Perspective

Contributions of the EMERALD project to assessing and improving microarray data quality

Vidar Beisvåg[1], Audrey Kauffmann[2], James Malone[2], Carole Foy[3], Marc Salit[4], Heinz Schimmel[5], Erik Bongcam-Rudloff[6], Ulf Landegren[6], Helen Parkinson[2], Wolfgang Huber[2], Alvis Brazma[2], Arne K. Sandvik[1,7], and Martin Kuiper[8]

[1]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, [2]European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute, Cambridge, UK, [3]LGC, Queens Road, Teddington, Middlesex, UK, [4]National Institute of Standards and Technology, Gaithersburg, MD, USA, [5]European Commission, Joint Research Centre, Institute for Reference Materials and Measurements (IRMM), Geel, Belgium, [6]Department of Genetics and Pathology and Science for Life Lab Uppsala, Uppsala University, Uppsala, Sweden, [7]Department of Gastroenterology and Hepatology, St. Olav's University Hospital, Trondheim, Norway, and [8]Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

BioTechniques 50:27-31 (January 2011) doi 10.2144/000113591

Keywords: microarray data; quality metrics; ontology; protocols; standards; external standards; meta-analysis

While minimum information about a microarray experiment (MIAME) standards have helped to increase the value of the microarray data deposited into public databases like ArrayExpress and Gene Expression Omnibus (GEO), limited means have been available to assess the quality of this data or to identify the procedures used to normalize and transform raw data. The EMERALD FP6 Coordination Action was designed to deliver approaches to assess and enhance the overall quality of microarray data and to disseminate these approaches to the microarray community through an extensive series of workshops, tutorials, and symposia. Tools were developed for assessing data quality and used to demonstrate how the removal of poor-quality data could improve the power of statistical analyses and facilitate analysis of multiple joint microarray data sets. These quality metrics tools have been disseminated through publications and through the software package arrayQualityMetrics. Within the framework provided by the Ontology of Biomedical Investigations, ontology was developed to describe data transformations, and software ontology was developed for gene expression analysis software. In addition, the consortium has advocated for the development and use of external reference standards in microarray hybridizations and created the Molecular Methods (MolMeth) database, which provides a central source for methods and protocols focusing on microarray-based technologies.

Over the last 15 years, microarray technology has grown into a leading approach for large-scale investigation of transcriptomes, genomes, and epigenomes (1–3). Microarray technologies are continually improving, with new applications being reported regularly. Although newer approaches based on massively parallel DNA sequencing are maturing as an alternative to high-throughput technology for nucleic acid analysis (4), low cost and practicality will sustain microarrays as important tools in both basic research and, increasingly, for diagnostic applications (5).

The quality of the data produced from microarray technology for transcriptome analysis has often been the subject of criticism. In particular, reproducibility has been a concern both within a given platform and in cross-platform comparisons (6). This has been a particular issue with "homemade" spotted arrays. However, recent reports seem to agree that the technology can provide reliable results (7,8), even when performed in high-throughput. This is largely due to the increased robustness sample preparation and labeling, as well as a switch from homemade arrays to commercial platforms provided by companies such as Affymetrix (Santa Clara, CA, USA), Agilent (Santa Clara, CA, USA), and Illumina (San Diego, CA, USA).

Microarray data produced in specific, small-scale experiments provide a rich source of information. However, when large amounts of data generated from various independent experiments are to be analyzed, data quality and proper experiment annotation (metadata) are critical. This notion triggered the formation of the Microarray Gene Expression and Data Society [MGED; now renamed to the Functional Genomics Data Society (FGED); www.mged.org]. In 2001, MGED published guidelines for experiment descriptions (minimum information about a microarray experiment; MIAME) (9) and proposed a structured data exchange

format (MAGE-TAB) (10). This work has subsequently served as a model to develop guidelines and standards for many other high-throughput genomics technologies (11). What has made MIAME successful is the principle that data supporting a scientific analysis must be made available in a way that makes these data usable for others (12). Currently, data from more than 15,000 different microarray studies have been deposited in MIAME-compliant public repositories such as ArrayExpress (13), Gene Expression Omnibus (GEO) (14), and the Center for Information Biology Gene Expression Database (CIBEX) (15).

Microarray data has become an important resource for data-driven analysis, and metadata and data quality in public microarray data repositories are the major determinants for success in information extraction and model building. This is of particular importance for systems biology approaches, in which microarray data are used to derive models of biological processes. Whereas the focus of MGED has been predominantly on data context and the quality of metadata, the focus of the EMERALD (Empowering the Microarray-based European Research Area to Take a Lead in Development and Exploitation) project has been on data content and the quality of the quantitative data produced by the technologies. The EMERALD project (www.microarray-quality.org), funded by the European Union (EU) 6th Framework Programme, was established with three specific aims: (*i*) to help improve microarray data quality, by (*ii*) establishing a quality metric (QM) tool able to measure quality, and (*iii*) to build a normalization and transformation ontology to archive methods of data preprocessing. Toward those objectives, the EMERALD consortium has worked closely with MGED, and in this article, we describe and discuss the results of the project.

Quality metrics

The EMERALD Coordination Action has continued the process toward standardization of microarray data, with a focus on data content. EMERALD was designed to help improve data quality, both by supporting ongoing initiatives of MGED and the External RNA Controls Consortium (ERCC) (16), and through dedicated activities described herein. An overview of the specific results of these activities is summarized in Table 1.

A fundamental part of EMERALD has been the development of specific QMs and the generation of publicly available software implementing these metrics (17). This development was expected to serve multiple purposes: (*i*) a QM tool could offer a first line of defense against the submission of poor-quality data or retrieval of such data from common repositories. Core facilities and microarray laboratories can use QM software to screen their new results and diagnose problematic array data sets that could then either be withheld from submission or replaced by better quality data. (*ii*) For data already submitted, a QM tool would allow users contemplating a meta-analysis of public data to consider only data of sufficient quality. The arrayQualityMetrics Bioconductor package (www.bioconductor.org/packages/release/bioc/html/arrayQualityMetrics.html) was created to fulfill this aim. This QM tool integrates various existing approaches to microarray data quality assessment and, in some aspects, has developed them further. It recognizes different microarray platforms, including Affymetrix, Agilent, Illumina, and homemade two-color arrays, and computes general and platform-specific QMs. The tool produces a comprehensive report reflecting individual array quality, including both relative measures of quality within the data set and absolute metrics. To facilitate meta-analyses of public data, the ArrayExpress Bioconductor package has also been developed (18). This package provides a bridge from the ArrayExpress repository to the R/Bioconductor data analysis environment, allowing users to perform a wide variety of customized, experiment-specific analyses of data quality beyond what is possible in a standard workflow. ArrayExpress also uses both of these packages in a pipeline to identify high-quality data for integration with European Bioinformatics Institute (EBI) databases, such as the ArrayExpress Gene Expression Atlas (19).

QA and significance of results

We extensively used the QMs approach ourselves to demonstrate the benefits of pruning larger data sets and removing poor-quality data before further analysis. One recent paper (20) addressed the possibility that important genes can be missed in a statistical analysis if a data set contains poor-quality arrays. We used the QM software to identify outlier arrays (i.e., arrays that contribute data of very different and therefore presumably low quality), thereby compromising the statistical and biological significance of the analysis. The removal of such outlier arrays could significantly enhance the sensitivity of the analysis, by yielding improved statistical and biological significance. We note that array data can appear to be outliers because of a genuine biological property of a sample or peculiarities of the protocol (e.g., for reasons that are beyond the reach of an automated analysis). Therefore, due to the risk of removing data that are of interest to the analysis, fully automated, unsupervised outlier removal is currently not advisable. Instead, we recommend that automatically identified outlier candidates be manually examined before a decision is made. The comprehensive report with the graphical display of results generated by arrayQualityMetrics helps users understand whether a particular array should be considered an outlier. Manual inspection can also provide useful feedback to improve experimental protocols.

The results and experiences from transcriptome microarray quality assurance

Table 1. Overview of the results of the EMERALD project.

Results	Available at
arrayQualityMetrics software	www.bioconductor.org/packages/stats/bioc/arrayQualityMetrics.html
ArrayExpress Bioconductor package	www.bioconductor.org/packages/release/bioc/html/ArrayExpress.html
Experimental Factor Ontology	www.ebi.ac.uk/efo
Gene expression software ontology	www.ebi.ac.uk/efo/swo
The Molecular Methods Database	www.molmeth.org
Project web site	www.microarray-quality.org

(QA) and quality control (QC) will create an example for emerging applications of other high-throughput molecular technologies, such as microarray-based protein analyses. Initial studies with the QM software on protein microarrays have shown that it can detect outlier data similar to its application on transcriptome microarrays (unpublished observations), because many of the same metrics (reproducibility of replicate measurements, spatial distribution of the signal, and dynamic range across replicates) apply also to protein arrays. We have subsequently demonstrated the utility of the developed microarray QMs approach in a systems biology application by building a global map of human gene expression from a compendium of gene expression data generated by 163 different laboratories (21). Over 9000 raw data files produced on the Affymetrix U133A human gene expression array were collected from the GEO (14) and ArrayExpress (13) repositories. An integrated analysis of this massive microarray data set compiled from many different laboratories was used to reveal the overall structure of gene expression space, after a careful process to remove outlier arrays and imposing stringent criteria for data quality.

Representing semantically rich data transformations

Whereas our development of QM software reflects an extension of the original MGED objectives toward data content, we also extended the original focus of MGED by enabling the upload of metadata. Fundamental to the reproducibility of experimental protocols is the clarity of the documentation of the experiments. This requires the use of a language that is precise, explicit, unambiguous, and understandable for the scientific community. The use of controlled vocabularies can go some way toward achieving this by providing a restricted terminology that defines important aspects of a given domain or application. However, such vocabularies lack the ability to formally relate concepts to one another, sometimes resulting in an ad hoc "bag of words." Ontologies have become an important method for agreeing on cross-domain concepts and for describing experimental processes and data and are being implemented in interfaces such as the ArrayExpress archive to provide queries based on technology, software used, and statistical analysis methods (19). Ontologies offer the advantage of modeling explicit relations among concepts, such as subclass or part-of, and can contain rules, in the form of axioms, about the use of concepts that can be computationally processed. In certain languages, such as the W3C-approved Web Ontology Language (OWL) (22), these rules can be used to check that models are consistent (i.e., that contradictory statements are not made) and also aid interoperability by standardizing syntax for creating ontologies.

To describe the types of data transformations used in gene expression analysis in an ontology, we used a three-step approach. First, we collected use cases that were initially used to identify requirements and, later, to test the developed ontology. Second, we considered the sorts of domain concepts and relations that ideally would be described in the ontology and that could fulfil these use cases. Third, we integrated our efforts with an existing framework, the Ontology of Biomedical Investigations (OBI; http://obi-ontology.org/page/Consortium), which is a community-driven ontology built by an international consortium of over 30 groups, so that it may increase interoperability with other ongoing community efforts.

The value of the data transformation ontology is further increased when it is integrated within the wider context of OBI. OBI has the scope of describing the elements of a biomedical investigation including protocols, instruments, and roles of participants in studies. When we couple descriptions of the data transformation parts of experiments with the broader scope of OBI, we are able to more richly describe the entire experiment (with regard to, for example, instrumentation used, specific assaying processes, the input material, and the organization or individual performing a study). This effort represents the first substantial cross-community attempt to support the annotation of the experimental context of biomedical data, and is an important achievement within the microarray and wider bioinformatics community. The ontology is already deployed as part of the Experimental Factor Ontology (EFO) (23) in the Gene Expression Atlas, in the Immune Epitope Database (www.immuneepitope.org), and in Integrative Tools for Protozoan Parasite Research (www.bioontology.org/node/604). In the near future, it is expected that OBI will become an Open Biological and Biomedical Ontologies (OBO) Foundry ontology; this certificate of quality will thereby promote OBI as a community standard.

Besides describing the protocols, it is also important to describe the software used to generate the data, and these terms are more commonly used in papers than detailed descriptions of analysis processes. We have, therefore additionally developed an ontology of gene expression software that contains details of commonly used software such as BioConductor packages (www.ebi.ac.uk/efo/swo), the algorithms implemented, and the purposes they can be used for. This has been shown to be useful for text mining full-text articles and experimental records in ArrayExpress.

External standards

It is well-established that the use of common external standards (e.g., spikes or reference RNAs) is helpful to standardize and evaluate experimental results. At the start of the EMERALD project, multiple independent approaches were being proposed to develop such standards. These included platform-specific controls developed by individual array manufacturers and more generic efforts to develop reference methods and materials by the grassroots array community, metrology institutes, regulatory agencies, and technology providers. Examples of these initiatives included the Microarray Controls Consortium (MAQC) (24), the External RNA Controls Consortium (ERCC), the Clinical Laboratory Genomic and Genetic Standards (CLGGS) consortium, the Clinical and Laboratory Standards Institute (CSLI), and array standardization initiatives under the UK Measurements for Biotechnology program (www.mfbprog.org.uk). A key role of EMERALD was to survey, liaise with, and assess the various approaches, and advocate the use of consistent standards by the microarray community. During the lifetime of EMERALD, a single consortium (ERCC) dedicated to developing microarray reference standards emerged as the major developer of materials in this field. We practically assessed the potential benefits of using the ERCC reference standards, and in several workshops, we disseminated the work of the ERCC and EMERALD in developing and evaluating the materials. The ERCC is developing a panel of RNA sequences (16) to serve as a certified reference material (RM; SRM-2374) in the form of 96 sequence-verified plasmids containing the sequences. These ERCC panel sequences were selected for their cross-platform performance and data consistency from an initial library of candidate sequences donated by consortium members. A publication outlining platform performance and data consistency is currently in preparation.

Certification of the ERCC clones

Normally, RMs must comply with the quality criteria cited in International Organization for Standardization (ISO) 15194:2002 and ISO Guide 34:2009 (e.g.,

characterization, stability, homogeneity, and commutability assessment). However, ISO 15194:2002 only fully applies to RMs that possess quantifiable values of either a differential or rational quantity, which also holds true for the revised ISO 15194:2009 standard, but where nominal and ordinal property-related aspects have been further clarified. As a result, new criteria for the quality review of nucleic acid as RMs are being established, and under the ISO Committee on Reference Materials (REMCO), a working group has been set up to develop a standard on requirements for RMs for qualitative analysis, including nominal properties. Recently, additional guidance has been made available for managing nominal properties, such as those generated by sequence base calls (25).

Recommendations for certifying materials based on sequence include: (*i*) quality-scored bidirectional sequencing, (*ii*) verification by alternative sequence assay and independent laboratory, and (*iii*) expression of uncertainty as the probability of a miscalled base. In response to the last point, the National Institute of Standards and Technology (NIST) has developed a best-practice approach that utilizes an ordinal scale to characterize confidence in regions of sequence. This comprises four orders of descending confidence: 1, most confident (all data agree); 2, very confident (ambiguity resolved with second strand); 3, confident (ambiguity resolved with judgment); and 4, ambiguous (ambiguity unresolved).

All of the 96 ERCC external RNA controls have been analyzed by conventional bidirectional Sanger sequencing and next-generation sequencing using the ABI SOLiD (Applied Biosystems, Foster City, CA, USA) and Illumina GA-IIx platforms. NIST will now distribute the ERCC external RNA controls as standard RM 2374 (NIST SRM 2374) with an estimate of sequence confidence at each insert base of all 96 clones.

The major array vendors have committed to including content for these sequences on their arrays, and the Clinical and Laboratory Standards Institute have published guidelines for incorporating external RNA controls in gene expression assays (26). The guidelines cover issues associated with the use of external RNA controls as a tool for verification of technical performance. Areas covered include preparation of control transcripts, design of primers and amplicons, quality control, use in final experiment, and analysis and interpretation of data obtained.

Requirements for production of RNA from the ERCC controls

The current ERCC controls will initially only be released as plasmid clones. It will be necessary for producers of RM or microarray end-users to express RNA from the clones which can then be spiked into microarray experiments. To function as a certified RM, the RNA produced from the clones will need to be produced and certified in accordance with ISO guidelines as discussed; however, the materials will also need to be certified for quantity. Evaluation of in vitro transcribed RNA in terms of size, purity, and concentration—particularly with respect to batch-to-batch variation—is essential for the generation of robust and reproducible data.

MolMeth database

As a further aid to establishing stable protocols for the production of microarray data, the EMERALD project has contributed to the development of the Molecular Methods (MolMeth) database (www.molmeth.org), which provides the research community with an up-to-date source of methods and protocols used in molecular biology and molecular medicine. All entries in the MolMeth database are manually curated, meaning that protocols are checked against published papers or verified by consulting the authors before addition to the database. The database complements other efforts to standardize molecular techniques by simplifying the documentation of experimental procedures. The database has a modular, easily searchable structure, and it allows for convenient time-stamped versioning of protocols with hyperlinks to resources, such as commercial products or research publications. While the database is freely available, the user can choose to password-protect an entry to confine it to a smaller group before making it public, while corresponding data are being published.

Other activities

The EMERALD project has provided a conduit for any ongoing initiatives aimed at improving microarray data quality. The consortium has been working together with prominent players in the field, including MGED (now FGED), NIST, ERCC, and critical assessment of microarray data analysis (CAMDA) to disseminate the results of microarray standardization and quality improvement initiatives to the different microarray technology stakeholders. The web portal at EBI hosts a web site that presents an overview of these initiatives (www.microarray-quality.org) and serves to disseminate the results and deliverables of the project. The arrayQualityMetrics software package (17) has been downloaded by 2841 distinct internet addresses during the time from August 2009 to July 2010 (http://bioconductor.org/packages/stats/bioc/arrayQualityMetrics.html), and it has been adopted by academic and commercial microarray core facilities for their report generation.

Future perspectives

Several studies have shown that cross-platform comparisons are more consistent when microarrays and protocols developed by commercial companies are used. In general, users of those arrays are more likely to adopt standardized protocols than users of arrays manufactured by core facilities. This will eliminate a significant source of variation across experiments. The results and experiences from transcriptome microarray QA/QC will further create a cornerstone for systems biology–based life science and also be of value as a model for standardizing many other emerging high-throughput analyses, first for use by the broader research community, including systems biology and medical research, and later as a base for applications in routine diagnostics. Examples of new analytical technologies in need of standardization include the rapidly expanding scope for image analysis and the various forms of next-generation sequencing. In fact, MIAME guidelines have already created an offspring—similar guidelines for experiments in which assays are performed using high-throughput sequencing—the minimum information about a high-throughput sequencing experiment (MINSEQE; http://mged.org/minseqe). These guidelines have been closely modeled after MIAME. However, while initial microarray experiments were focused almost exclusively on gene expression, high-throughput assays have a much broader application range, which adds extra complexity. It is important that major scientific journals adopt MINSEQE guidelines soon, if the achievements of data openness that were facilitated by MIAME are not to be lost.

Acknowledgments

The EMERALD project was funded by the EU commission and the 6th Framework Programme. The authors wish to thank Ewa Sugajska for her administrative role in the initial phase of the project, Joaquin Dopazo for linking the EMERALD dissemination activities with the CAMDA meetings, and Astrid Lægreid for her driving role in the inception of the project.

Competing interests

The authors declare no competing interests.

References

1. **Gregory, B.D., J. Yazaki, and J.R. Ecker.** 2008. Utilizing tiling microarrays for whole-genome analysis in plants. Plant J. *53*:636-644.
2. **Gresham, D., M.J. Dunham, and D. Botstein.** 2008. Comparing whole genomes using DNA microarrays. Nat. Rev. Genet. *9*:291-302.
3. **Hoheisel, J.D.** 2006. Microarray technology: beyond transcript profiling and genotype analysis. Nat. Rev. Genet. *7*:200-210.
4. **Ansorge, W.J.** 2009. Next-generation DNA sequencing techniques. Nat. Biotechnol. *25*:195-203.
5. **Wu, L., P.M. Williams, and W. Koch.** 2005. Clinical applications of microarray-based diagnostic tests. BioTechniques *39*:S577-S582.
6. **Wilkes, T., H. Laux, and C.A. Foy.** 2007. Microarray data quality—review of current developments. OMICS *11*:1-13.
7. **Guo, L., E.K. Lobenhofer, C. Wang, R. Shippy, S.C. Harris, L. Zhang, N. Mei, T. Chen, et al.** 2006. Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nat. Biotechnol. *24*:1162-1169.
8. **Ioannidis, J.P., D.B. Allison, C.A. Ball, I. Coulibaly, X. Cui, A.C. Culhane, M. Falchi, C. Furlanello, et al.** 2009. Repeatability of published microarray gene expression analyses. Nat. Genet. *41*:149-155.
9. **Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, et al.** 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. *29*:365-371.
10. **Rayner, T.F., P. Rocca-Serra, P.T. Spellman, H.C. Causton, A. Farne, E. Holloway, R.A. Irizarry, J. Liu, et al.** 2006. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC Bioinformatics *7*:489.
11. **Taylor, C.F., D. Field, S.A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C.A. Ball, P.A. Binz, et al.** 2008. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat. Biotechnol. *26*:889-896.
12. **Brazma, A.** 2009. Minimum information about a microarray experiment (MIAME)—successes, failures, challenges. ScientificWorldJournal *9*:420-423.
13. **Brazma, A., H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, et al.** 2003. ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. *31*:68-71.
14. **Edgar, R., M. Domrachev, and A.E. Lash.** 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*:207-210.
15. **Ikeo, K., J. Ishi-i, T. Tamura, T. Gojobori, and Y. Tateno.** 2003. CIBEX: center for information biology gene expression database. C. R. Biol. *326*:1079-1082.
16. **Baker, S.C., S.R. Bauer, R.P. Beyer, J.D. Brenton, B. Bromley, J. Burrill, H. Causton, M.P. Conley, et al.** 2005. The External RNA controls consortium: a progress report. Nat. Methods *2*:731-734.
17. **Kauffmann, A., R. Gentleman, and W. Huber.** 2009. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics *25*:415-416.
18. **Kauffmann, A., T.F. Rayner, H. Parkinson, M. Kapushesky, M. Lukk, A. Brazma, and W. Huber.** 2009. Importing ArrayExpress datasets into R/Bioconductor. Bioinformatics *25*:2092-2094.
19. **Kapushesky, M., I. Emam, E. Holloway, P. Kurnosov, A. Zorin, J. Malone, G. Rustici, E. Williams, et al.** 2010. Gene Expression Atlas at the European Bioinformatics Institute. Nucleic Acids Res. *38(Database issue)*:D690-D698.
20. **Kauffmann, A. and W. Huber.** 2010. Microarray data quality control improves the detection of differentially expressed genes. Genomics *95*:138-142.
21. **Lukk, M., M. Kapushesky, J. Nikkilä, H. Parkinson, A. Goncalves, W. Huber, E. Ukkonen, and A. Brazma.** 2010. A global map of human gene expression. Nat. Biotechnol. *28*:322-324.
22. **Horrocks, I., P.F. Patel-Schneider, and F. van Harmelen.** 2003. From SHIQ and RDF to OWL: the making of a web ontology language. J. Web Semant. *1*:7-26.
23. **Malone, J., E. Holloway, T. Adamusiak, M. Kapushesky, J. Zheng, N. Kolesnikov, A. Zhukova, A. Brazma, and H. Parkinson.** 2010. Modeling sample variables with an experimental factor ontology. Bioinformatics *26*:1112-1118.
24. **MAQC Consortium.** 2006. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat. Biotechnol. *24*:1151-1161.
25. **Rousseau, F., D. Gancberg, H. Schimmel, M. Neumaier, A. Bureau, C. Mamotte, R. van Schaik, D. Payne, et al.** 2009. Considerations for the development of a reference method for sequencing of haploid DNA—an opinion paper on behalf of the IFCC Committee on Molecular Diagnostics. Clin. Chem. Lab. Med. *47*:1343-1350.
26. **Clinical and Laboratory Standards Institute.** Use of External RNA Controls in Gene Expression Assays; Approved Guideline. CLSI Document MM16-A. Vol. 26. [ISBN 1-56238-617-4]. CLSI, Wayne, PA.

Received 6 September 2010; accepted 30 November 2010.

Address correspondence to Martin Kuiper, Department of Biology, Norwegian University of Science and Technology (NTNU), Høgskoleringen 5, 7491 Trondheim, Norway. e-mail: kuiper@nt.ntnu.no