Clinical Chemistry 59:2
392–400 (2013)

Data Submission and Quality in Microarray-Based MicroRNA Profiling

Kenneth W. Witwer[1]*

BACKGROUND: Public sharing of scientific data has assumed greater importance in the omics era. Transparency is necessary for confirmation and validation, and multiple examiners aid in extracting maximal value from large data sets. Accordingly, database submission and provision of the Minimum Information About a Microarray Experiment (MIAME)[3] are required by most journals as a prerequisite for review or acceptance.

METHODS: In this study, the level of data submission and MIAME compliance was reviewed for 127 articles that included microarray-based microRNA (miRNA) profiling and were published from July 2011 through April 2012 in the journals that published the largest number of such articles—*PLOS ONE*, the *Journal of Biological Chemistry*, *Blood*, and *Oncogene*—along with articles from 9 other journals, including *Clinical Chemistry*, that published smaller numbers of array-based articles.

RESULTS: Overall, data submission was reported at publication for <40% of all articles, and almost 75% of articles were MIAME noncompliant. On average, articles that included full data submission scored significantly higher on a quality metric than articles with limited or no data submission, and studies with adequate description of methods disproportionately included larger numbers of experimental repeats. Finally, for several articles that were not MIAME compliant, data reanalysis revealed less than complete support for the published conclusions, in 1 case leading to retraction.

CONCLUSIONS: These findings buttress the hypothesis that reluctance to share data is associated with low study quality and suggest that most miRNA array investigations are underpowered and/or potentially compromised by a lack of appropriate reporting and data submission.

Irreproducibility of scientific findings is a cause for ongoing concern. The majority of apparently positive results are likely to be false positives *(1)*. Confirmation of reported results was achieved for only 6 of 53 "landmark" preclinical oncology studies published in journals with impact factors of 5 or greater *(2)* and, separately, for approximately 20%–25% of 67 published studies *(3)*. With such dismal results for influential and presumably closely scrutinized studies, it is possible that opportunities for misinterpretation or misrepresentation are even greater when experimental design, data, or analyses are of low quality. The authors of a recent analysis *(4)* presented a parsimonious surrogate for poor quality: a "reluctance to share published research data," presumably stemming from investigators' "fear that reanalysis may expose errors in their work or may produce conclusions that contradict their own" *(4)*.

For publication of research that involves array technology, mandatory full data deposition is already the norm, based on the Minimum Information About a Microarray Experiment (MIAME)[2] standard *(5)*. First published in 2001 and rapidly adopted by most journals *(6)*, MIAME describes the minimum information that is needed to interpret "the results of the experiment unambiguously and potentially to reproduce the experiment" *(5)*. This includes raw data, normalized data, sample annotation, experimental design, feature descriptors, and a detailed account of preprocessing and normalization. In truth, MIAME simply "stated the obvious" *(6)*. Its guidelines should be common sense and, although designed for hybridization microarray results, the standard should also apply to quantitative PCR arrays and even next generation sequencing. In addition to interpretation and reproduction, MIAME compliance facilitates derivation of maximal scientific benefit from experiments that are often resource intensive. For example, data may be used to answer questions not addressed by the original investigators, or to inform new study design. Databases such

[1] Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD.
* Address correspondence to the author at: Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, 733 North Broadway, Suite 831, Baltimore, MD 21205. Fax 410-955-9823; e-mail kwitwer1@jhmi.edu.
Received July 26, 2012; accepted November 30, 2012.
Previously published online at DOI: 10.1373/clinchem.2012.193813
[3] Nonstandard abbreviations: MIAME, Minimum Information About a Microarray Experiment; GEO, Gene Expression Omnibus; miRNA, microRNA.

as the Gene Expression Omnibus (GEO) *(7)*, CIBEX *(8)*, and ArrayExpress *(9)* are curated, user-friendly, free-to-the-depositor options for MIAME-compliant data reporting. Of course, adherence to MIAME requires the efforts of individual editors, reviewers, and authors, and profiling studies are not uniformly MIAME compliant. Only 10 of 18 *Nature Genetics* microarray studies that were examined in a 2009 study were found to link to MIAME-compliant data *(10)*.

I noticed recently that many microRNA (miRNA) microarray–profiling publications do not reference public data, suggesting that the miRNA field may be a special area of concern. Interest in microRNAs as regulators and biomarkers of clinical conditions *(11, 12)* has led to a rapid increase of miRNA-profiling studies and funding opportunities. The influx of new investigators has been facilitated by the relatively small number of canonical miRNAs in comparison with, for example, protein-coding transcripts, combined with the availability of off-the-shelf profiling systems and mail-in services. Even experience with analysis of large data sets may appear to be an unnecessary prerequisite to miRNA profiling, because data analysis is offered by service companies, and vendors provide largely automated analysis workflows, obviating direct data manipulation by the investigator. Although access to research options is positive, black-box services and software also present potential pitfalls and may contribute to irreproducible results and confusion in the miRNA-profiling field—as in any work that involves large data sets.

To assess the current level of MIAME compliance in miRNA array studies, I reviewed articles that reported array-based miRNA profiling in the 4 journals that published the largest number of such studies during a 10-month period in 2011–2012. To provide a sampling of the wider literature as well, I examined all such articles published during a 2-week period within these 10 months (chosen because this period included a publication in *Clinical Chemistry*).

Methods

INITIAL LITERATURE SEARCH

Candidate articles were first identified by a PubMed search with these requirements: (*a*) Publication during the 10-month period from July 1, 2011, to April 30, 2012, inclusive, with publication date defined as the earliest listed date (usually e-publication date); (*b*) use of at least 1 of the following terms in the title or abstract: "miR," "miRNA," "microRNA," "small RNA," or "noncoding RNA," and use of at least 1 of the terms "array," "microarray," or "TLDA" (TaqMan low-density array) in the article text; (*c*) English language; and (*d*) primary research publication type.

CANDIDATE SCREENING

PubMed and journal websites were used to examine articles in *PLOS ONE, Blood,* the *Journal of Biological Chemistry*, and *Oncogene* to remove false positives and identify true positives that may not have appeared in the original search. For example, articles that were removed included those that were published outside the specified date range (for unknown reasons, a small number of extraneous results were returned), or false positives, containing keywords but not miRNA profiling (for example, articles that discussed miRNAs but reported transcriptome array results). All articles that were published during the 2-week period surrounding publication of an miRNA-profiling manuscript in *Clinical Chemistry* were also identified, and any articles duplicating those found above were discarded.

VALIDATED ARTICLE DATABASE

A database was created with Microsoft Excel. For each publication, the title, first author, publication date, academic editor (*PLOS ONE* only, and later removed as uninformative because most articles had a different editor), and URL were recorded, along with the following information:

(*a*) Type of miRNA profiling platform: hybridization or real-time quantitative PCR array.
(*b*) Sample: tissue, cells, body fluid.
(*c*) Validation of results by a separate method: yes or no.
(*d*) Were the data deposited in a public database? If so, the accession code was recorded.
(*e*) Did the authors specify the number of biological and technical array replicates?
(*f*) Number (or range) of biological replicates per study condition.
(*g*) Sufficient data-processing description, e.g., threshold determination, signal cutoff or background determination, QC?
(*h*) Adequate data normalization description: controls, exact normalization methods. (For example, "We normalized the data to internal controls" would be insufficient unless the internal controls were specified, their values were reported, and the exact methods of control averaging and normalization were described.)
(*i*) Sufficient description of statistical analyses to facilitate replication.
(*j*) Specification of software programs and/or contracted service companies used to generate the data and analyses.
(*k*) Use of a global normalization method.
(*l*) Use of multiple comparison correction for significance testing (or other methods appropriate for large data sets).

(*m*) Overall ruling on MIAME compliance (liberally interpreted as availability of raw and normalized data, description of technical and biological replicates, and some combination of information on data processing, normalization, and analysis): yes or no.
(*n*) Notes.

Data submission and MIAME compliance were assessed for each article as it existed at the time of publication. Note that, for some articles, authors may have since deposited or provided links to data because of postpublication requests.

ASSIGNMENT OF QUALITY SCORE

An overall quality score was given to each study and comprised 8 component scores. These scores were assigned on the basis of study characteristics and factors important for independent replication of the results (see Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue2.). Minimum and maximum possible overall scores were 0 and 19. A review of potential weaknesses of this scoring system is presented in the Discussion. The following components were used in the scoring system:

Component 1, sample size (where n was the smallest number of samples per experimental or control group; 0–5 points with with no 2-point score):
5 points: n based on a reported power calculation (no study received this score);
4 points: n = 10 or more;
3 points: n = 3 to 9 (3 is the minimum number for identification of outliers);
1 point: n = 2 replicates (minimal replication; does not allow identification of outliers);
0 points: n = 1 or not reported/no indication of replicates for all conditions.

If different numbers of replicates were included for different conditions, the lower number was used to calculate the score. For example, 2 experimental samples compared with 1 control would receive 0 points, because no meaningful biological information can be derived from this comparison.

Components 2, data processing; 3, normalization; and 4, statistical procedures, 0, 1, or 3 points each:
3 points: procedures adequately described (e.g., background correction, thresholding, exclusion criteria for data processing);
1 point: some procedures described, but incompletely or with insufficient detail to allow faithful replication;
0 points: procedures not described or inadequately described.

Component 5, software score (0 to 1):
1 point: all programs reported;
0.5 points: some reported;
0 points: not reported.

Component 6, global normalization score (0 or 1)
A point was awarded if a global normalization strategy was used. (Global normalization may be superior to normalization to just 1 or a small number of controls.)

Component 7, multiple comparison correction score (0 or 2):
Two points were awarded if statistical procedures were chosen on the basis of the presence of multiple comparisons. (It is questionable to consider an unadjusted *P* value of 0.05 "significant" when hundreds of comparisons are made.)

Component 8, validation score (0 or 1):
One additional point was awarded if some form of result validation was provided: technical validation of array results with the same samples and a different technique or biological validation with different samples.

STATISTICS

Microsoft Excel, XLStat, and GraphPad Prism were used for statistical analyses. Comparisons of multiple groups were done by 1-way ANOVA with Tukey's posttest for multiple comparisons.

ASSESSMENT OF GEO DATA SETS

Several GEO data sets, including but not limited to those of Allantaz et al. *(13)*, Chen et al. *(14)*, Mellios et al. *(15)*, and Bignami et al. *(16)*, were downloaded in full from the GEO and closely examined or reanalyzed. Additionally, postpublication requests for missing data were made for multiple articles, including but not limited to Huang et al. *(17)*, Gupta et al. *(18)*, Rahman et al. *(19)*, and Mohan et al. *(20)*. When data were made available, they were accessed and reanalyzed if possible, first by following the authors' described methods as faithfully as possible and then by different approaches, for example, as outlined in previous publications *(21–24)*.

Results

DISTRIBUTION OF ARTICLES ACROSS JOURNALS

According to my PubMed-based literature search, more than 750 research articles that reported array-based miRNA profiling were first published during t 10-month period from July 1, 2011, through April 30, 2012. Articles that reported deep sequencing were omitted because of the relatively inchoate nature of reporting standards for these studies. The journals with the largest share of miRNA array articles were *PLoS*

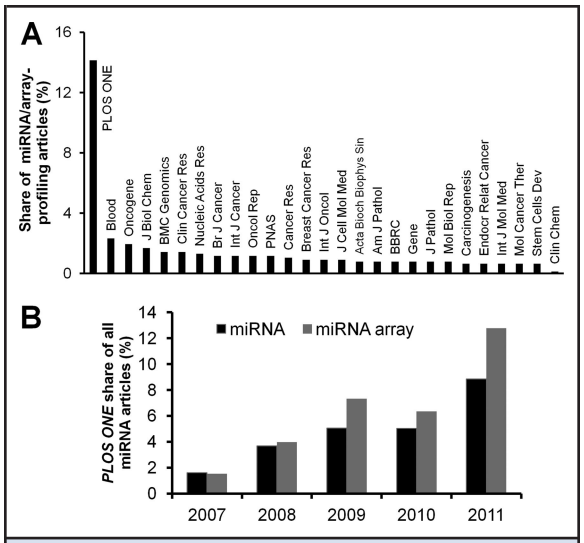

Fig. 1. Distribution of articles reporting miRNA microarray results and predominance of *PLOS ONE* in this subject area.

(A), Share of miRNA array articles for journals that published 5 or more miRNA microarray articles from July, 2011, through April, 2012, along with *Clinical Chemistry* (far right column). (B), Increasing percentage of all miRNA articles and share of miRNA array profiling studies published by *PLOS ONE*.

ONE (now officially titled and henceforth referred to as *PLOS ONE*), *Blood, Oncogene*, and the *Journal of Biological Chemistry* (Fig. 1A). As per the search results, 25 journals included 5 or more miRNA array articles. Many articles appeared in journals that published fewer than 5 such articles during the specified time period. This category included *Clinical Chemistry*, which published 1 miRNA microarray profiling study in the 10-month span.

The most miRNA array–profiling articles were published in *PLOS ONE*, with a share of over 14% of returned articles (Fig. 1A). *PLOS ONE* published more than 6 times as many miRNA array articles as the next most represented journal and approximately as many as the next 14 combined. Since the inception of *PLOS ONE* in late 2006, the journal has grown rapidly, and, in 2011, contributed over 2% of the total English-language, primary literature listings on PubMed (13 624 of 662 224). During the same time, the expansion of *PLOS ONE* miRNA articles has occurred even faster, reaching about 9% of all primary miRNA reports in 2011 (Fig. 1B). Among these reports, there has been a special concentration of array-profiling publications (Fig. 1B), because *PLOS ONE* captured almost 13% of the share in 2011 and over 14% in the 10-month period examined here.

The inexact nature of literature searches and the different reporting habits of some journals mean that the original search likely returned some false positives and eliminated some true positives. With a focus on the top 4 journals, PubMed and journal websites were used to assemble and curate lists of articles that could be verified to report on array-based miRNA profiling. For 3 journals (*PLOS ONE, Blood*, and *Oncogene*), true positives corresponded to 60%–76% of the original search results, whereas the *Journal of Biological Chemistry* was found to have published 1 article more than was indicated in the initial search. The ratios of articles in the first- and second-ranked journals, as well as in the top journal vs the next 3 combined, were unchanged.

DATA SUBMISSION AND MIAME COMPLIANCE POLICIES

All 4 of the top-publishing journals have editorial policies that mandate or strongly encourage public data submission and/or MIAME compliance (see the Journal Policies section of the online Supplemental Text file that accompanies this article). Because these journals publish more miRNA-profiling studies than most journals, the editorial staff and reviewers may be disproportionately practiced in handling such submissions. Each of these journals also has an impact factor higher than the mean for journals in the biological sciences, possibly indicating higher quality. These 3 factors might combine to skew the results of this study toward higher apparent MIAME compliance than actually exists in the wider literature, so I examined publications that appeared in journals other than the top 4 publishers during a 2-week period in August/September, 2011. This2-week period was centered on the September 2 publication date of a *Clinical Chemistry* miRNA microarray profiling report *(25)*.

Altogether, 127 verified miRNA array-reporting publications appeared in the top 4 journals or in the 9 journals from the 2-week period (see Supplemental Table 2 and the references in the online Supplemental Text file that accompanies this article); each article was reviewed and categorized on the basis of indications of raw and/or normalized data deposition with a public database. Each article was also assigned a multicomponent "quality score" as detailed in Methods, considering experimental design and assessment of components of the MIAME guidelines in addition to data submission. Where accession numbers were provided, the links were followed to examine the submission for completeness (all submissions were to either GEO or ArrayExpress). For a subset of submissions, raw and processed data were downloaded and spot-checked or reanalyzed. Finally, articles were judged to be MIAME-compliant or not.

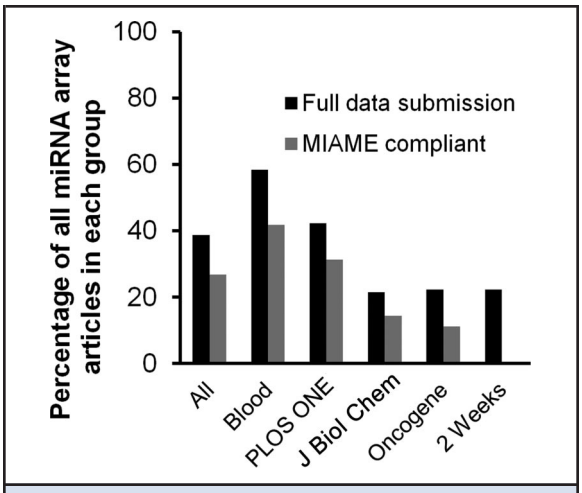

Fig. 2. Rates of data submission and MIAME compliance.

Percentage of all articles examined here ("All") or of articles in specific journals or the "2 Weeks" group, that included links to publicly available data ("Submission," black) and that were deemed MIAME-compliant (gray).

LOW LEVELS OF MIAME COMPLIANCE AND DATA SUBMISSION ACROSS JOURNALS

Of the 127 articles examined here, 93 (73%) were judged to be MIAME noncompliant. Fifteen articles (12%) included incomplete public database submissions (either raw or normalized data but not both), and 78 articles (61%) had no public database links. Of 127 articles, 34 (27%) were associated with full data set submission and were deemed MIAME compliant. Interestingly, the presence of an accession number or even a statement asserting that data were "fully MIAME compliant" *(26)* did not necessarily mean that data had been properly submitted or that the MIAME checklist had been followed. Accession numbers cited in some articles, even 1 stating that "all of the microarray data have been deposited in the Gene Expression Omnibus" *(27)* did not contain miRNA data [examples include but are not necessarily limited to *(27–29)*]. Among the 4 top-publishing journals, there were large differences in reporting and MIAME compliance (Fig. 2). Over 60% of *Blood* publications reported public database submissions, and over 40% were apparently MIAME compliant. More than 40% of articles in *PLOS ONE* reported data submission, and almost one-third appeared to satisfy MIAME. The *Journal of Biological Chemistry* had the lowest level of public reporting, at just over 20%, whereas *Oncogene* had the lowest MIAME compliance, at around 11%. Among the 2-week period group, 20% reported data accession numbers and none adhered to MIAME (Fig. 2), supporting the hypothesis that journals with a record of publishing more articles that report array results also tend to publish articles of higher quality.

ASSOCIATION OF QUALITY SCORE, IMPACT FACTOR, DATA SUBMISSION, AND MIAME COMPLIANCE

For journals represented by more than 1 article, there was no clear association between mean quality score and impact factor (Fig. 3A). It is possible that such an association exists, but that my focus on a small number of journals precluded its identification. However, there was a clear association between quality score and data submission (Fig. 3B). Although several individual exceptions were found—some articles included full data submission but scored very low on the quality scale, and several apparently high-quality and meticulously reported studies did not submit data—the highest mean quality scores were associated with full data submission and MIAME compliance, followed by partial data submission, and finally by no data submission. The mean scores were significantly different between categories (Fig. 3B, $P < 0.0001$, ANOVA), and each pairwise category comparison was also significant (Fig. 3B, $P < 0.05$, Tukey's multiple comparison test).

EXPERIMENTAL DESIGN REPORTING AND SAMPLE SIZE

Studies that involved larger sample groups also tended to include full descriptions of experimental design (Fig. 3C). The description of technical and biological replicates was recorded as adequate or vague for 85 and 42 studies, respectively; studies with no description of experimental design were also classified as "vague." Furthermore, the number of biological replicates was recorded when sufficient information was available. For approximately one-fifth of studies, it was not, and for most of these, it appeared that at least 1 examined condition had an experimental n of 1. Most the studies (36 of 42) with vague or no experimental design descriptions had only 1 experimental sample in at least 1 category (e.g., control or treated). In contrast, of the 85 articles found to provide an adequate description of technical and biological replicates, most *(58)* had an n of 3 or more, 14 had n of 1, and 13 included experimental duplicates.

EXAMPLES OF HIGH AND LOW QUALITY AND APPARENTLY ERRONEOUS CONCLUSIONS

When MIAME-compliant data were available, spot-checking of data sets generally revealed high quality and supported confidence in results. To provide an excellent example, I reanalyzed the data sets of Allantaz et al. *(13)* and closely replicated the authors' conclusions. In contrast, the true quality of MIAME-noncompliant articles may be difficult or impossible to judge when data are unavailable. However, in some cases partial data were available for review or full data

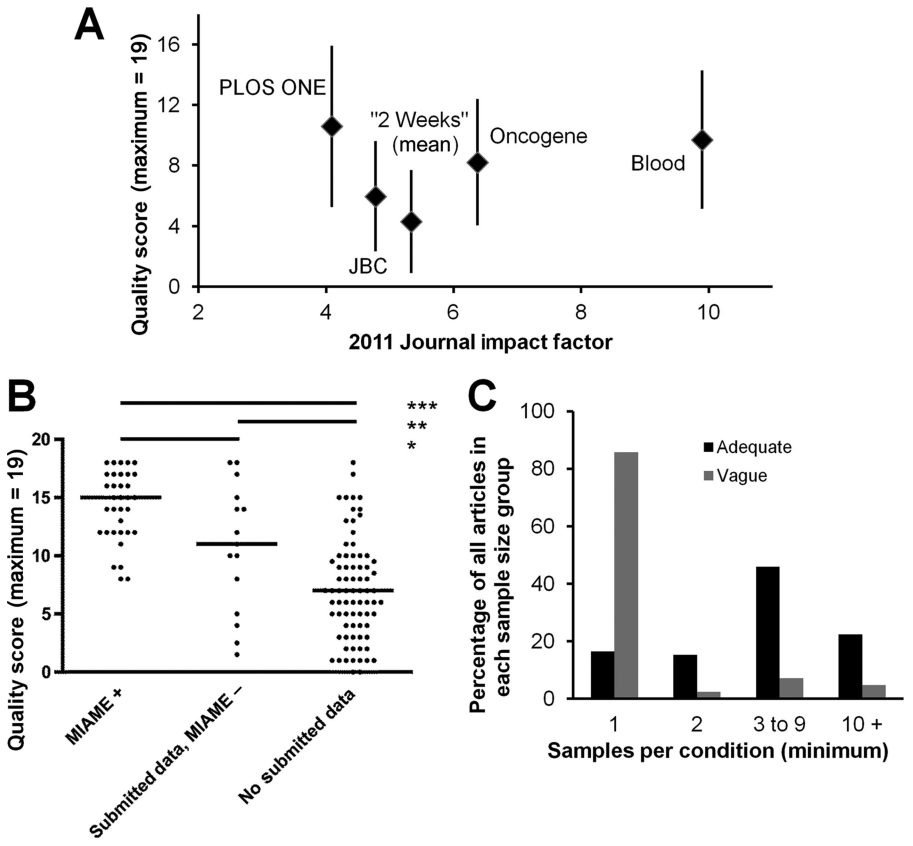

Fig. 3. Associations of impact factor, data submission, experimental design, and quality.

(A), Mean quality scores (with SD) for specific journals plotted against journal impact factor or, for the "2 weeks" group, against the group mean without the highest and lowest values. (B), The quality score was plotted for articles with full, partial, or no data submission. *P* values (*<0.05, **<0.01, ***<0.001) are from 1-way ANOVA with Tukey's post-test for multiple comparisons. (C), Studies that provided adequate descriptions of experimental design also tended to include larger numbers of experimental/biological repeats. The percentage of all articles with adequate (n = 85) or vague (n = 42) experimental design descriptions is shown by the minimum number of biological replicates. Most studies with vague descriptions relied on microarray experiments with an n of 1 in at least 1 condition.

sets were uploaded at postpublication, either at my request or at the behest of the journal. In the cases *(16, 18–20)* I examined most closely, the original data were found to provide limited support for the conclusions of the articles (see online Supplemental Text file, Section III). One of these articles has since been retracted *(18, 30)*.

Discussion

The results of this investigation indicate that levels of data reporting and MIAME compliance in miRNA array articles are cause for concern despite journal policies that mandate data submission and/or MIAME compliance as a prerequisite for review or acceptance. It is not encouraging that *Blood*, the journal with the most stringently worded policies and the best marks in this study, had just 60% data submission and approximately 40% MIAME compliance. Reporting and quality issues were found for articles in journals with impact factors ranging from approximately 1 to 30, with no obvious association between impact factor and quality score, indicating the endemic nature of the problem. However, other associations were clear. MIAME noncompliant studies were twice as likely to arise from array experiments with n of 1. Articles with vague descriptions of experimental design were disproportionately those with few experimental replicates. Studies with fully submitted data received significantly higher mean quality scores

than articles with partial submitted data or no data deposition.

For at least several articles, the quality issues may have affected the conclusions. Among articles chosen for more in-depth follow-up, 1 has since been retracted and the conclusions of others appeared to have a tenuous connection with the data. An apparent retraction rate of 1 of 83 articles without submitted data is much higher than the mean retraction rate for scientific publications; however, it is as impossible to make firm conclusions from this observation as it is to make biological conclusions from an underpowered array study. Additional monitoring is therefore needed.

This study had several weaknesses. I focused mainly on articles in the 4 journals that published the largest number of microarray-based miRNA profiles during the study period. This could skew results in favor of apparent article quality because of the higher-than-average impact factors of these journals and the relative familiarity of the journal staff with processing array submissions. These journals could also be "dumping grounds" for low-quality studies; however, no evidence to support this theory was uncovered here. By chance, the 2-week period fell at the end of the traditional summer vacation season, when relatively fewer publications appear. As a result, the small number of articles in this group may not have been representative of the wider literature. The study also examined only English-language articles, but PubMed-listed non-English articles would have represented fewer than 2% of miRNA microarray results.

My interpretation of the MIAME criteria was somewhat liberal in that I did not require complete fulfillment of every point to consider a study compliant. A more stringent interpretation would diminish even further the number of studies found to be MIAME compliant, already a minority. On the other hand, the MIAME criteria are not laws of physics, and some projects may have greater need of adherence than others. A study designed to indicate what is present in a given sample, rather than relative quantities between multiple samples, might not require strict normalization. Nevertheless, to avoid another layer of subjective judgments, I applied the same criteria to all studies.

The "quality" metric I used here to assess thoroughness of reporting and appropriate sample size, processing, and analysis may be imperfect and, in some respects, subjective. The choice of numerical values and weighting of criteria is debatable. For example, I awarded an extra point to all studies that presented some form of independent validation, but this was done whether 2 miRNAs or 40 were measured, whether in the same samples analyzed by array or in other samples. It could also be argued that the focus on array quality was unwarranted for "small n" array studies that included nonarray validation. However, because few statistically meaningful conclusions may be drawn from array studies with n of 1, performing validation studies based on such results is unlikely to be more efficient than selecting follow-up candidates randomly. Additional important factors (e.g., for 2-color hybridizations, performance of dye-swap experiments) were not considered. Finally, because I performed this study alone and over several sessions, it is possible that my application of criteria and my assignment of scores were imperfectly uniform.

Despite these weaknesses, I believe that the study is reliable and that the overall quality of miRNA microarray articles may have been overestimated because of my almost exclusive focus on the most prolific journals and my fairly liberal assessment of MIAME compliance.

In the interest of maximizing the utility of miRNA biomarker studies and the efficiency of the scientific review process, I make the following recommendations that, if implemented, might help to ensure needed improvements in the quality of miRNA microarray-based studies.

1. On the part of journals and reviewers, renewed adherence to existing data submission policies or implementation of mandatory submission policies where they do not exist. Specific endorsement of MIAME is encouraged if not already included in journal policies. Although this recommendation applies to all journals, the publishers specifically reviewed in this report—and especially *PLOS ONE*—could make the greatest contributions because of the large numbers of publications for which they are responsible.
2. At least 1 scientist with experience with large data set analysis should be involved in the review process for manuscripts reporting miRNA (and other) profiling results. This individual should verify the raw and normalized data or, ideally, perform a rapid analysis check. A review should not be considered complete until this is done.
3. On the part of researchers, the acceptance of the need for public submission of data and encouragement of maximal use of public data. This is particularly important in academic science. Unless I have personally and fully funded my laboratory and research out-of-pocket, my data do not belong to me. They belong to my institution and to the taxpayer, and I have no right to withhold them to prevent another laboratory from analyzing my data in a way I did not consider. Indeed, "integrators" of existing data—informaticists who can provide insight into what appears to be an expensively expansive morass of under- or low-powered studies—should be encouraged and fostered, as eloquently

stated in a recent plaidoyer by Andrew Moore of *BioEssays (31)*.

4. Availability or introduction of a letter-to-the-editor publication category for journals that do not already offer this publication type (e.g., *PLOS ONE*) to facilitate open, public communication about missing data and methodological information. Although online comments may be helpful in some cases, and are certainly less of an editorial burden, they carry correspondingly little weight and are not available to the reader in the same way as a letter or a formal correction.

5. Emphasis on statistically meaningful experiments. Scientists who lack experience with large data set generation and processing must recognize the need to collaborate with biostatisticians on experimental design and analysis, even when apparently attractive profiling services and vendor-supplied start-to-finish analysis software programs are available. Performing and publishing a study with an n of 1, or a study in which data are improperly processed, normalized, and analyzed, is scientifically uninformative and a waste of valuable resources, especially when precious patient samples are involved and in an era in which important public health concerns are juxtaposed with talk of funding sequestration.

6. Researchers should remain closely involved in all stages of their projects. In many cases of low quality and inadequate reporting, array-based profiling and data analysis were performed by a remote company [e.g., *(20, 32–38)*]. This arrangement, in which the (usually academic) researcher is not involved with data generation or analysis, and may not even have full access to the raw data, may be necessary in some cases. However, it also seems to create high risk for misunderstandings and errors. In addition to the communication disconnect, the goals and motivations of academic researcher and companies are simply not aligned in the same way as those of the researcher and, say, an academic collaborator or core facility.

7. *Caveat lector*. With under 40% data submission, just over 25% MIAME compliance, widespread data normalization issues, lack of multiple comparison corrections, and fully half of all experiments conducted with an n of 1 or 2, most published claims about miRNA profiles are probably erroneous and, I would predict, will not be independently verified. Exceptions might include large, well-designed cohort studies such as those reviewed by Nair et al. *(38)*. It may be wise not to draw conclusions from published miRNA profiling unless the results are independently experimentally verified or at least derived from a high-quality, publicly available data set.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *No authors declared any potential conflicts of interest.*

Role of Sponsor: No sponsor was declared.

References

1. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.
2. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. Nature 2012;483:531–3.
3. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 2011;10:712.
4. Wicherts JM, Bakker M, Molenaar D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. PloS One 2011;6:e26828.
5. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001;29:365–71.
6. Brazma A. Minimum Information About a Microarray Experiment (MIAME)–successes, failures, challenges. ScientificWorldJournal 2009;9:420–3.
7. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207–10.
8. Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y. CIBEX: center for information biology gene expression database. C R Biol 2003;326:1079–82.
9. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003;31:68–71.
10. Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, et al. Repeatability of published microarray gene expression analyses. Nat Genet 2009;41:149–55.
11. Fabbri M. miRNAs as molecular biomarkers of cancer. Expert Rev Mol Diagn 2010;10:435–44.
12. Krutovskikh VA, Herceg Z. Oncogenic microRNAs (OncomiRs) as a new class of cancer biomarkers. Bioessays 2010;32:894–904.
13. Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, et al. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. PLoS One 2012;7:e29979.
14. Chen J, Zhang X, Lentz C, Abi-Daoud M, Pare GC, Yang X, et al. miR-193b regulates Mcl-1 in melanoma. Am J Pathol 2011;179:2162–8.
15. Mellios N, Sugihara H, Castro J, Banerjee A, Le C, Kumar A, et al. miR-132, an experience-dependent microRNA, is essential for visual cortex plasticity. Nat Neurosci 2011;14:1240–2.
16. Bignami F, Pilotti E, Bertoncelli L, Ronzi P, Gulli M, Marmiroli N, et al. Stable changes in CD4+ T-lymphocyte microRNA expression following exposure to HIV-1. Blood 2012;119:6259–67.
17. Huang L, Lin JX, Yu YH, Zhang MY, Wang HY, Zheng M. Downregulation of six microRNAs is associated with advanced stage, lymph node metastasis and poor prognosis in small cell carcinoma of the cervix. PLoS One 2012;7:e33762.
18. Gupta A, Nagilla P, Le HS, Bunney C, Zych C, Thalamuthu A, et al. Comparative expression profile of miRNA and mRNA in primary peripheral blood mononuclear cells infected with human immunodeficiency virus (HIV-1). PLoS One 2011;6:e22730.
19. Rahman S, Quann K, Pandya D, Singh S, Khan ZK, Jain P. HTLV-1 tax mediated downregulation of miRNAs associated with chromatin remodeling factors in T cells with stably integrated viral promoter. PLoS One 2012;7:e34490.
20. Mohan KV, Devadas K, Sainath Rao S, Hewlett I,

Atreya C. Identification of XMRV infection-associated microRNAs in four cell types in culture. PLoS One 2012;7:e32853.
21. Sisk JM, Clements JE, Witwer KW. miRNA profiles of monocyte-lineage cells are consistent with complicated roles in HIV-1 restriction. Viruses 2012;4:1844–64.
22. Witwer KW, Clements JE. Evidence for miRNA expression differences of HIV-1-positive, treatment-naive patients and elite suppressors: a re-analysis. Blood 2012;119:6395–6.
23. Witwer KW, Sarbanes SL, Liu J, Clements JE. A plasma microRNA signature of acute lentiviral infection: biomarkers of CNS disease. AIDS 2011; 204:1104–14.
24. Witwer KW, Watson AK, Blankson JN, Clements JE. Relationships of PBMC microRNA expression, plasma viral load, and CD4+ T-cell count in HIV-1-infected elite suppressors and viremic patients. Retrovirology 2012;9:5.
25. Palacin M, Reguero JR, Martin M, Diaz Molina B, Moris C, Alvarez V, Coto E. Profile of microRNAs differentially produced in hearts from patients with hypertrophic cardiomyopathy and sarcomeric mutations. Clin Chem 2011;57:1614–6.
26. Lang MF, Yang S, Zhao C, Sun G, Murai K, Wu X, et al. Genome-wide profiling identified a set of miRNAs that are differentially expressed in glioblastoma stem cells and normal neural stem cells. PLoS One 2012;7:e36248.
27. Li Z, Huang H, Li Y, Jiang X, Chen P, Arnovitz S, et al. Up-regulation of a HOXA-PBX3 homeobox-gene signature following down-regulation of miR-181 is associated with adverse prognosis in patients with cytogenetically abnormal AML. Blood 2012;119:2314–24.
28. Saba R, Gushue S, Huzarewich RL, Manguiat K, Medina S, Robertson C, Booth SA. MicroRNA 146a (miR-146a) is over-expressed during prion disease and modulates the innate immune response and the microglial activation state. PLoS One 2012;7:e30832.
29. Yang MH, Lin BR, Chang CH, Chen ST, Lin SK, Kuo MY, et al. Connective tissue growth factor modulates oral squamous cell carcinoma invasion by activating a miR-504/FOXP1 signalling. Oncogene 2012;31:2401–11.
30. Retraction: Comparative expression profile of miRNA and mRNA in primary peripheral blood mononuclear cells infected with human immunodeficiency virus (HIV-1). PLoS One 2012;7(8).
31. Moore A. Have we produced enough results yet, Sir? BioEssays 2012;34:163.
32. Liu S, Guo W, Shi J, Li N, Yu X, Xue J, et al. MicroRNA-135a contributes to the development of portal vein tumor thrombus by promoting metastasis in hepatocellular carcinoma. J Hepatol 2012;56:389–96.
33. Tilghman SL, Bratton MR, Segar HC, Martin EC, Rhodes LV, Li M, et al. Endocrine disruptor regulation of microRNA expression in breast carcinoma cells. PLoS One 2012;7:e32754.
34. Pasha Z, Haider H, Ashraf M. Efficient non-viral reprogramming of myoblasts to stemness with a single small molecule to generate cardiac progenitor cells. PLoS One 2011;6:e23667.
35. Tano N, Kim HW, Ashraf M. microRNA-150 regulates mobilization and migration of bone marrow-derived mononuclear cells by targeting Cxcr4. PLoS One 2011;6:e23114.
36. Xu J, Kang Y, Liao WM, Yu L. MiR-194 regulates chondrogenic differentiation of human adipose-derived stem cells by targeting Sox5. PLoS One 2012;7:e31861.
37. Raychaudhuri S. MicroRNAs overexpressed in growth-restricted rat skeletal muscles regulate the glucose transport in cell culture targeting central TGF-beta factor SMAD4. PLoS One 2012; 7:e34596.
38. Biyashev D, Veliceasa D, Topczewski J, Topczewska JM, Mizgirev I, Vinokour E, et al. miR-27b controls venous specification and tip cell fate. Blood 2012;119:2679–87.
39. Nair VS, Maeda LS, Ioannidis JP. Clinical outcome prediction by microRNAs in human cancer: a systematic review. J Natl Cancer Inst 2012;104:528–40.