. (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 102: 3738-3743. doi:10.1073/pnas.0409462102
11. Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, et al. (2006) Gene signature evaluation as a prognostic tool: Challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* 3: 540-551.
12. Paik S (2007) Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. *Oncologist* 12: 631-635. doi:10.1634/theoncologist.12-6-631
13. Ntzani E, Ioannidis J (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment. *Lancet* 362: 1439-1444.
14. Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: A multiple random validation strategy. *Lancet* 365: 488-492.
15. Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2005) Outcome signature genes in breast cancer: Is there a unique set? *Bioinformatics* 21: 171-178. doi:10.1093/bioinformatics/bth469
16. Jafari P, Azuaje F (2006) An assessment of recently published gene expression data analyses: Reporting experimental design and statistical factors. *BMC Med Inform Decis Mak* 6: 27. doi:10.1186/1472-6947-6-27
17. Dupuy A, Simon R (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99: 147-157.
18. Ferguson L (2004) External validity, generalizability, and knowledge utilization. *J Nurs Scholarsh* 36: 16-22.
19. Rhodes D, Barrette T, Rubin M, Ghosh D, Chinnaiyan A (2002) Meta-analysis of microarrays: Inter-study validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62: 4427-4433.
20. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression analysis of human 490 genes across many microarray data sets. *Genome Res* 14: 1085-1094. doi:10.1101/gr.1910904
21. Pilarsky C, Wenzig M, Specht T, Saeger HD, Grützmann R (2004) Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. *Neoplasia* 6: 744-750. doi:10.1593/neo.04277
22. Rhodes D, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 101: 9309-9314.
23. Wang J, Coombes KR, Highsmith WE, Keating MJ, Abruzzo LV (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. doi:10.1093/bioinformatics/bth381
24. Grützmann R, Boriss H, Ammerpohl O, Lüttges J, Kalthoff H, et al. (2005) Meta-analysis of microarray data on pancreatic cancer defines a set of

- commonly dysregulated genes. *Oncogene* 24: 5079-5088. doi:10.1038/sj.onc.1208696
25. Mehra R, Varambally S, Ding L, Shen R, Sabel MS, et al. (2005) Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 65: 11259-11264. doi:10.1158/0008-5472.CAN-05-2495
  26. Bianchi F, Nuciforo P, Vecchi M, Bernard L, Tizzoni L, et al. (2007) Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. *J Clin Invest* 117: 3436-3444. doi:10.1172/JCI32007
  27. Kim SY, Kim JH, Lee HS, Noh SM, Song KS, et al. (2007) Meta- and gene set analysis of stomach cancer gene expression data. *Mol Cells* 24: 200-209.
  28. Silva GL, Junta CM, Mello SS, Garcia PS, Rassi DM, et al. (2007) Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. *Ann NY Acad Sci* 1110: 33-46. doi:10.1196/annals.1423.005
  29. Choi J, Yu U, Kim S, Yoo O (2003) Combining multiple microarray studies and modelling inter-study variation. *Bioinformatics* 19 (Suppl 1): i84-i90.
  30. Smid M, Dorsers LCJ, Jenster G (2003) Venn mapping: Clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. *Bioinformatics* 19: 2065-2071.
  31. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302: 249-255. doi:10.1126/science.1087447
  32. Choi J, Choi J, Kim D, Choi D, Kim B, et al. (2004) Integrative analysis of multiple gene expression profiles applied to liver cancer study. *FEBS Lett* 565: 93-100.
  33. Parmigiani G, Garrett-Mayer E, Anbazhagan R, Gabrielson E (2005) A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clin Cancer Res* 10: 2922-7.
  34. Warnat P, Eils R, Brors B (2005) Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics* 6: 265. doi:10.1186/1471-2105-6-265
  35. Yang X, Bentink S, Spang R (2005) Detecting common gene expression patterns in multiple cancer outcome entities. *Biomed Microdevices* 7: 247-251. doi:10.1007/s10544-005-3032-7
  36. Aggarwal A, Guo DL, Hoshida Y, Yuen ST, Chu KM, et al. (2006) Topological and functional discovery in a gene coexpression meta-network of gastric cancer. *Cancer Res* 66: 232-241. doi:10.1158/0008-5472.CAN-05-2232
  37. DeConde R, Hawley S, Falcon S, Clegg N, Knudsen B, et al. (2006) Combining results of microarray experiments: A rank aggregation approach. *Stat Appl Genet Mol Biol* 5: Article15.
  38. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, et al. (2006) RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22: 2825-2827. doi:10.1093/bioinformatics/btl476
  39. Wang Y, Joshi T, Zhang XS, Xu D, Chen L (2006) Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics* 22: 2413-2420. doi:10.1093/bioinformatics/btl396
  40. Zintzaras E, Ioannidis JPA (2008) Meta-analysis for ranked discovery datasets: Theoretical framework and empirical demonstration for microarrays. *Comput Biol Chem* 32: 38-46. doi:10.1016/j.compbiolchem.2007.09.003
  41. Sutton A, Abrams K, Jones D, Sheldon T, Song F (2000) *Methods for meta-analysis in medical research*. New York: John Wiley & Sons.
  42. Deeks JJ, Altman DG, Bradburn MJ (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman D, editors. *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Publishing Group. pp. 285-312.
  43. Whitehead A (2002) *Meta-analysis of controlled clinical trials*. 1st edition. Chichester (United Kingdom): Wiley. 352 p.
  44. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, et al. (2003) ArrayExpress—A public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31: 68-71.
  45. Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y (2003) CIBEX: Center for Information Biology gene EXpression database. *C R Biol* 326: 1079-1082.
  46. Edgar R, Domrachev M, Lash A (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210.
  47. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME)—Toward standards for microarray data. *Nat Genet* 29: 365-371.
  48. Ball CA, Brazma A, Causton H, Chervitz S, Edgar R, et al. (2004) Submission of microarray data to public repositories. *PLoS Biol* 2: e317. doi:10.1371/journal.pbio.0020317
  49. Rhodes D, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) ONCOMINE: A cancer microarray database and integrated data-mining platform. *Neoplasia* 6: 1-6.
  50. Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, et al. (2007) The Stanford Microarray Database: Implementation of new analysis tools and open source release of software. *Nucleic Acids Res* 35: D766-D770.
  51. Suárez-Fariñas M, Noggle S, Heke M, Hemmati-Brivanlou A, Magnasco MO (2005) Comparing independent microarray studies: The case of human embryonic stem cells. *BMC Genomics* 6: 99. doi:10.1186/1471-2164-6-99
  52. Ioannidis JPA, Rosenberg PS, Goedert JJ, O'Brien TR, International Meta-analysis of HIV Host Genetics (2002) Commentary: Meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 156: 204-210.
  53. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettinger M, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5: R80. doi:10.1186/gb-2004-5-10-r80
  54. Bunes A, Huber W, Steiner K, Sültmann H, Poustka A (2005) Arraymagic: Two-colour cDNA microarray quality control and preprocessing. *Bioinformatics* 21: 554-556. doi:10.1093/bioinformatics/bti052
  55. Wilson CL, Miller CJ (2005) Simpleaffy: A bioconductor package for Affymetrix quality control and data analysis. *Bioinformatics* 21: 3683-3685. doi:10.1093/bioinformatics/bti605
  56. Bolstad B (2006) affyPLM: Methods for fitting probe-level models. R package version 1.10.0. Available: <http://bmbolstad.com/>. Accessed 4 August 2008.
  57. Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 (Suppl 1): S96-S104.
  58. Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J (2005) Independence and reproducibility across microarray platforms. *Nat Methods* 2: 337-344. doi:10.1038/nmeth757
  59. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, et al. (2005) Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2: 345-350. doi:10.1038/nmeth756
  60. Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, et al. (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2: 351-356.
  61. Consortium MAQC, Shi L, Reid LH, Jones WD, Shippy R, et al. (2006) The microarray quality control [maqc] project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24: 1151-1161. doi:10.1038/nbt1239
  62. Benito M, Parker J, Du Q, Wu J, Xiang D, et al. (2004) Adjustment of systematic microarray data biases. *Bioinformatics* 20: 105-114.
  63. Kruskal JB, Wish M (1978) *Multidimensional scaling*. Beverly Hills: SAGE Publications.
  64. Venables WN, Ripley BD (2002) *Modern applied statistics with S*. 4th edition. New York: Springer.
  65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
  66. Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, et al. (2004) An overview of ensembl. *Genome Res* 14: 925-928. doi:10.1101/gr.1860604
  67. Morris J, Yin G, Baggerly K, Wu C, Zhang L (2005) Pooling information across different studies and oligonucleotide microarray chip types to identify prognostic genes for lung cancer. In: *Methods of microarray data analysis*. 4th edition. Springer-Verlag. pp. 51-66. Available: [http://works.bepress.com/cgi/viewcontent.cgi?article=1005&context=jeffrey\\_s\\_morris](http://works.bepress.com/cgi/viewcontent.cgi?article=1005&context=jeffrey_s_morris). Accessed 4 August 2008.
  68. Carter SL, Eklund AC, Meacham BH, Kohane IS, Szallasi Z (2005) Redefinition of Affymetrix probe sets by sequence overlap with cdna microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics* 6: 107. doi:10.1186/1471-2105-6-107
  69. Wheeler D, Church D, Federhen S, Lash A, Madden T, et al. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31: 28-33.
  70. Diehn M, Sherlock G, Binkley G, Jin H, Matese J, et al. (2003) SOURCE: A unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 31: 219-223. Available: <http://source.stanford.edu/>. Accessed 4 August 2008.
  71. Tsai J, Sultana R, Lee Y, Pertea G, Karamycheva S, et al. (2001) Resourcerer: A database for annotating and linking microarray results within and across species. *Genome Biol* 2: SOFTWARE0002.
  72. Lennon G, Au-ray C, Polymeropoulos M, Soares M (1996) The I.M.A.G.E. Consortium: An integrated molecular analysis of genomes and their expression. *Genomics* 33: 151-152.
  73. Noth S, Benecke A, Systems Epigenomics Group (2005) Avoiding inconsistencies over time and tracking difficulties in applied biosystems ab1700/panther probe-to-gene annotations. *BMC Bioinformatics* 6: 307.
  74. Perez-Iratxeta C, Andrade MA (2005) Inconsistencies over time in 5probe-to-gene annotations. *BMC Bioinformatics* 6: 183. doi:10.1186/1471-2105-6-183
  75. Zeeberg BR, Riss J, Kane DW, Bussey KJ, Uchio E, et al. (2004) Mistaken identifiers: Gene name errors can be introduced inadvertently when using excel in bioinformatics. *BMC Bioinformatics* 5: 80. doi:10.1186/1471-2105-5-80
  76. Bushman BJ, Cooper H, Hedges LV (1994) Vote counting methods in meta-analysis. In: *The handbook of research synthesis*. New York: Russell Sage Foundation Publications. pp. 193-214.
  77. Venn J (1880) On the diagrammatic and mechanical representation of propositions and reasonings. *Dublin Philos Mag J Sci* 9: 1-18.

78. Breitling R, Herzyk P (2005) Rank-based methods as a non-parametric alternative of the t-statistic for the analysis of biological microarray data. *J Bioinform Comput Biol* 3: 1171-1189.
79. Fisher R (1932) *Statistical methods for research workers*. 4th edition. London: Oliver and Boyd.
80. Elo LL, Lahti L, Skottman H, Kyläniemi M, Lahtesmaa R, et al. (2005) Integrating probe-level expression changes across generations of Affymetrix arrays. *Nucleic Acids Res* 33: e193. doi:10.1093/nar/gni193
81. Cochran W (1937) Problems arising in the analysis of a series of similar experiments. *J R Stat Soc*: 102-118.
82. Fleiss JL (1993) The statistical basis of meta-analysis. *Stat Methods Med Res* 2: 121-145.
83. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. 2nd edition. New Jersey: Lawrence Erlbaum.
84. Rosenthal R (1994) Parametric measures of effect size. In: *The handbook of research synthesis*. New York: Russell Sage Foundation Publications. pp. 231-244.
85. R Development Core Team (2004) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available: <http://www.R-project.org>. Accessed 4 August 2008.
86. Lewis S, Clarke M (2001) Forest plots: Trying to see the wood and the trees. *BMJ* 322: 1479-1480.
87. Aldred MA, Morrison C, Gimm O, Hoang-Vu C, Krause U, et al. (2003) Peroxisome proliferator-activated receptor gamma is frequently downregulated in a diversity of sporadic nonmedullary thyroid carcinomas. *Oncogene* 22: 3412-3416. doi:10.1038/sj.onc.1206400
88. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, et al. (2005) Reverse engineering of regulatory networks in human b cells. *Nat Genet* 37: 382-390. doi: 10.1038/ng1532
89. Beer DG, Kardias SLR, Huang CC, Giordano TJ, Levin AM, et al. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8: 816-824. doi:10.1038/nm733
90. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, et al. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 98: 13790-13795. doi:10.1073/pnas.191502998
91. Chen X, Cheung ST, So S, Fan ST, Barry C, et al. (2002) Gene expression patterns in human liver cancers. *Mol Biol Cell* 13: 1929-1939. doi:10.1091/mbc.02-02-0023
92. Chen X, Leung SY, Yuen ST, Chu KM, Ji J, et al. (2003) Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell* 14: 3208-3215. doi:10.1091/mbc.E02-12-0833
93. Couvelard A, O'Toole D, Leek R, Turley H, Sauvaget A, et al. (2005) Expression of hypoxia-inducible factors is correlated with the presence of a fibrotic focus and angiogenesis in pancreatic ductal adenocarcinomas. *Histopathology* 46: 668-676. doi:10.1111/j.1365-2559.2005.02160.x
94. Dyrskjöt L, Kruhoffer M, Thykjaer T, Marcussen N, Jensen JL, et al. (2004) Gene expression in the urinary bladder: A common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res* 64: 4040-4048. doi:10.1158/0008-5472.CAN-03-3620
95. Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, et al. (2002) Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res* 62: 233-240.
96. Hu J, Bianchi F, Ferguson M, Cesario A, Margaritora S, et al. (2005) Gene expression signature for angiogenic and nonangiogenic non-small-cell lung cancer. *Oncogene* 24: 1212-1219. doi:10.1038/sj.onc.1208242
97. Huang Y, Prasad M, Lemon WJ, Hampel H, Wright FA, et al. (2001) Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci U S A* 98: 15044-15049. doi:10.1073/pnas.251547398
98. Jones MH, Virtanen C, Honjoh D, Miyoshi T, Satoh Y, et al. (2004) Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet* 363: 775-781. doi:10.1016/S0140-6736(04)15693-6
99. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretto G, et al. (2001) Gene expression profiling of B-cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory b cells. *J Exp Med* 194: 1625-1638.
100. Kuriakose MA, Chen WT, He ZM, Sikora AG, Zhang P, et al. (2004) Selection and validation of differentially expressed genes in head and neck cancer. *Cell Mol Life Sci* 61: 1372-1383. doi:10.1007/s00018-004-4069-0
101. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 101: 811-816. doi:10.1073/pnas.0304146101
102. Lenburg ME, Liou LS, Gerry NP, Frampton GM, Cohen HT, et al. (2003) Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data. *BMC Cancer* 3: 31. doi:10.1186/1471-2407-3-31
103. Pellagatti A, Cazzola M, Giagounidis AAN, Malcovati L, Porta MGD, et al. (2006) Gene expression profiles of cd34+ cells in myelodysplastic syndromes: Involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood* 108: 337-345. doi:10.1182/blood-2005-12-4769
104. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, et al. (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 98: 15149-15154. doi:10.1073/pnas.211566398
105. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, et al. (2002) Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1: 203-209.
106. Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, et al. (2001) Analysis of gene expression identities candidate markers and pharmacological targets in prostate cancer. *Cancer Res* 61: 5974-5978.
107. Winter SC, Buffa FM, Silva P, Miller C, Valentine HR, et al. (2007) Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* 67: 3441-3449. doi:10.1158/0008-5472.CAN-06-3322
108. Irazarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, et al. (2003) Summaries of Affymetrix genechip probe level data. *Nucleic Acids Res* 31: e15.
109. Smyth GK, Speed T (2003) Normalization of cDNA microarray data. *Methods* 31: 265-273.
110. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, et al. (2002) Normalization of cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30: e15.
111. Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc Ser B* 64: 479-498.
112. Ventura B (2005) Mandatory submission of microarray data to public repositories: How is it working? *Physiol Genomics* 20: 153-156. doi:10.1152/physiolgenomics.00264.2004
113. Larsson O, Sandberg R (2006) Lack of correct data format and comparability limits future integrative microarray research. *Nat Biotechnol* 24: 1322-1323. doi:10.1038/nbt1106-1322
114. Piwowar HA, Day RS, Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2: e308. doi:10.1371/journal.pone.0000308
115. Ioannidis JPA, Polyzos NP, Trikalinos TA (2007) Selective discussion and transparency in microarray research findings for cancer outcomes. *Eur J Cancer* 43: 1999-2010. doi:10.1016/j.ejca.2007.05.019
116. Dickersin K, Min YI, Meinert CL (1992) Factors influencing publication of research results. follow-up of applications submitted to two institutional review boards. *JAMA* 267: 374-378.
117. Egger M, Smith GD (1998) Bias in location and selection of studies. *BMJ* 316: 61-66.
118. Mondry A, Loh M, Giuliani A (2007) DNA expression microarrays may be the wrong tool to identify biological pathways. *Nature Precedings*. doi:10.1038/npre.2007.1036.1
119. Loh M, Mondry A (2007) Diagnostic robustness of DNA microarrays in the classification of acute leukemia. *Nature Precedings*. doi:10.1038/npre.2007.1056.1