Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study

Pieter Mestdagh[1], Nicole Hartmann[2], Lukas Baeriswyl[2], Ditte Andreasen[3], Nathalie Bernard[4], Caifu Chen[4], David Cheo[5], Petula D'Andrade[6], Mike DeMayo[7], Lucas Dennis[8], Stefaan Derveaux[9], Yun Feng[5], Stephanie Fulmer-Smentek[6], Bernhard Gerstmayer[10], Julia Gouffon[7], Chris Grimley[8], Eric Lader[11], Kathy Y Lee[4], Shujun Luo[12], Peter Mouritzen[3], Aishwarya Narayanan[13], Sunali Patel[4], Sabine Peiffer[10], Silvia Rüberg[10], Gary Schroth[12], Dave Schuster[5], Jonathan M Shaffer[11], Elliot J Shelton[4], Scott Silveria[9], Umberto Ulmanella[4], Vamsi Veeramachaneni[13], Frank Staedtler[2], Thomas Peters[2], Toumy Guettouche[14] & Jo Vandesompele[1]

MicroRNAs are important negative regulators of protein-coding gene expression and have been studied intensively over the past years. Several measurement platforms have been developed to determine relative miRNA abundance in biological samples using different technologies such as small RNA sequencing, reverse transcription–quantitative PCR (RT-qPCR) and (microarray) hybridization. In this study, we systematically compared 12 commercially available platforms for analysis of microRNA expression. We measured an identical set of 20 standardized positive and negative control samples, including human universal reference RNA, human brain RNA and titrations thereof, human serum samples and synthetic spikes from microRNA family members with varying homology. We developed robust quality metrics to objectively assess platform performance in terms of reproducibility, sensitivity, accuracy, specificity and concordance of differential expression. The results indicate that each method has its strengths and weaknesses, which help to guide informed selection of a quantitative microRNA gene expression platform for particular study goals.

Intensive study in different research domains has demonstrated a prominent role for microRNAs (miRNAs) in virtually every aspect of cell biology. At the basis of these discoveries are alterations in the miRNA expression profile. More recently, miRNAs also have been detected in various body fluids[1]. Expression profiles of such miRNAs are currently being established to identify noninvasive biomarkers for human disease. Because of the small size of mature miRNAs, the high degree of homology between miRNA family members, and the low abundance of miRNAs in body fluids, miRNA expression profiling is technically challenging. Several platforms have been developed and successfully applied for quantification of miRNA expression. Unfortunately, quantifiable performance metrics for these platforms are often ill-defined or simply non-existing, which hampers informed selection of the most appropriate method, considering the particular study objectives. Published platform comparisons evaluate performance[2–6] or concordance in differential miRNA expression[7–9] but address only a limited number of platforms, RNA samples and performance parameters (**Supplementary Fig. 1**). Here we aimed to comprehensively assess quantitative miRNA gene expression platforms. We therefore initiated the microRNA quality control (miRQC) study, involving all major vendors of miRNA profiling technologies on the basis of hybridization, sequencing and RT-qPCR. All vendors agreed with the miRQC study design, sample selection and data-analysis methods. Ultimately, we included 12 platforms from 9 different vendors in the study, each platform profiling 16 (mandatory) and 4 (optional) standardized positive and negative control samples (**Fig. 1**). To enable robust and objective cross-platform comparisons irrespective of underlying technological differences, we developed performance metrics.

RESULTS

miRQC study design

We defined the samples (16 mandatory and 4 optional samples) in such a way that seven different aspects of platform performance could be evaluated, each by means of one or multiple performance metrics. Optional samples comprised RNA derived from human serum and were profiled on 9 of the 12 platforms. The miRQC study results are based on expression level information for 196 miRNAs measured by all 12 platforms (**Supplementary Table 1**). Results taking into account all miRNAs detected using an individual platform are available in platform-specific reports (**Supplementary Notes 1–12**). Each vendor also had the

[1]Center for Medical Genetics, Ghent University, Ghent, Belgium. [2]Novartis Institutes of Biomedical Research, Basel, Switzerland. [3]Exiqon, Vedbaek, Denmark. [4]Life Technologies, South San Francisco, California, USA. [5]Quanta Biosciences, Inc., Gaithersburg, Maryland, USA. [6]Agilent Technologies, Inc., Santa Clara, California, USA. [7]Affymetrix, Santa Clara, California, USA. [8]NanoString Technologies, Seattle, Washington, USA. [9]WaferGen, San Francisco, California, USA. [10]Miltenyi Biotec, Bergisch Gladbach, Germany. [11]Qiagen, Fredrick, Maryland, USA. [12]Illumina, Hayward, California, USA. [13]Strand Life Sciences, Bangalore, India. [14]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, USA. Correspondence should be addressed to J.V. (joke.vandesompele@ugent.be).

RECEIVED 27 FEBRUARY; ACCEPTED 22 MAY; PUBLISHED ONLINE 29 JUNE 2014; DOI:10.1038/NMETH.3014

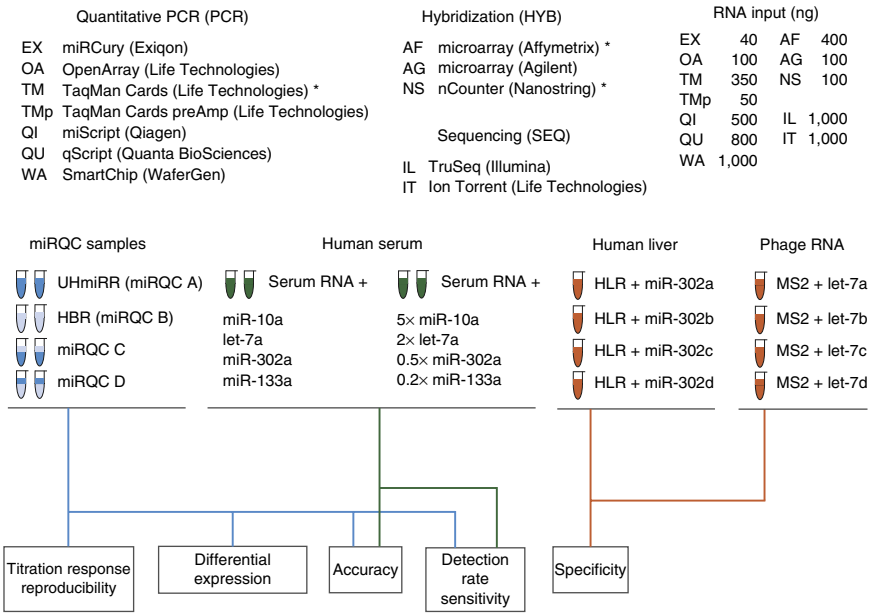

Figure 1 | MiRQC study overview. Lists of the 12 platforms used for miRNA expression profiling (top) of 16 or 20 samples (middle) to assess different aspects of platform performance (bottom). *, serum RNA samples were not measured. HLR, human liver RNA; MS2, MS2-phage RNA. Total amount of input RNA for nonserum samples is indicated for each platform (top right).

opportunity to provide platform-specific comments (**Supplementary Note 13**).

Titration response

The use of RNA sample titrations, originally proposed by the microarray quality control (MAQC) study, has become a standard approach to evaluate quantitative platform performance[10]. We used a set of samples modeled after those used in the original MAQC study. These samples (termed miRQC A–D) consist of 100% Universal Human miRNA Reference RNA (UHmiRR; A), 100% human brain RNA (HBR; B) and two titrations thereof (C = 0.75A + 0.25B and D = 0.25A + 0.75B). This titration implies that, if for gene i, $A_i > B_i$ (expression of i is higher in A than in B), then $A_i > C_i > D_i > B_i$ and conversely if $B_i > A_i$, then $B_i > D_i > C_i > A_i$. The ability to detect the correct sample order is defined as the titration response. Titration response for any given miRNA will strongly depend on the expression difference between sample A and B; a twofold higher expression in A versus B results in a difference of only 14% between samples A and C. We determined the titration response for each of the 196 commonly measured miRNAs for all platforms using only one replicate for each miRQC sample (**Fig. 1**). We binned miRNAs according to expression difference in miRQC A and B (A – B) and plotted the percentage of 'titrating miRNAs' (miRNAs for which the correct sample order is detected in the titration experiment) per bin (**Fig. 2a**). The titration response generally increased with increasing expression difference between A and B, with clear differences between platforms. To quantify these platform differences, we rescaled titration response curves, as the actual miRNA expression difference between A and B may be measured differently by the various platforms. To exclude that bias, we ranked all miRNAs based on their absolute expression difference in miRQC A and B and calculated the percentage of titrating miRNAs for increasing bin sizes, starting with a bin containing only the most differentially expressed miRNA, then added lower ranked miRNAs, until the final bin contained all miRNAs. Then we plotted the titration response as a function of the cumulative bin size with the area under the curve (AUC) as a single scale-invariant measure of platform titration response (**Fig. 2b**,**c**). If all miRNAs are titrating, AUC = 1. The AUC value clearly demonstrates differences among platforms (**Fig. 2d**). Although sequencing and hybridization technologies consistently showed high overall titration response (AUC > 0.8), results for RT-qPCR platforms were somewhat variable with TaqMan Card preamp (TMp), TaqMan Card (TM) and OpenArray (OA) showing poorest titration response (all from Life Technologies). High titration response reflects the platform's capacity to detect small expression changes, which in turn requires high reproducibility.

Reproducibility

We evaluated platform reproducibility by means of all duplicated miRQC samples (samples A–D) (**Fig. 2e**). Expression correlation plots comparing miRQC A–D to miRQC A–D replicates revealed two populations of miRNAs: those detected in both replicates (double positives) and those detected in only one replicate (single positives, visualized by imputing missing values). Single positives were associated with low-abundance miRNA expression and differed in frequency among platforms (**Fig. 2f**). We observed no association between the percentage of single positives and the range of measured miRNA expression values. To quantify the reproducibility of double positives, we plotted the cumulative distribution of absolute replicate expression differences (**Fig. 2g**,**h**). We quantified reproducibility in a single measure by calculating the area between the theoretical cumulative distribution representing perfect reproducibility (optimal curve) and the cumulative distribution of the platform. This area left of the cumulative distribution curve (ALC) decreases with increasing reproducibility and is approximated by the mean absolute replicate expression difference. Other reproducibility measures are shown in **Supplementary Figure 2**. Although there are clear differences in ALC values among platforms, the highest ALC value (lowest reproducibility) was 0.376 (TM), which corresponded to a mean replicate expression difference of only 1.3-fold. For this platform, 6.6% of all measurements had a replicate expression difference greater than twofold compared to 0% for the most reproducible platforms (Agilent microarray (AG) and WaferGen SmartChip (WA)) (**Supplementary Fig. 2**). The single positives and ALC values derive from the common set of miRNAs measured by all platforms. As this set is biased toward highly expressed miRNAs (**Supplementary Fig. 3**), the percentage of single positives and the reproducibility increase and decrease, respectively, when considering all measured miRNAs on a given platform (**Supplementary Notes 1–12**). As expected, high reproducibility significantly correlated with a better titration response

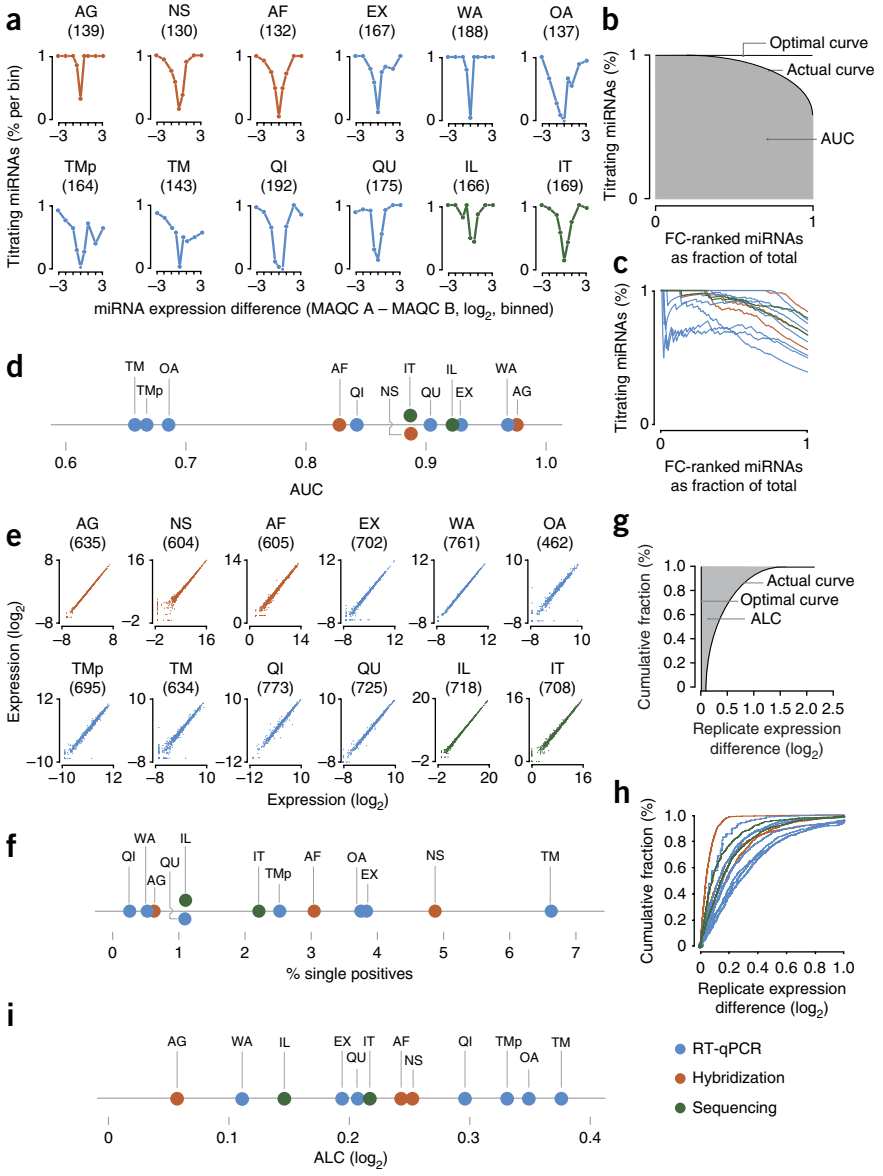

Figure 2 | Titration response and reproducibility. (**a**) Titrating miRNAs as a function of expression difference between miRQC A and miRQC B. miRNAs were binned according to expression difference between miRQC A and miRQC B. Percentage of titrating miRNAs (miRNAs for which the correct sample order is detected in the titration response experiment) was calculated per bin. Platform abbreviations are as defined in **Figure 1**, and numbers of miRNAs used for analysis (in **a**,**c**,**d**) are listed below platform abbreviations. (**b**) Schematic of titration-response data transformation. FC, fold change. (**c**) Transformed titration response data for each platform. (**d**) AUC values representing titration response for each platform. High AUC value denotes high titration response. (**e**) Correlation of miRNA expression between all four miRQC samples and their replicates. Numbers of data points used for analysis (in **h**,**i**) are listed below platform abbreviations. (**f**) Percentage of observed single positive data points for each platform. Single positives are shown in **e** by imputing the missing replicate (data imputation is described in Online Methods). (**g**) ALC as a measure for reproducibility. The point at which the cumulative distribution curve reaches 100% was set arbitrarily. (**h**) Cumulative distribution of replicate expression difference for each platform. For visualization purposes, the *x* axis is limited to 1 $\log_2$ unit. Therefore, only part of the cumulative distribution curve is shown, which explains why the curves do not reach 100% on the *y* axis. (**i**) ALC value for each platform. Low ALC value denotes high reproducibility.

($P < 0.001$, Spearman's rank test; **Fig. 2d**,**i** and **Supplementary Fig. 4a**).

Accuracy

To assess the accuracy of the expression differences, we examined those miRNAs in the miRQC samples that are exclusively expressed in either A or B. Expression of such miRNAs should be exactly threefold higher or lower, respectively, in C versus D. This threefold expression difference was underestimated by almost all platforms (**Fig. 3a**). We observed this underestimation mainly for those miRNAs that were of low abundance in A or B. This suggests that these miRNAs may not be exclusive to either A or B, but are expressed in both, albeit below the detection level in one of these samples. As expected, excluding the least abundant miRNAs (first quartile) increased the median fold difference (**Supplementary Fig. 5**). Although the median fold difference was close to that expected (threefold) for several platforms, the variation can differ (**Fig. 3a**). To quantify this, we calculated the median deviation from the expected ratio for each platform (**Fig. 3b** and **Supplementary Fig. 5**). This revealed substantial differences between platforms with deviations ranging from as low as 17% up to 42%.

To assess accurate quantification of low-abundance miRNAs, we spiked in four synthetic miRNAs at low but variable copy number in human serum RNA (**Fig. 1**). We evaluated accuracy by comparing observed (Δ_o) and expected expression difference (Δ_e) between both serum samples for each of the four miRNAs (**Fig. 3c**). With a mean 1.36-fold difference between observed and expected expression difference ($2^{(\Delta o - \Delta e)}$), qPCR platforms generally accurately quantified low-copy miRNAs. One qPCR platform (WA) performed substantially worse, displaying a mean 2.48-fold difference between observed and expected expression difference. Sequencing platforms were less accurate with a mean 1.80-fold difference, whereas the only hybridization platform to measure the serum samples accurately quantified one miRNA but failed to detect the others. Of note, certain inhibitors in serum might still be present in the isolated RNA and could have a different impact on the tested platforms. Overall, these results demonstrate differences in accuracy, both for high- and low-abundance miRNAs, and also point at differences in sensitivity between platforms.

Detection rate and sensitivity

To assess the apparent differences between platforms in terms of miRNA detection sensitivity, we first evaluated the detection rate in the miRQC samples by counting the number of unique

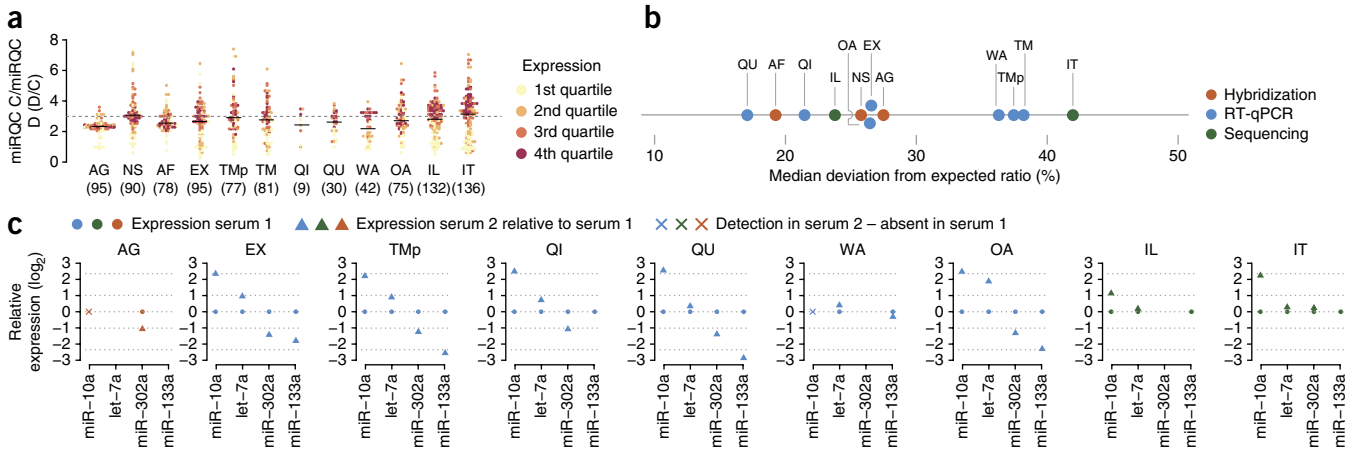

Figure 3 | Platform accuracy. (**a**) Fold change (miRQC C/D or miRQC D/C) for miRQC A or miRQC B specific miRNAs. The dotted line indicates the expected threefold expression change. Individual miRNAs are color-coded according to their expression amount. Horizontal black lines indicate median fold change. Total number of miRNAs per platform is listed below the platform identifiers (defined in **Fig. 1**). Analyses were performed using one replicate from each miRQC sample. (**b**) Median deviation from the expected C/D or D/C ratio for each platform, calculated as $(2^{\text{median}(|\log_2 3 - \log_2|C/D||)} - 1) \times 100$. (**c**) Relative mean miRNA expression values for synthetic miRNAs spiked in serum RNA samples ($n = 2$ technical replicates). Expression in serum sample 2, which contains varying concentrations of synthetic miRNAs, is shown relative to expression in serum sample 1, which contains fixed concentrations of synthetic miRNAs. Instances in which miRNA expression was only detected in serum sample 2 were set arbitrarily at 0.

miRNAs reproducibly detected (double positives) in the miRQC replicates (**Fig. 4a**). We observed lower detection rates for hybridization platforms as compared to most of the qPCR and sequencing platforms, in line with the results for the low-copy miRNA spike-in experiments (**Fig. 3c**). Then we evaluated rates of miRNA detection in the serum RNA samples by counting those miRNAs detected in at least two of four replicates (**Fig. 4b**). Detection rates in serum RNA were much more variable between platforms, with up to a 12-fold difference between the highest and lowest number of detected miRNAs. In line with the detection rates in the miRQC samples, qPCR platforms displayed higher sensitivity when quantifying serum RNA compared to hybridization and sequencing platforms, with TMp detecting the highest number of miRNAs in serum.

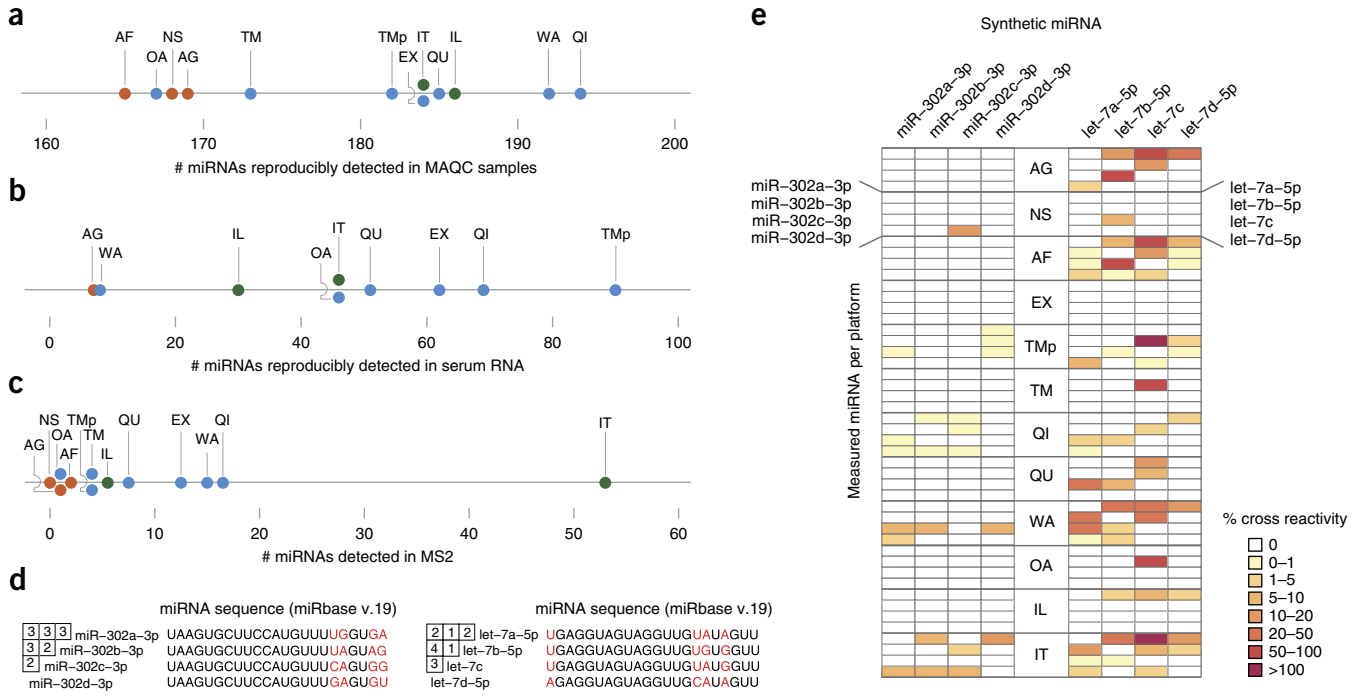

Figure 4 | Detection rate, sensitivity and specificity. (**a**) Number of miRNAs reproducibly detected in any of four miRQC samples for each platform (abbreviations defined in **Fig. 1**). (**b**) Number of miRNAs detected in at least two of four serum RNA samples for each platform. (**c**) Number of miRNAs detected in MS2-phage RNA, calculated as the median across all four MS2-phage RNA samples. (**d**) Mature miRNA sequence of miR-302 and let-7 family members. Variable nucleotides are indicated in red. The number of nucleotides that differ between each member is indicated in the matrix. (**e**) Cross-reactivity between miRNA family members measured for synthetic miRNAs (columns) and corresponding measured miRNA signals for each platform (rows). Expression values for the perfect match are not indicated in the heatmap. Cross-reactivity was calculated relative to the exact match for each miRNA.

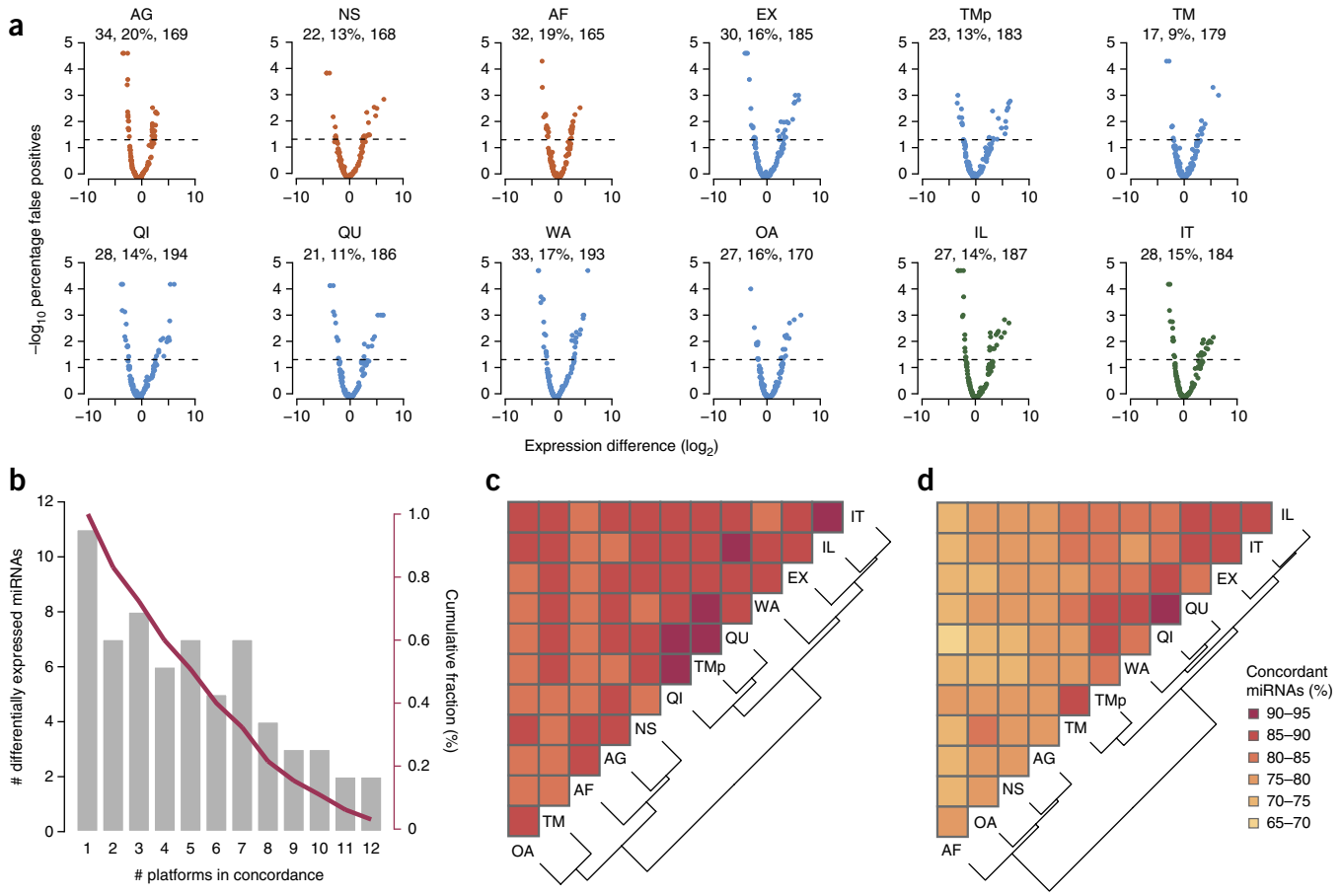

Figure 5 | Differential miRNA expression. (**a**) Volcano plot showing $-\log_{10}$ of the rank products percentage of false positives value as a function of the mean expression difference for miRQC A and C ($n = 4$) versus miRQC B and D ($n = 4$) samples. For each platform (abbreviated as in **Fig. 1**), total number of differentially expressed miRNAs, is indicated above the plot as an absolute number and as a percentage relative to the total number of miRNAs included in the analysis, also indicated above the plot. (**b**) Number of differentially expressed miRNAs identified by at least one or multiple platforms. (**c**) Hierarchically clustered heatmap indicating miRNA concordance between all platform combinations. (**d**) Hierarchically clustered heatmap indicating miRNA concordance between all platform combinations, taking into account detection rate.

Specificity

We analyzed platform specificity by means of MS2-phage RNA, which does not contain any miRNAs (**Fig. 4c**). Opposite to what we observed for sensitivity, hybridization platforms displayed higher specificity, as evidenced by the low detection rate of miRNAs in MS2-phage RNA. In fact, we observed a significant inverse correlation between background detection rate in MS2-phage RNA samples and miRNA detection rate in miRQC samples (**Supplementary Fig. 4a**, Spearman's rank $P < 0.05$, rho = −0.77), which suggested that increased detection rate might, in part, be nonspecific. The exceptionally high miRNA detection rate for the Ion Torrent (IT; Life Technologies) platform might be related to library contamination and should therefore be interpreted with caution (**Supplementary Fig. 6**).

To further evaluate specificity, we spiked in 8 synthetic miRNAs from two miRNA families into human liver RNA (miR-302) or MS2-phage RNA (let-7) (**Fig. 1**). Sequence differences between the let-7 family members vary from one to four nucleotides while miR-302 family members have a two or three nucleotide difference (**Fig. 4d**). A cross-reactivity heatmap displaying signal intensity for mismatched miRNA combinations relative to signal intensity of the perfect match demonstrates major differences between platforms and between both miRNA families (**Fig. 4e**). One platform (miRCury (EX); Exiquon) showed absolute specificity for both miRNA families. Whereas most platforms showed little or no cross-reactivity between miR-302 family members, cross-reactivity between let-7 family members was markedly higher and predominantly occurred between members differing in only one nucleotide. In some instances, the measured signal for the mismatch approximated or even surpassed that of the match. For sequencing platforms, cross-reactivity can be impacted by the number of allowed mismatches during read mapping. Mapping of TruSeq (IL; Illumina) reads using 0 instead of 1 mismatch reduced both the frequency and the level of cross-reactivity between let-7 family members (**Supplementary Fig. 7**).

Differential expression

Ultimately, the goal of most miRNA expression profiling studies is to quantify differences between sample groups. To evaluate the differential expression concordance among individual platforms, we identified differentially expressed miRNAs between two sample groups, miRQC A and miRQC C (group 1), and miRQC B and miRQC D (group 2), for each platform (**Fig. 5a**). We selected a total of 66 miRNAs, identified as differentially expressed by at least one platform, and evaluated platform concordance (**Fig. 5b**). To our surprise, only two miRNAs (3%) were differentially

expressed by all platforms; about half of the miRNAs (48%) were concordant for half of the platforms. We then calculated miRNA concordance between any two platform combinations. In a first analysis, a miRNA was scored concordant when both platforms agreed on the differential expression status. Hierarchical clustering of the concordance matrix revealed two main clusters, one consisting only of qPCR and sequencing platforms and another containing qPCR and hybridization platforms (**Fig. 5c**). Concordance varied between 80% and 95%, and was higher in the cluster containing the sequencing platforms. In a second analysis, detection rate was taken into account and miRNAs that were detected by only one of both platforms were scored as non-concordant. Hierarchical clustering again revealed two clusters similar in content as in the first analysis (**Fig. 5d**). However, concordance was much lower, down to 70%, especially among the platforms in the cluster containing only hybridization and qPCR platforms. In a typical miRNA expression study, results obtained by one platform are often validated by means of another platform. In this study, the average concordance between any two platforms was 86.7% (95% CI, 86.0–87.3%), or 79.2% (95% CI, 77.0–80.4%) when we took into account the detection rate. However, in most cases, validation studies only focused on the differentially expressed miRNAs. For those, the average validation rate between any platform combination was only 54.6% (95% CI, 52.5–56.7%). To evaluate platform recall rates (number of truly differentially expressed miRNAs retrieved per platform), we defined truly 'differential' miRNAs (49 miRNAs) as those called differentially expressed by at least two different technologies (PCR, hybridization and sequencing). Recall rates varied from 35% (TM) to 63% (AG and WA) and were significantly correlated to reproducibility and titration response ($P < 0.05$, Spearman's rank test; **Supplementary Fig. 8**).

DISCUSSION

Our choice of calculating quality metrics using the 196 miRNAs measured on all platforms has implications for several of the metrics, most notably detection rate. In the end, detection rate is a combination of sensitivity, detection cutoff and platform content (number of assays or probes available on the platform). Although sequencing platforms theoretically have unlimited content, hybridization and qPCR platforms are limited to the number of probes or assays available, which in turn is limited by the content of miRBase. Nevertheless, content can only be exploited if sensitivity is high enough, exemplified by the higher overall detection rates for qPCR versus hybridization platforms. It did not surprise us that detection rates for sequencing platforms increased with increasing sequencing depth (**Supplementary Fig. 9**). Assessment of sensitivity should also take into account the amount of input RNA required as this varies substantially between platforms and could be a limiting factor when choosing a platform. The apparent bias of the common set toward more highly expressed miRNAs also affects metrics that depend heavily on expression level such as reproducibility and titration response. Of note, differences in reproducibility between platforms could also be attributed to the presence of replicate probes or assays as evidenced by some of the platforms that incorporate replicate probes or assays (AG and WA) that show exceptional reproducibility. We only evaluated titration response using high-RNA-content samples. Therefore, results might not apply for samples with limited RNA content such as body fluids.

Our approach to use titration samples to quantify accuracy is cost-effective as it does not rely on synthetic miRNA pools (often used to evaluate accuracy) and can be applied for any type of transcript, coding or noncoding. Nevertheless, it is limited to those transcripts that are exclusively expressed in either miRQC A or miRQC B and detected well above background. Low-abundance transcripts might lead to underestimation of the true difference between miRQC C and miRQC D. We frequently observed overestimation of this difference for many platforms. This overestimation can only be explained by a difference in total miRNA content between miRQC A and miRQC B. Differences in miRNA content among different tissue types and different RNA content between the original MAQC samples have been documented previously[10–11] (C. Tissot, Analysis of miRNA content in total

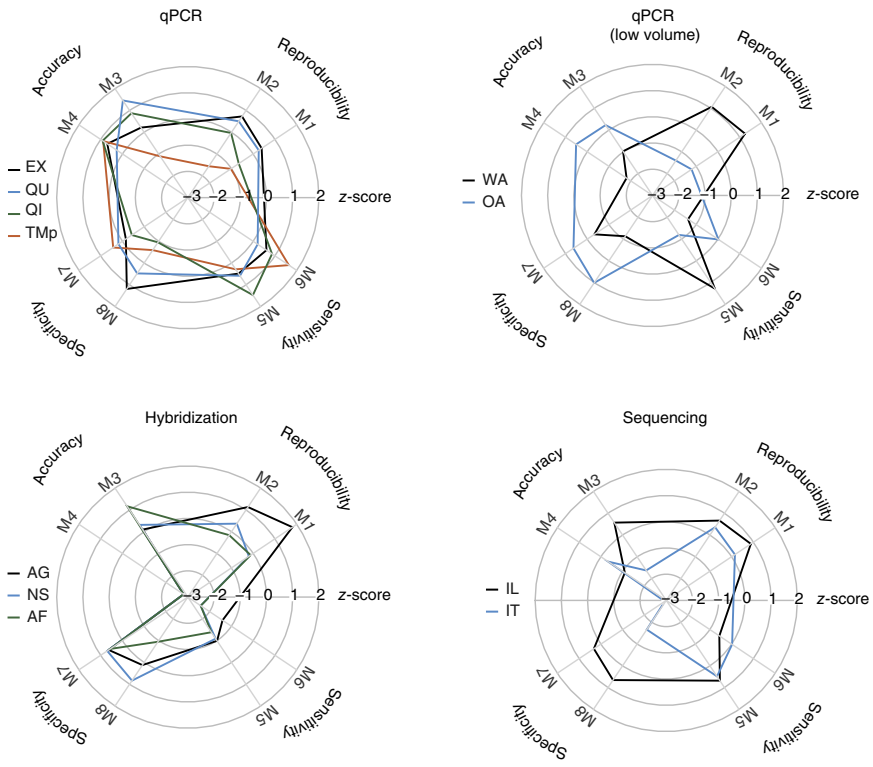

Figure 6 | Radial plot of performance metric *z*-scores. Platforms are divided into four technologies: qPCR, low-volume qPCR, hybridization and sequencing (as TM and TMp are essentially based on the same technology, only TMp is shown). *Z*-scores for eight metrics are shown. M1, reproducibility; M2, titration response; M3, accuracy; M4, accuracy low-input RNA; M5, sensitivity; M6, sensitivity low-input RNA; M7, specificity MS2 RNA; M8, assay cross-reactivity. Metrics were transformed in such way that a higher metric value (*z*-score) corresponds to a better performance (Online Methods). Each radial plot has an identical scale, which makes plots directly comparable.

RNA preparations using the Agilent 2100 bioanalyzer; 2008) and should be examined in more detail to better understand some of these observations.

Some of the calculated metrics are significantly correlated (**Supplementary Fig. 4a**). We expected a number of these correlations (e.g., titration response and reproducibility), but others, such as the inverse correlation between sensitivity and specificity (when specificity was evaluated by means of background detection in MS2) surprised us.

When looking at platform correlation based on all metrics, we observed no obvious clustering of technologies (**Supplementary Fig. 4b**). We also observed substantial interplatform differences when evaluating differential miRNA expression. As different technologies are often applied for validation purposes, the choice of platform could dramatically impact the validation rate. With an average validation rate for differentially expressed miRNAs of only 54.6% between any two platform combinations, we strongly advise that screening studies are followed by targeted validation using an alternative platform or technology. Data from this study could be applied in future studies to model other factors that might impact differential expression analysis, such as normalization[9] and data-imputation methods. To help guide platform choice, we grouped platform performance measures into four categories that reflect different research questions: reproducibility, specificity, sensitivity and accuracy for experiments with high and low RNA input amount (**Fig. 6**). We transformed each metric into a *z*-score, which allowed direct platform comparison. This analysis demonstrated that platforms based on the same technology can have very different performance. This is most obvious for reproducibility and specificity among qPCR platforms. In contrast, sensitivity is very much technology-related with qPCR platforms having an overall better score, especially when it comes down to low input RNA samples (with the exception of OA and WA). This superior sensitivity is accompanied by high accuracy, which results in reliable quantitative measurements. Hybridization-based platforms show lower sensitivity, even when input RNA is not limiting, whereas sequencing platforms are sensitive when RNA is not limiting but lose sensitivity for low-input-amount RNA samples. Few platforms can capture small expression differences (reflected by platform reproducibility and titration response), with the top three platforms being AG, WA and IL; this feature appears to be independent of technology. When working with low-input-amount RNA samples such as body fluids, these metrics might be secondary to sensitivity, especially if the goal of the study is detection rather than accurate quantification of miRNAs.

Although we expected some of these results, unexpected findings included (i) low specificity for several platforms, (ii) low concordance of differential expression, (iii) poor titration response and reproducibility for several qPCR platforms, which implied that these qPCR platforms should not be used to quantify small changes in expression in small size sample cohorts, (iv) the observation that some performance parameters are technology-related (PCR, hybridization or sequencing), whereas others are platform-related, and (v) the strong and significant inverse correlation between sensitivity and specificity (e.g., platforms that detect a lot of miRNAs also detect a lot of signal in the negative control MS2 samples). Each platform has specific strengths and weaknesses, which suggests platform should be chosen on the basis of the experimental setting and the specific research questions.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus: GSE49453 (Agilent miRNA expression data), GSE49661 (Affymetrix miRNA expression data), GSE49600 (NanoString miRNA expression data) and GSE49816 (Illumina miRNA expression data). ArrayExpress: E-MTAB-1815 (Ion Torrent miRNA expression data).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Nuytens for preparing all RNA samples and E. Lefebvre and T. Demoor for statistical and mathematical support. P. Mestdagh is supported by the Fund for Scientific Research Flanders (FWO); J.V. is supported by the Ghent University Research Fund (BOF).

AUTHOR CONTRIBUTIONS

P. Mestdagh, J.V. and T.G. designed the study; P. Mestdagh and J.V. performed data analysis and wrote the manuscript; and N.H., L.B., D.A., N.B., C.C., D.C., P.D., M.D., L.D., S.D., Y.F., S.F.-S., B.G., J.G., C.G., E.L., K.Y.L., S.L., P. Mouritzen, A.N., S. Patel, S. Peiffer, S.R., G.S., D.S., J.M.S., E.J.S., S.S., U.U., V.V., F.S. and T.P. produced miRNA expression data. All authors approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Cortez, M.A. *et al.* MicroRNAs in body fluids–the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* **8**, 467–477 (2011).
2. Jensen, S.G. *et al.* Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. *BMC Genomics* **12**, 435 (2011).
3. Wang, B. *et al.* Systematic evaluation of three microRNA profiling platforms: microarray, beads array, and quantitative real-time PCR array. *PLoS One* **6**, e17167 (2011).
4. Kelly, A.D. *et al.* Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance. *Genomics* **102**, 8–14 (2013).
5. Kolbert, C.P. *et al.* Multi-platform analysis of microRNA expression measurements in RNA from fresh frozen and FFPE tissues. *PLoS One* **8**, e52517 (2013).
6. Leshkowitz, D., Horn-Saban, S., Parmet, Y. & Feldmesser, E. Differences in microRNA detection levels are technology and sequence dependent. *RNA* **19**, 527–538 (2013).
7. Pradervand, S. *et al.* Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs. *Biotechniques* **48**, 219–222 (2010).
8. Git, A. *et al.* Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* **16**, 991–1006 (2010).
9. Meyer, S.U., Kaiser, S., Wagner, C., Thirion, C. & Pfaffl, M.W. Profound effect of profiling platform and normalization strategy on detection of differentially expressed microRNAs–a comparative study. *PLoS ONE* **7**, e38946 (2012).
10. Shippy, R. *et al.* Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat. Biotechnol.* **24**, 1123–1131 (2006).
11. Dittmar, K.A., Goodenbour, J.M. & Pan, T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* **2**, e221 (2006).

ONLINE METHODS

RNA samples. Universal Human miRNA reference RNA (Agilent Technologies, #750700), human brain total RNA (Life Technologies, #AM6050), human liver total RNA (Life Technologies, #AM7960) and MS2-phage RNA (Roche, #10165948001) were diluted to a platform-specific concentration. RNA integrity and purity were evaluated using the Experion automated gel electrophoresis system (Bio-Rad) and Nanodrop spectrophotometer. All RNA samples were of high quality (miRQC A: RNA quality index (RQI, scale from 0 to 10) = 9.0; miRQC B: RQI = 8.7; human liver RNA: RQI = 9.2) and high purity (data not shown). RNA was isolated from serum prepared from three healthy donors using the miRNeasy mini kit (Qiagen) according to the manufacturer's instructions, and RNA samples were pooled. Informed consent was obtained from all donors (Ghent University Ethical Committee). Different kits for isolation of serum RNA are available; addressing their impact was outside the scope of this work. Synthetic miRNA templates for let-7a-5p, let-7b-5p, let-7c, let-7d-5p, miR-302a-3p, miR-302b-3p, miR-302c-3p, miR-302d-3p, miR-133a and miR-10a-5p were synthesized by Integrated DNA Technologies and 5′ phosphorylated. Synthetic let-7 and miR-302 miRNAs were spiked into MS2-phage RNA and total human liver RNA, respectively, at 5×10^6 copies/μg RNA. These samples do not contain endogenous miR-302 or let-7 miRNAs, which allowed unbiased analysis of cross-reactivity between the individual miR-302 and let-7 miRNAs measured by the platform and the different miR-302 and let-7 synthetic templates in a complex RNA background. Synthetic miRNA templates for miR-10a-5p, let-7a-5p, miR-302a-3p and miR-133a were spiked in human serum RNA at 6×10^3 copies per microliter of serum RNA or at 5-times higher, 2-times higher, 2-times lower and 5-times lower concentrations, respectively. All vendors received 10 μl of each serum RNA sample.

miRNA expression profiling. Ten vendors of miRNA expression profiling platforms accepted the invitation to participate in the miRQC study (Affymetrix, Agilent, Exiqon, Illumina, Life Technologies, Nanostring, Qiagen, Quanta Biosciences, Toray and WaferGen). Each received 20 anonymized RNA samples and was asked to profile at least 384 miRNAs according to their standard procedures. RNA concentration was adjusted according to the standard specifications for each platform. One vendor withdrew its data (Toray), and this data set was excluded from the study.

Agilent microarray (AG). The sample labeling and hybridization was performed according to the Agilent miRNA microarray system with miRNA complete labeling and hyb kit protocol v.2.4 (Agilent publication G4170-90011). The assay was performed with the use of the microRNA spike-ins as described in the miRNA complete labeling and hyb kit protocol v.2.4 using the Agilent microRNA spike-in kit (PN 5190-1934). This protocol required the use of 100 ng of total RNA not derived from body fluids. Each 100-ng (2 μl, 50 ng/μl) sample was labeled using components from the Agilent miRNA complete labeling and hyb kit (PN 5190-0456). For the serum samples, 10 μl of RNA was dried and resuspended in 2 μl of nuclease-free water. Once resuspended, the serum samples were labeled in accordance to the protocol just as the 100-ng total RNA samples. A hybridization solution for each of the labeled samples was prepared using the components of the gene expression hybridization kit (PN5188-5242), and samples were hybridized to Agilent human microRNA microarrays on the basis of the miRBase v 16.0 release (microarray design ID 031181; PN G4870A). The microarrays were hybridized rotating in a 55 °C oven for 20 h. After hybridization, the microarrays were disassembled and washed in the Agilent gene expression wash buffer 1 at room temperature for 5 min and then washed in preheated Agilent gene expression wash buffer 2 (Agilent gene expression wash buffer kit; PN 5188-5327) at 37 °C for another 5 min. The dried microarray slides were scanned using the Agilent microarray scanner system. The data captured on the scanned images were extracted using the Agilent feature extraction software. The Feature Extraction files were loaded into GeneSpring GX software where the data were $\log_2$ transformed.

Affymetrix custom microarray (AF). An amount of 400 ng RNA per sample was labeled without amplification using the Affymetrix FlashTag Biotin HSR kit (PN 901911). The hybridization cocktails were prepared using the Affymetrix hybridization, wash and stain kit (PN 900720). The entire labeled sample was then hybridized on an Affymetrix custom HsMir-v1s520779F microarray (HsMiR-v1s520779F_rev1 Part No. 520779 Rev 1. Affymetrix Inc. and Chip Library File (1LQ) Chip Library File Format Specification 3.0) for 42 h at 45 °C in a rotating oven. The HsMir-v1s520779F custom array is primarily based on the commercially available Affymetrix microRNA microarray version 2.0 (Affymetrix GeneChip miRNA 2.0 Array datasheet P/N EXP00180 Rev 1) with (i) microRNA probe sequences updated using miRBase version 17, (ii) all human pre-microRNAs removed, (iii) all non-human sequences removed and (iv) all viral target sequences removed. In summary, the array contains 1,738 mature microRNAs and 2,333 other small RNAs (such as small nucleolar RNAs or small cajal body–specific RNAs), probe sequences plus the Affymetrix hybridization and normalization control probe sets. The microarrays were washed and stained, respectively, on the Affymetrix fluidics station 450 and scanned in a 3000 7G scanner. The scanned image's features were converted into numerical values of the probe intensity (signal) and stored as a 'CEL-file' data using the Affymetrix GeneChip Command Console software version 3.0.1. The array.CEL files ran through the Affymetrix Power Tools (APT) version 1.14.3 "probeset-summarize" function. APT is a set of command line programs that applies standard algorithms used to analyze Affymetrix microarrays. The microarray data are prepared for further analysis using the APT "rma-bg" function that performs an RMA style background adjustment as described[12,13]. The perfect match option (perfect match only, Pm-only) then uses the unmodified Pm intensity values of the probes and calculates the median value. As a consequence, the median values of each of the probesets represent the summarized expression values of the transcripts for a particular chip.

Nanostring nCounter (NS). 2 μl (100 ng) of each sample was prepared according to the manufacturer's instructions. Mature miRNAs were ligated to a species-specific tag sequence (miRtag) via a thermally controlled splinted ligation. After enzymatic purification of unligated miRtags, prepared samples were hybridized with an nCounter Human (V2) miRNA Expression Assay CodeSet overnight at 65 °C. Unhybridized CodeSet was removed via an automated purification performed on an nCoutner Prep Station,

doi:10.1038/nmeth.3014

and resulting target:probe complexes were deposited and bound to a imaging surface as previously described[14]. Reporter counts were tabulated for each sample by the nCounter Digital Analyzer and output as raw data in .csv format. Raw data was imported into nSolver (http://www.nanostring.com/products/nSolver).

Internal negative control probes included in each assay were used to determine a background threshold (3 s.d. above the mean negative control probe count value) for each sample. Background was subtracted from raw count values for each probe and set to 1 for all probes at or below the background threshold. Positive control count values were then used to normalize samples for any differences in sample preparation, hybridization and Prep Station processing efficiency.

Exiqon miRCury (EX). For samples 1–16, 2 µl RNA of 20 ng/µl was reverse-transcribed in 40-µl reactions. For samples 17–20 (serum) 10 µl RNA was reverse-transcribed in 75-µl reactions. Reverse transcription (RT) was performed using the miRCURY LNA Universal RT microRNA PCR, Polyadenylation and cDNA synthesis kit (Exiqon). cDNA was diluted 100× (samples 1–16) or 50× (samples 17–20) and assayed in 10 µl PCR reactions according to the protocol for miRCURY LNA Universal RT microRNA PCR; each microRNA was assayed once by qPCR on the microRNA Ready-to-Use PCR, human panel I and panel II. The amplification was performed in a LightCycler 480 Real-Time PCR System (Roche) in 384-well plates. The amplification curves were analyzed using the Roche LC software, both for determination of the quantification cycle (Cq; by the second derivative method) and for melting-curve analysis. The amplification efficiency was calculated using algorithms similar to the LinRegPCR software. All assays were inspected for distinct melting curves, and the melting temperature was checked to be within known specifications for the assay. Furthermore, assays must be detected with Cq < 37 to be included in the data analysis. Data that did not pass these criteria were omitted from any further analysis. MiRNA expression data are available in rdml format[15] (**Supplementary Data 1**).

Life Technologies TaqMan Array Human MicroRNA cards (TM and TMp). Megaplex Primer Pools A and B v3.0 (PN 4444750) were used in conjunction with the matching TaqMan MicroRNA Array Cards (PN 4444913) to analyze the samples. Each sample was run separately with pool A and B according to the recommended protocol (Life Technologies Application Note: "Optimized blood plasma protocol for profiling human miRNAs using the OpenArray Real-Time PCR System," 2011, PN 4399721 rev C). The cDNA for MAQC samples A, B, C and D was prepared using 350 ng total RNA per Megaplex pool. For samples that included the preamplification step, only 50 ng total RNA was used. The reverse transcriptase (RT) reaction was performed using TaqMan microRNA Reverse Transcription Kit (PN 4366596) and Megaplex RT primer pools A or B in 7.5 µl final volume. The RT reaction was thermal cycled (2 min at 16 °C, 1 min at 42 °C, 1 s at 50 °C, for 40 cycles) and the enzyme inactivated at 85 °C for 5 min. 6 µl of RT product was diluted 1:150 in 1× TaqMan Universal Master Mix II (PN 4324018) and loaded onto a TaqMan miRNA Array Card (A or B). For samples that included the preamplification step, 2.5 µl of the RT reaction was combined with the matching Megaplex PreAmp Primer Pool and TaqMan PreAmp Master Mix (PN 4391128) in a final volume of 25 µl. Preamplification was run using the following cycling conditions: 10 min at 95 °C; 2 min at 55 °C; 2 min at 72 °C; 15 s at 95 °C, 4 min at 60 °C for 12 cycles; 99 °C for 10 min. The preamplification product was diluted 1:4 in TE then diluted 1:100 in 1× TaqMan Universal Master Mix II before loading on the matching TaqMan MicroRNA Array Card. The TaqMan Array Cards were spun, sealed and then run on a 7900 HT Real Time PCR system with TaqMan Array Block using universal cycling conditions (10 min at 95 °C; 15 s at 95 °C, 1 min at 60 °C, 40 cycles). For serum samples, a modified protocol that included a preamplifcation step was used (LTC Application Note: "Optimized blood plasma protocol for profiling human miRNAs using the OpenArray Real-Time PCR System"; 2011). Briefly, 3 µl of serum total RNA sample was reverse-transcribed with the Megaplex RT Primer Pool in a 10-µl volume as described above. The entire RT reaction was used in a 50-µl preamplification reaction. The preamplification reaction was run under the conditions described above except it was cycled for 14 cycles. The final preamplification product was diluted 1:100 in 1× TaqMan Universal Master Mix II then loaded onto the matching TaqMan MicroRNA Array Card. Raw data files (.sds) were imported and analyzed in Expression Suite, a software data analysis tool that can quickly analyze large sets of data files (Life Technologies), using threshold setting at 0.2 and auto baseline. MiRNA expression data are available in rdml format (**Supplementary Data 1**).

Life Technologies TaqMan OpenArray Human MicroRNA panel (OA). Reverse transcription (RT) and preamplification was performed on all samples using Megaplex Primer Pools A and B v3.0 (PN 4444750) with the recommended protocol (LTC publication PN 4461306 Rev. B) for TaqMan OpenArray microRNA panel (PN 4461104). The cDNA was prepared with 100 ng of total RNA per pool using TaqMan MicroRNA Reverse Transcription Kit (PN 4366596) and Megaplex RT Primer Pools in 7.5 µl final volume. The cycling condition for RT is the same as mentioned above. Preamplification for all nonserum samples was performed using 2.5 µl of RT product in a final volume of 25 µl and was run under the same cycling condition as TaqMan Array Card. For serum samples, 3 µl was used in the reverse transcriptase reaction in a 7.5-µl volume. The entire 7.5 µl of RT product per pool was used for preamplification in a final volume of 40 µl. 16 cycles of preamplification were performed on serum samples compare to 12 cycles for all other samples. The preamplified product for each pool was diluted 1:40 in TE and then diluted 1:2 with TaqMan OpenArray Real-Time PCR Master Mix (PN 4462164). 5 µl of the diluted reaction mixture for each samples were aliquoted into each of eight wells on an OpenArray 384 well sample plate where each well corresponds to one subarray on OpenArray Plate. (Refer to the plate loading section; protocol PN 4461306 Rev. B.) OpenArray Plates are loaded with the help of automation process AccuFill using standard AccuFill method (OpenArray AccuFill System User Guide PN 4456986). A TaqMan OpenArray MicroRNA Panel can accommodate three samples that cover the full microRNA profile with Megaplex primer pools A and B. Three OpenArray plates can be run at one time on OpenArray NT cycler using appropriate .tpf file corresponding to the barcode for each plate. Raw data files (.cvs) were imported and analyzed in DataAssist software (Life Technologies). Global mean normalization

was used to calculate relative fold changes. MiRNA expression data are available in rdml format (**Supplementary Data 1**).

Qiagen miScript (QI). cDNA synthesis and real-time PCR were performed using the miScript PCR system. cDNA was synthesized from either 450–500 ng of total RNA (for cellular samples) or 5 µl of RNA eluate (for cell-free or serum samples) using the miScript II RT Kit with HiSpec buffer. The reactions were incubated for 60 min at 37 °C followed by a heat-inactivation step for 5 min at 95 °C. Each 20-µl cDNA synthesis was then diluted to a final volume of 110 µl using RNase-free water, and the diluted cDNA (1 µl per 4.1 wells) was analyzed using the miScript SYBR Green PCR Kit, consisting of QuantiTect SYBR Green PCR Master Mix and miScript Universal Primer, and the content associated with plate one of the three-plate Human miRNome miScript miRNA PCR Array (MIHS-3216Z). For the study, the miScript Primer Assay for hsa-miR-302 was also added to this plate. Real-time PCR was performed on an ABI-7900HT (Life Technologies) using the miScript cycling program, which consists of an initial hold at 95 °C for 15 min followed by 40 cycles of 94 °C for 15 s, 55 °C for 30 s and 70 °C for 30 s. Baseline and threshold were set at 3–15 and 0.2, respectively, and the Cq values were exported for analysis. MiRNA expression data are available in rdml format (**Supplementary Data 1**).

Quanta Biosciences qScript (QU). miRQC RNA samples were profiled using the qScript microRNA Quantification System from Quanta Biosciences. A total of 489 PerfeCta microRNA assay primers and control assays were arrayed into a set of six 96-well plates that were used to prepare eight sets of six 384-well qPCR plates. Two picomoles of each assay primer were added to four adjacent (quadrant) wells of each 384-well qPCR plate. The qPCR assay plates were dried down and stored at room temperature before use. For each nonserum RNA sample, 800 ng of RNA was used to prepare cDNA using the qScript microRNA cDNA Synthesis Kit (Quanta PN 95107). After cDNA synthesis, an equivalent of 1 ng of the original RNA sample was mixed with Perfecta SYBR Green SuperMix (Quanta PN 95054) and Universal PCR Primer (Quanta PN 95109) in 10 µl qPCR reactions. For the serum RNA samples, 10 µl of sample was used to prepare cDNA as described above and an equivalent of 12.8 nl of the original sample was added per well in the qPCR plates. Four cDNA samples were run in adjacent wells of each 384 well qPCR plate. The qPCR plates were run in a LightCycler 480 instrument (Roche) using a two-step cycling protocol (95 °C for 2 min followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s) and concluded by a melting curve. Raw data from the LightCycler 480 instrument were analyzed using the Fit Points method (LightCycler 480 Software) to calculate the Cq value using a threshold setting of a user defined multiple of the s.d. of the noise. MiRNA expression data are available in rdml format (**Supplementary Data 1**).

WaferGen Smartchip (WA). Steps involved in assessing miRNA expression on the SmartChip system involve RNA isolation, attachment of a poly(A) tail to the mature miRNAs and production of cDNAs via reverse transcriptase and a proprietary modified poly(T) primer. Once a cDNA library is made, it is mixed with a SYBR Green–based master mix and dispensed onto wells containing specific primer pair for each miRNA. Expression analysis is done via qPCR and melt-curve analysis.

Polyadenylation of miRNA and cDNA synthesis. cDNA from the miRNA samples was prepared using components from the Poly(A) Polymerase Tailing Kit P/N PAP5104H (Epicentre Biotechnologies) and the High Capacity cDNA Reverse Transcription Kit with RNAse Inhibitor P/N 4374966 (Life Technologies).

Polyadenylation. A single polyadenylation reaction was prepared in a 200 µl PCR grade tube. 2.0 µl of PCR Grade Water P/N 03315932001 or equivalent (Roche Applied Science) was combined with 1.5 µl of 10× RT Buffer from the High Capacity cDNA Kit, 1.0 µl Epicentre 10 mM ATP, 0.5 µl Epicentre poly(A) polymerase and 5 µl of miRNA sample (200 ng/µl). The entire solution was mixed gently and quickly centrifuged. Polyadenylation was performed in a standard thermal cycler using the following thermal profile: 30 min at 37 °C, 5 min at 70 °C and 4 °C until the next step.

cDNA synthesis. The following components were mixed in a sterile 200 µl PCR-grade tube: 5.7 µl of PCR-grade water, 0.5 µl of 10× RT buffer, 0.8 µl 25× dNTP Mix, 1.0 µl of SmartChip Universal miRNA RT Primer (20×) P/N 430-000087 (WaferGen Biosystems), 1.0 µl MultiScribe RT and 1.0 µl of RNase inhibitor. The entire solution was mixed gently, quickly centrifuged and combined with the 10-µl polyadenylation reaction. Reverse transcription was performed using the following thermal profile: 60 min at 40 °C, 5 min at 85 °C and 4 °C until the next step.

Real-time PCR on the SmartChip Human miRNA panel v3. In a sterile 1.5-ml tube, the following components were mixed: 424.0 µl of PCR-grade water, 500.0 µl of 2× LightCycler 480 DNA SYBR Green I Mix Roche P/N 04707516001, 50.0 µl of glycogen (20 mg/ml) Roche P/N 10901393001, 4.0 µl of M13mp18-ssDNA (250 µg/ml) P/N N4040S (New England BioLabs), 2.0 µl of WaferGen Yeast Control (50×) P/N 430-000007 and 20 µl of the cDNA products created in the previous steps. The entire solution was mixed gently and quickly centrifuged before dispensing onto a SmartChip Human miRNA Panel v3 chip WaferGen P/N 430-000084 using the SmartChip MultiSample Nanodispenser (WaferGen Biosystems).

Once a sample was dispensed, the panel containing the sample was cycled in the SmartChip Cycler using the following thermal parameters: 6 min at 95 °C, a two-step cycle composed of 1 min at 95 °C and 70 s at 52 °C, and 39 two-step cycles composed of 1 min at 95 °C and 70 s at 60 °C. Immediately after amplification, melt-curve analysis was performed from 60 °C to 97 °C. MiRNA expression data are available in rdml format (**Supplementary Data 1**).

Ilumina TruSeq (IL). The samples were prepared using the TruSeq Small RNA Prep kit (Illumina) using an input amount of 1 µg of total RNA. The kit was used as specified by the manufacturer without any modifications. All samples were prepared with a unique index, after which the libraries were mixed and sequenced across eight lanes of a GAIIx instrument with single-end 51-bp reads. The total number of reads used for analysis of each of the 20 experiments, along with the indexes that were used during library preparation, are given in **Supplementary Table 2**. Quality control, alignment and quantification of the data were performed using the Avadis NGS Software (v 1.4) from Strand Life Sciences. The Avadis NGS aligner was used to detect and remove the 3′ adaptor sequence after allowing for two mismatches

doi:10.1038/nmeth.3014

in the adaptor sequence. The trimmed reads were aligned against the hg19 reference sequence allowing for one mismatch. Reads matching more than ten locations in the genome were discarded. For all other reads, at most five best matches were used for each read. miRBase (version 18) annotations were used to assign reads to miRNA genes. A read was assigned to a miRNA only if the 5′ end of the read matched the 5′ end of the miRNA.

Life Technologies Ion Torrent (IT). Small RNA libraries were constructed with Ion Total RNA-Seq Kit v2 following manufacturer's instructions (Life Technologies, PN 4475936). Generally, 1 µg of total RNA (10 µl, 100 ng/µl) was used for small RNA enrichment, followed by overnight ligation, reverse transcription and PCR. For the serum samples, 10 µl RNA was dried and resuspended in 3 µl of nuclease-free water. The samples were then directly used for ligation without small RNA enrichment. Small RNA libraries were quantitated using the DNA 1000 assay on the Agilent 2100 Bioanalyzer.

miRNA libraries were qualified and quantified using High Sensitivity DNA chips on Bioanalyzer 2100 (Agilent). Templated spheres were generated using Ion OneTouch 200 Template Kit v2 DL and enriched with the Ion OneTouch ES following manufacturer's protocols. Enriched spheres were loaded on an Ion 318 chip (one library per chip) and sequenced with the Ion PGM 200 Sequencing Kit for 200 flows on the Ion Torrent Personal Genome Machine (PGM) running Torrent Suite Software 3.2.1. for automated signal analysis, base calling and mapping to human genome (hg19 build 37.2).

The libraries were sequenced on the Ion PGM System using 3NN chips and analyzed using Partek Flow. Briefly, quality analysis and trimming were done on all samples using Partek Flow. Each sample was run on its own chip (total number of reads per sample are listed in **Supplementary Table 3**). Quality analysis identified adaptors, which were trimmed by removing the first 11 base pairs (bp) from each read. Reads with lengths less than 16 bp after trimming were discarded. For the specificity and sensitivity control analysis, reads were aligned to synthetic miRNAs using Bowtie version 0.12.7 with sensitive settings (-v 3,–try-hard,–best, -k 1). Reads were aligned to the miRBase mature miRNAs (version 18) extracted from hg19 using SHRiMP 2.2.3 with miRNA default options settings (-M) using a gapped-seed index. Quantification was performed to determine reads abundance counts for each miRNA using Partek isoform expression quantification.

Data normalization. All data sets were normalized according to the manufacturer's instructions. Agilent microarray data were normalized using the 90 percentile shift method. Affymetrix microarray data were normalized using an RMA style background adjustment as described above. Nanostring nCounter data were normalized based on positive spike-in controls as detailed above. Illumina and Ion Torrent small RNA sequencing data were normalized using the total read count per sample and multiplied by the median total read count across all samples. All RT-qPCR data sets were normalized using the global mean[16].

Data processing. Unless specified by the manufacturer, a detection cutoff was calculated for each platform by means of the single positives in the replicate miRQC samples. The cutoff was defined such that it reduces the single positives fraction by at least 95% (**Supplementary Fig. 10**). This approach was applied for the Agilent, Affymetrix, Illumina, Ion Torrent, Qiagen, Quanta Biosciences and WaferGen platform data. Detection cutoff values for each of these platforms are listed in **Supplementary Notes 5**–**9**, **11** and **12**. For the remaining platforms, a cutoff was supplied by the manufacturer (Exiqon, 37 cycles; TaqMan Cards, 35 cycles; TaqMan Cards with preamplification, 32 cycles; OpenArray, 25 cycles). The common set of miRNAs was defined by means of the miRBase v18 MIMAT accession numbers.

Data analysis and statistics. All metrics are calculated using expression data from multiple miRNAs (miRNAs measured by all platforms, $n = 196$) to ensure a robust estimate representative of the overall platform performance. One sample was excluded from the analysis (OA miRQC D replicate two corresponding to sample 6; **Supplementary Note 2**) because of poor technical data quality. For reproducibility analysis, all replicate miRQC samples were included except for the OA platform, where both miRQC D replicates were excluded. For titration response analysis and accuracy analysis one replicate of each miRQC sample was used: samples 1, 3, 5 and 7 for the OA platform (**Supplementary Note 2**) and samples 2, 4, 6 and 8 for the remaining platforms (**Supplementary Notes 1** and **3**–**12**). Only the miRNAs expressed in all four miRQC samples (on a per-platform basis) were included for titration-response analysis. The miRQC samples (miRQC A–D) consist of 100% Universal Human miRNA Reference RNA (UHmiRR, A), 100% human brain RNA (HBR, B) and two titrations thereof (C = ¾A + ¼B and D = ¼A + ¾B). A titration response implies that, if for gene *i* $A_i > B_i$ (gene i is higher expressed in A versus B), then $A_i > C_i > D_i > B_i$ and conversely if $B_i > A_i$, then $B_i > D_i > C_i > A_i$. The ability to detect the correct sample order for any given miRNA will strongly depend on the expression difference between sample A and B. MiRNAs for which the correct sample order is maintained are called titrating miRNAs.

Absence or presence of miRNA expression is determined on the basis of the platform-specific detection cutoff. For accuracy analysis, all measured miRNAs per individual platform were considered as the common set of miRNAs only contained few miRQC A–specific or miRQC B–specific miRNAs. As a consequence, the number of data points used for this analysis differs between platforms. For specificity analysis using MS2-phage RNA, quadruplicate MS2-phage RNA samples were profiled and the median number of miRNAs detected was applied as a measure of specificity (or background detection rate). Differentially expressed miRNAs between miRQC A + miRQC C and miRQC B + miRQC D were identified using rank products[17] analysis (nonparametric test with no assumptions on data distribution) with 1,000 permutations. Only miRNAs detected in at least two of eight samples were included. Missing values were imputed based on the lowest expression of the respective miRNA minus one $\log_2$ unit. 'Significant' miRNAs were selected based on a percentage-false-positives value <0.05. Hierarchical clustering was performed using distance 'Manhattan' and clustering method 'Ward'. Correlations were calculated using Spearman's rank. Missing expression values were imputed by means of the lowest expression for the respective miRNA minus 1 $\log_2$ unit. For correlation analysis between

individual metrics, metrics were transformed (if needed) such that a higher metric value corresponds to a better performance

M1 = −ALC

M3 = $-\left|(\log_2 C/D) - (\log_2 3)\right|$ for miRNAs exclusively expressed in miRQC A

M3 = $-\left|(\log_2 D/C) - (\log_2 3)\right|$ for miRNAs exclusively expressed in miRQC B

M4 = $\frac{\sum_{k=1}^{n} |\Delta^{o}_{\mathrm{miR}_k} - \Delta^{e}_{\mathrm{miR}_k}|}{n}$ with $\Delta^{o}_{\mathrm{miR}_k}$ = observed difference in expression of microRNA k between serum sample 1 and serum sample 2 and $\Delta^{e}_{miR_k}$ = expected difference

M7 = − MS2 phage detection rate

M8 = − # of mismatches with cross-reactivity

M9 = − median cross-reactivity

All metrics were subsequently transformed to z-scores for correlation analysis. All analyses were performed in R Bioconductor. Analysis results for each platform based on all measured miRNAs per platform are available in **Supplementary Notes 1–12**.

12. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
13. Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
14. Geiss, G.K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).
15. Lefever, S. *et al.* RDML: structured language and reporting guidelines for real-time quantitative PCR data. *Nucleic Acids Res.* **37**, 2065–2069 (2009).
16. Mestdagh, P. *et al.* A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* **10**, R64 (2009).
17. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **573**, 83–92 (2004).

doi:10.1038/nmeth.3014