Reproducibility and Visual Inspection of Data

To the Editor:

In 2011, Moreau *et al.* (1) reported in *Biological Psychiatry* findings concerning patterns of microRNA (miRNA) expression in postmortem samples from prefrontal cortex (Brodmann area 9) from the Stanley Medical Research Institute (SMRI) (Chevy Chase, Maryland). They used advanced statistical techniques and miRNA assays by real-time quantitative polymerase chain reaction (2) (TaqMan probes; Applied Biosystems, Foster City, California) for the canonical sequences in early versions of miRBase (http://www.mirbase.org/) (3). All demographics, sample processing, and data analysis steps were carefully described in their article. The main conclusion was that certain miRNAs were distinguished in comparisons of samples from unaffected control subjects versus samples from subjects with schizophrenia and versus samples from subjects with bipolar disorder. They identified 24 miRNAs in particular with distinctive fold changes [Figure 1 in Moreau *et al.* (1)].

Moreau *et al.* also identified several issues with potential to cause discordant findings as different laboratories seek to identify informative miRNAs. Confounders might include variable tissue source and quality; demographics; lifetime habits of subjects regarding alcohol, tobacco, or illicit drug use; and use of medications. Moreover, different laboratories might employ different protocols for RNA isolation, different platforms for assays, and different data processing strategies at many levels starting with cleaning and normalization.

We hypothesized that visualization might identify suspicious data in some publications and, conversely, consistency among others. Graphs of raw or normalized data can be inspected for implausible anomalies. Anomalies would include sharply defined and distinct clusters of values for sets of samples that are seemingly unrelated to groups of cases and controls. An association of clusters with order of sample processing—presumably arbitrary—could also be regarded as suspicious. Markers that are chemically very similar or closely related in pathways might be expected to generate related graphs over samples. Moreover, if two markers are highly correlated over all samples except one or a few, assays in those few samples could be investigated.

As described by Moreau *et al.* (1), miRNA assays were performed on postmortem tissue from Brodmann area 9 from three groups: 35 donors who were unaffected control subjects, 35 patients with schizophrenia, and 35 patients with bipolar disorder. Two control assays and two bipolar assays were discarded from their report because of processing issues. In the present analysis, a third control assay was also set aside because of extreme values for certain miRNAs, leaving a total of 100 assays.

Visualization of the "Final Data— Delta Ct" from SMRI prompted us to question data validity. Figure 1 reveals unexpected bimodal clustering of certain miRNA values over the 100 samples. The two modes of values are unrelated to membership in the three groups but are related to sample code numbers.

Another concern in the study by Moreau *et al.* (1) is a pattern of values for certain pairs of similar miRNAs that provide some of the bimodal values in Figure 1. For example, as shown in Figure 2, two pairs of canonical mature sequences from miRBase (with distinguishing nucleotides underlined) are as follows:
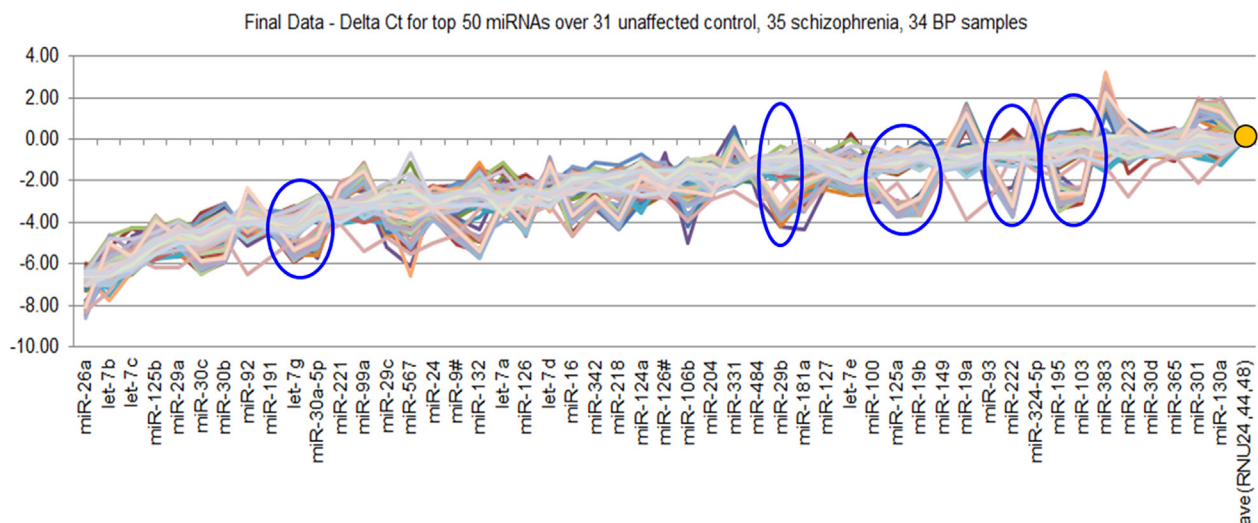

Figure 1. Cq (formerly called Ct) data from Moreau *et al.* (1) ("Final Data—Delta Ct") for 100 samples in three groups and the 50 most strongly expressed (in the 31 unaffected control subjects) miRNAs. The 51st value (orange dot, far right) represents the average of three small nucleolar RNAs used as a normalization factor and hence is exactly zero. Of concern are radically bimodal values for several miRNAs (circled in blue). The three sample groups do not coincide with the two levels in the bimodalities.
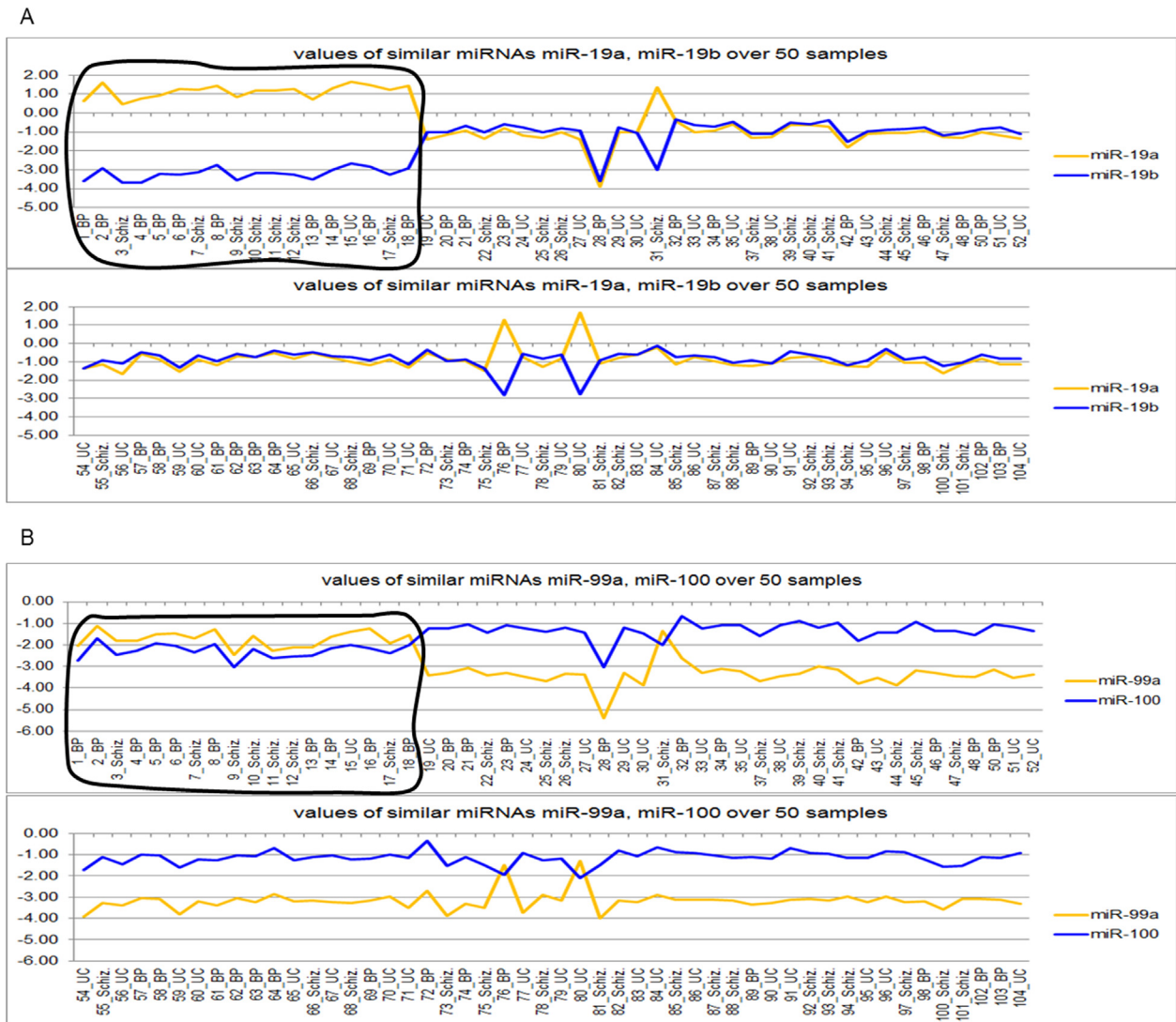
A

B

Figure 2. "Final Data—Delta Ct" (1) values for two pairs of miRNAs in order of sample code numbers. **(A)** The first 18 (ordered by code number and within black outline) are also most of the 21 samples for which miR-19a-3p, miR-19b-3p values greatly differ. Three other samples are anomalous. **(B)** Although miR-99a-5p and miR-100-5p sequences also differ by only one nucleotide, exactly the first 18 by code number are distinctive, but this time by being similar, not dissimilar. Also as in **(A)**, the same three other samples defy the pattern.

miR-19a-3p UGUGCAAAUCUAUGCAAAACUGA
miR-19a-3p UGUGCAAAUCUAUGCAAAACUGA

miR-99a-5p AACCCGUAGAUCCGAUCUUGUG
miR-100-5p AACCCGUAGAUCCGAACUUGUG

Other miRNA pairs with similar sequences (miR-132-3p, miR-212; miR-30b-5p, miR-30c-5p; and miR-103-3p, miR-107) are likewise discordant for the same 21 samples (data not shown).

To the best of our knowledge, fold change results in the study by Moreau *et al.* (1) have not been replicated. By contrast, data from two other studies (4,5) using different platforms but the same Stanley Medical Research Institute resources are quite consistent (data analyses not shown).

Concern about the reproducibility of experiments in medical research has become widespread (6). The purpose of this

correspondence is to demonstrate the utility of graphic representations of data in detection of anomalous or suspicious biomarker values. Researchers who have published biomarker studies pertaining to psychiatry would serve the interest of medical progress by rechecking their data with simple (as herein) or sophisticated visualization tools, especially when findings have not been successfully replicated.

Clark D. Jeffries
Diana O. Perkins

References

1. Moreau MP, Bruse SE, David-Rus R, Buyske S, Brzustowicz LM (2011): Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. Biol Psychiatry 69:188–193.

2. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. (2005): Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33:e179.

3. Kozomara A, Griffiths-Jones S (2014): miRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73.

4. Miller BH, Zeier Z, Xi L, Lanz TA, Deng S, Strathmann J, et al. (2012): MicroRNA-132 dysregulation in schizophrenia has implications for both neurodevelopment and adult brain function. Proc Natl Acad Sci U S A 109:3125–3130.

5. Kim AH, Reimers M, Maher B, Williamson V, McMichael O, McClay JL, et al. (2010): MicroRNA expression profiling in the prefrontal cortex of individuals affected with schizophrenia and bipolar disorders. Schizophr Res 124:183–191.

6. Collins FS, Tabak LA (2014): Policy: NIH plans to enhance reproducibility. Nature 505:612–613.