

Methodology Report

Automation of cDNA Synthesis and Labelling Improves Reproducibility

**Daniel Klevebring,¹ Marcus Gry,^{1,2} Johan Lindberg,^{1,3}
Anna Eidefors,⁴ and Joakim Lundeberg¹**

¹Division of Gene Technology, School of Biotechnology, Royal Institute of Technology, AlbaNova University Center, 10691 Stockholm, Sweden

²Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, 75185 Uppsala, Sweden

³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden

⁴Nordiag AB, P.O. Box 70036, 100 44 Stockholm, Sweden

Correspondence should be addressed to Joakim Lundeberg, joalun@kth.se

Received 31 March 2009; Accepted 26 July 2009

Recommended by Momiao Xiong

Background. Several technologies, such as in-depth sequencing and microarrays, enable large-scale interrogation of genomes and transcriptomes. In this study, we assess reproducibility and throughput by moving all laboratory procedures to a robotic workstation, capable of handling superparamagnetic beads. Here, we describe a fully automated procedure for cDNA synthesis and labelling for microarrays, where the purification steps prior to and after labelling are based on precipitation of DNA on carboxylic acid-coated paramagnetic beads. *Results.* The fully automated procedure allows for samples arrayed on a microtiter plate to be processed in parallel without manual intervention and ensuring high reproducibility. We compare our results to a manual sample preparation procedure and, in addition, use a comprehensive reference dataset to show that the protocol described performs better than similar manual procedures. *Conclusions.* We demonstrate, in an automated gene expression microarray experiment, a reduced variance between replicates, resulting in an increase in the statistical power to detect differentially expressed genes, thus allowing smaller differences between samples to be identified. This protocol can with minor modifications be used to create cDNA libraries for other applications such as in-depth analysis using next-generation sequencing technologies.

Copyright © 2009 Daniel Klevebring et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The field of gene expression analysis has evolved dramatically in recent years. With a basis in microarray technology and the ongoing transition into next-generation sequencing technologies, gene expression analysis is a widely used assay. The microarray technology provides a way of obtaining huge quantities of genome and transcriptome data; it has developed from relatively small-scale experiments using in-house platforms and protocols to more robust studies using what are essentially genome-wide commercially manufactured arrays. Today, several commercial and academic platforms exist, with different array manufacturing and sample preparation approaches. A common criticism of the field of global gene expression analysis has been its lack of

standardised experimental protocols and well-defined reference studies comparing different platforms and procedures. To address these issues, the Microarray Quality Control Consortium (MAQC) performed a study where relative gene expression measurements, using one and two-colour platforms, were compared within and between platforms, as well as with TaqMan real-time PCR data [1–3]. The reference data set, based on standardised commercially available RNA sources, was made publicly available, enabling researchers to benchmark new procedures and platforms.

One overall purpose for developing new protocols for global gene expression analysis is to improve the quality of the results produced. During recent years the quality of microarrays has been further improved in terms of content and fabrication procedures [4–6], ensuring limited

slide-to-slide and batch-to-batch variability. Equally important is sample preparation, during which several of the steps may introduce additional variability into the experiment. To minimise variation within an experiment, an experienced technician and a good laboratory protocol are required. In general, the statistical power increases as the variance in the experiment decreases, increasing the likelihood that, for example, differentially expressed genes and subtle differences between samples will be detected. In order to achieve these goals, we present a method where all the major steps in a typical cDNA library preparation protocol are automated using a robotic workstation capable of handling superparamagnetic beads. Our procedure performs cDNA synthesis, purification and subsequent labelling using NHS-modified fluorophores, and is capable of handling up to 48 samples per instrument in parallel. Studies have indicated that automation of sample preparation for single-colour microarray experiments can reduce variation and increase throughput [7, 8]. We present a similar approach, to benchmark our protocol against the freely available data from MAQC to validate our results, and compare the automated approach to our own manual procedure as well as to several academic and commercial platforms.

Several methods for the purification of nucleic acids are available today, including ethanol precipitation and spin-column-based methods. These methods generally require manual input, thereby limiting the throughput of the experiment and increasing variability. In order to reduce human input and increase reproducibility, an automated protocol for cDNA synthesis, labeling, and purification has been developed using a dedicated instrument capable of handling superparamagnetic beads. The procedure is outlined in Figure 1, where the two purification steps are highlighted.

For a full run of 48 samples, the current protocol takes about five hours, with most of the time being incubation steps (e.g., cDNA synthesis takes two hours). Performing 24 similar reactions manually can be done in approximately the same time, but manual handling introduces variation that can be avoided using the automated approach. The automated protocol outperforms the manual procedure when it comes to within-experiment correlation between replicates.

2. Results

An overview of the automated approach for total RNA followed by cDNA synthesis and labelling is illustrated in Figure 1. This procedure consists of several critical steps that were optimised, including evaporation from open wells, investigation of bead capacity, and proper clean-up from free fluorophores. In conclusion, we show that an automated approach for cDNA synthesis performs better than a similar manual procedure when interrogated using DNA microarrays. We believe that this conclusion is valid for other read-out platforms as well, including RNA-sequencing [9–11].

2.1. Bead Capacity and Evaporation. In order to increase the yield from the purification steps, we used a double capture approach to increase purification yield. Briefly, following

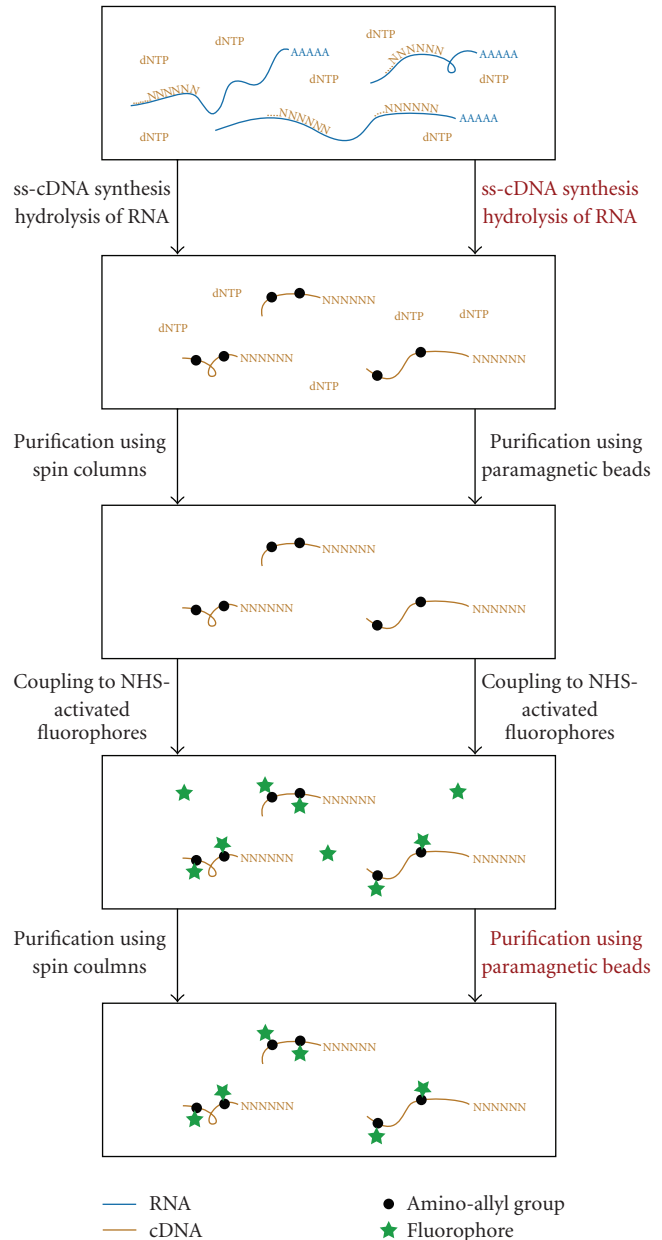


FIGURE 1: Outline of the cDNA synthesis and labelling workflow. The manual workflow is presented on the left and the automated workflow on the right. Differences between the manual and automated approaches are highlighted in red, and include the purification steps and neutralization after RNA hydrolysis.

the first elution, the beads are returned to the supernatant, enabling capture of any residual cDNA. This approach increases the yield from each purification step by approximately 15%, which has a significant impact on the overall yield of labelled cDNA, given that the process makes use of two purification steps (data not shown). Only small amounts of beads are necessary for a high yield: approximately 1.8 μg of labelled cDNA can be purified using as little as 20 μg of beads if there are two captures during each purification step (Figure 2). To facilitate complete capture, we use 150 μL

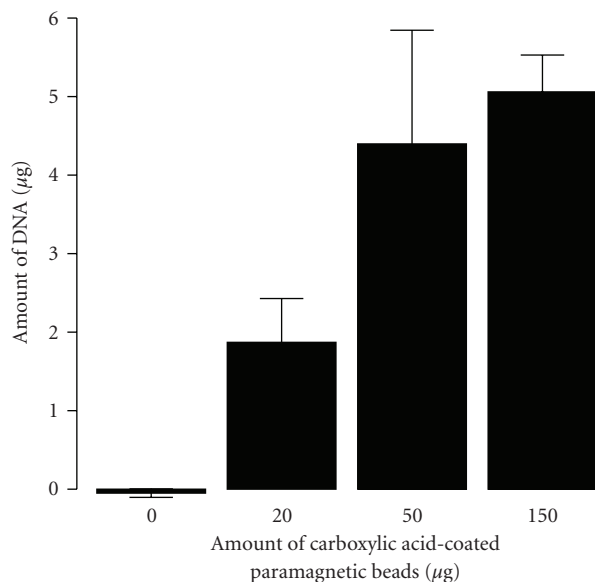


FIGURE 2: Binding capacity of the carboxylic acid-coated paramagnetic beads. Concentration of cDNA is measured before coupling to fluorophores.

of beads, thus purifying approximately 5 µg of labelled cDNA. Precipitation is carried out in an ethanol/TEG buffer, after which the bead pellet is washed five times in 80% ethanol. A large number of washes are necessary to remove unincorporated fluorophores, but does not affect the yield of sample cDNA (data not shown). Elution can easily be performed in a low salt buffer, such as water.

2.2. Intraplatform Correlation Analysis. In order to test the performance and reproducibility of the automated process, we took advantage of the well-documented model system used by the MAQC Consortium [1–3]. The aim of the MAQC study was to create a freely-available large reference data set and to compare multiple transcription profiling platforms using a standard set of RNA samples. We chose a subset of these platforms for our analysis. For within-experiment comparisons, the two-colour experiments, NCI 3 and 4 (National Cancer Institute), OPN (Operon), and NMC (Norwegian Microarray Consortium) were selected. These platforms use the human probe set version 3 from Operon, which is also used on the slides in our study. We used the same sources of RNA, Total Human Reference RNA (Stratagene), and FirstChoice Human Brain Reference Total RNA (Ambion) to evaluate and compare the automated protocol with the manual procedure for cDNA synthesis and labelling prior to microarray analysis. Ten replicates were run in two separate automated experiments and were compared with ten replicates prepared according to the manual procedure. The differential gene expression between the two sources of RNA was measured in these experiments as outlined for the previous MAQC experiments. Figure 3 shows the overall results by comparing the correlation between the M -values ($\log_2(\text{Sample A}/\text{Sample B})$) from our experiments (KTHAuto and KTHMan) and M -values from different

experiments in previous studies [2, 3], including only two-colour platforms. The samples from our two automated experiments cluster together closely and cluster less closely to the manual samples. Within the “automated” cluster, two samples (KTHAuto_B2 and KTHAuto_A4) are set slightly apart from the other samples but are close to each other; they are still, however, within the same super-cluster (blue cluster, Figure 3(a)). In both of these samples, the cDNA synthesis and labelling procedure yielded about two thirds the amount of labelled cDNA compared to the other samples (this was due to manual handling-related issues), while the degree of labelling was about the same (data not shown). This may indicate that the total amount of hybridised material could play an important role in creating reproducible results. The median Spearman correlation of all available M -values between samples labelled using the automated protocol was 0.92 and 0.91 for the two parallel experiments (KTHAuto1 and KTHAuto2, resp.), compared to 0.86 for the manual procedure (KTHMan, Figure 3(b)) indicating an increase in well-to-well reproducibility. This is in concordance with previous studies [7, 8]. In the wider context, a higher correlation between replicates means greater statistical ability to detect differentially expressed genes. When comparing the top 200 differentially expressed genes, 175 (87.5%) are common between KTHAuto1 and KTHAuto2, which can be compared to 155 (77.5%) between NCI3 and NCI4 (Supplementary Figures 1(a) and 1(b) in Supplementary Material available online at doi:10.1155/2009/396808).

2.3. Interplatform Analysis. In the MAQC study, a list of genes common to all platforms was created. After mapping the probes to RefSeq [12] and AceView [13], a complete list of 12091 probes for each platform was created, representing 12091 RefSeq entries in 12091 Entrez genes [3, 14]. After filtering out genes with stable expression levels across all experiments (see Section 2), we used this list to calculate the Spearman’s rank correlation coefficient between mean M -values from each experiment. When compared to other platforms investigated in the MAQC project, the data generated using the automated approach cluster together with the two-colour arrays used by NCI. This is to be expected since, as aforementioned, both NCI and the arrays used in the automation experiments made use of the human probe set version 3 from Operon. As expected from previous studies [2], all one-colour platforms (GE Healthcare, Applied Biosystems, Agilent One-colour, Illumina, and Affymetrix, see [3] for more information on these platforms) cluster together (Figure 4). When comparing the Spearman correlation values, the correlation between the KTH Auto experiments is 0.97, whereas the highest correlation between the different experiments performed by NCI is 0.94. A higher correlation between experiments (lower technical noise) means that samples can be added later, while introducing relatively little variability to the data.

3. Discussion

Here, we chose to take advantage of an efficient in-tip magnetic separation system [15, 16] to assess throughput and

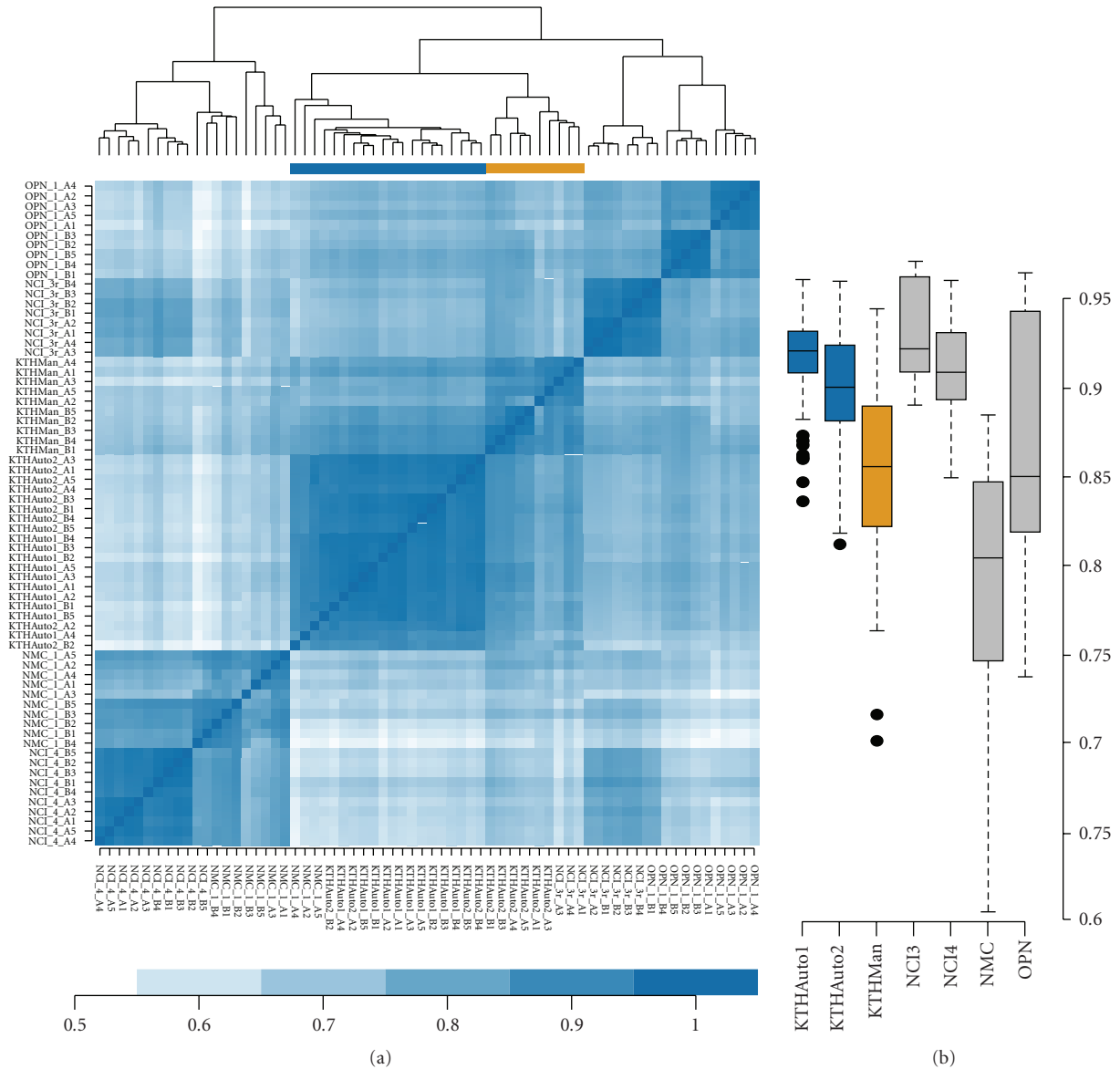


FIGURE 3: Automation of cDNA synthesis and labelling produces a higher correlation between technical replicates. (a) Dendrogram and heatmap of Spearman's rank correlation between manually labelled (gold cluster, KTH_Man) and automatically labelled (blue cluster, KTH_Auto1, and KTH_Auto2) samples. (b) Boxplot of the well-to-well reproducibility within experiments. Automatically labelled samples (blue boxes) are more highly correlated than manually labelled samples (gold box). Our automated approach also exhibited a higher correlation than between NMC and OPN experiments, but not higher than between the two NCI experiments.

variance within a gene expression experiment. There are several approaches available for nucleic acid purification using paramagnetic beads; these include the use of streptavidin-coated beads and a biotinylated primer in the reverse transcription step. However, unless strictly controlled, free biotinylated primers or free biotinylated nucleotides can rapidly saturate the beads, leading to a low purification yield. Here we show that, by precipitating the first-strand cDNA on carboxylic acid coated paramagnetic beads, a high yield can be achieved in the purification steps, giving a large quantity of purified product. Using this automated procedure, the technical variation is significantly decreased

when compared to a corresponding manual experiment. Thus, automation of cDNA library preparation for analysis on microarrays or using massive-scale sequencing [9–11] leads to decreased variance and greater statistical power to detect for example differentially expressed genes or alternative splicing patterns.

We describe an automated platform that can be used to increase the robustness of the overall performance of cDNA library preparation, including target labelling. The MAQC study indicated two types of variation: array content and array performance, the latter relating to the manual variation associated with performing the experiments. We show that

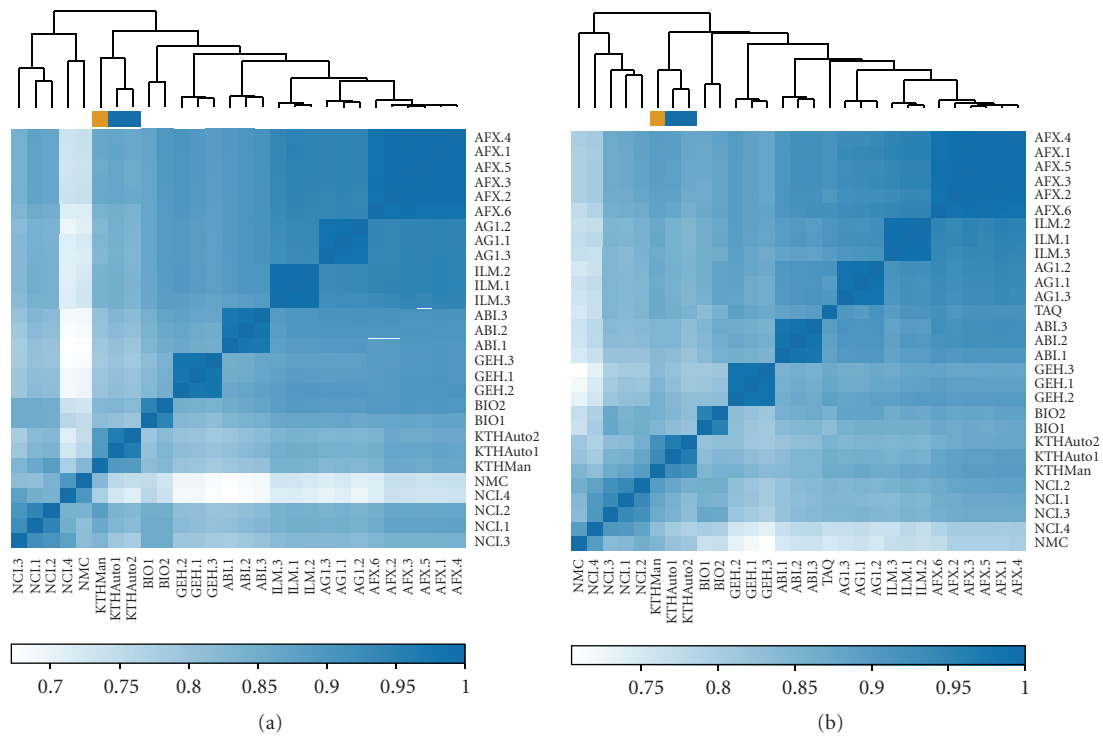


FIGURE 4: Dendrogram and heat maps of interplatform correlations. (a) is based on the 12091 probes common to all platforms (based on the mapping procedure in the MAQC study). A distinct clustering of platforms based on identical probes manufactured by Operon is apparent (KTH_Auto, KTH_Man, NCI.1–3). The one-colour platforms (GEH, ABI, AG1, ILM, AFX) were generally more highly correlated than two-colour platforms (NMC, NCI, BIO, KTH). (b) is based on a subset 906 probes for which TaqMan-based relative expression levels are available. Generally, the one-colour platforms are clustered closer to the TaqMan experiment. Interestingly, the NCI.4 experiment is positioned somewhat ambiguously in the two dendrograms, possibly indicating some uncertainty in the dendrograms.

this variation can be minimised by using an automated procedure, employing a standard microtiter plate. The number of samples that can be run in parallel can also be increased using this protocol from one up to 48. A single row in the microtiter plate (1 to 12 samples) takes approximately 4 hours and 40 minutes and the time increase for running four rows (48 samples) is marginal. Some modifications, other than changing the purification method, were necessary when automating the process. These were mainly due to reactions associated with the microtiter wells, which have no lids, so that evaporation occurs and is pronounced at elevated temperatures. The amount of evaporated water at a given temperature, however, is relatively constant. During the initial total RNA denaturing step, about 3.5 μL of water evaporates, and this can easily be taken into account in subsequent steps. Prior to optimisation, the amount of cDNA obtained from the first strand synthesis on the workstation was approximately 20% less than from synthesis in closed tubes. This effect was completely eradicated when water was added at regular intervals during the cDNA synthesis (data not shown). The adjusted volume of the samples was determined empirically.

To investigate the performance of our automated procedure, we chose DNA microarrays because of the access to a well established reference set of RNA samples and the standardized statistical procedure to address and demon-

strate technical variation. In general, well-to-well correlation increased when compared to a manual protocol. Correlation between different experiments (i.e., the same experiment performed at different days), we noted a higher correlation using our automated procedure when compared to the procedure carried out by NCI. Our results indicate that automation of sample preparation improves technical reproducibility, and should be general indifferent of the platform for readout, DNA microarrays or RNA-sequencing.

4. Materials and Methods

4.1. Experimental Design. In all experiments, Total Human Reference RNA (Stratagene) (sample A) and FirstChoice Human Brain Reference Total RNA (Ambion) (sample B) were used. The two RNA samples were labelled as described below, and subsequently hybridised to microarrays. In each experiment, a total of 10 microarrays was used. On five of these, sample A labelled with Cy3 and sample B labeled with Cy5 were hybridised. On the remaining five slides, the dye setting was reversed, giving a dye-balanced direct design, corresponding to the two-colour microarray experiments that were carried out within the MAQC study. Using the optimised protocol, we performed two experiments following the automated approach and one following our manual procedure.

4.2. Sample Preparation. For the automated sample preparation, the process from total RNA to purified labelled cDNA was performed on a Magnatrix 1200 (Magnetic Biosolutions, Sweden) robotic workstation. This workstation is equipped with a heating block and a system for in-tip magnetic separation. For initial cooling in the instrument, we used a cooling block (PCR cooler, Eppendorf). The protocol takes approximately 4 hours and 40 minutes for a single microtiter row of 12 samples, and only slightly longer for the double amount of samples. Twenty micrograms of total RNA was primed with 5 μg of random hexamers (Invitrogen, USA) in a total volume of 22 μL DEPC-treated sterile deionised water. After denaturing the RNA for 10 minutes at 70°C, the mixture was cooled in a PCR-cooler (Eppendorf) in the instrument, maintaining 0°C in the wells for 5 minutes prior to starting the first strand cDNA synthesis. After this step, approximately 3.5 μL of liquid had evaporated, resulting in a total volume of about 18.5 μL . A volume of 11.5 μL of a reverse transcription master mix was then added, setting the reaction composition to 1x first-strand buffer (Invitrogen), 0.01 mM DTT (Invitrogen), 0.4 mM aminoallyl-modified dUTP (Biotium), 0.1 mM dTTP (Sigma-Aldrich), 0.5 mM dATP, dCTP, and dGTP (Sigma-Aldrich) and 400 units of Superscript III reverse transcriptase (Invitrogen). An initial incubation of 10 minutes at room temperature (approximately 22°C) was followed by two hours at 46°C. During the incubation, 2.5 μL DEPC-treated sterile deionised water was added every 15 minutes to compensate for evaporation. The cDNA synthesis reaction was stopped by the addition of 3 μL 0.2 M EDTA after which RNA was hydrolysed by the addition of 5 μL 1 M NaOH together with incubation at 70°C for 15 minutes. pH neutralisation was achieved by the addition of 15 μL 1 M HEPES, pH 7.0, after the hydrolysis step. Prior to purification, the storage buffer was removed from 150 μg of DynaBeads MyOne Carboxylic acid (Invitrogen) by magnetic separation. The beads were resuspended in the neutralised cDNA synthesis mixture, after which three volumes (150 μL) of binding buffer, containing 80% ethanol and 6.7% TEG, were added. After a single 10-minute incubation at room temperature, the beads were collected from the supernatant and washed five times in 45 μL 80% ethanol. Elution was carried out by resuspending the beads in 10 μL sterile deionised water and mixing for 1 minute. To increase the yield, collected beads were resuspended in 5 μL of sterile deionised water and returned to the supernatant for a second capture. The incubation, washing and elution steps were then repeated as described above, resulting in a final volume of 20 μL . In order to set the pH to facilitate coupling to fluorophores, 2 μL 1 M NaHCO₃, pH 9.0, was added to the eluted cDNA, after which the mixture was transferred to a 5 μL aliquot containing one tenth of the contents of mono-functional NHS-ester Cy3 or Cy5 dye tubes (GE Healthcare) dissolved in DMSO (Sigma-Aldrich). After a single 30-minute incubation at room temperature, 2 μL of 1 M HEPES, pH 7.0, was added to neutralise the mixture prior to purification of fluorophore-coupled cDNA using carboxylic acid coated paramagnetic beads as described above.

The manual procedure was similar, but there were minor differences. Briefly, the pre-cDNA synthesis incubation was performed at 25°C instead of room temperature, and the purification steps were carried out using the MinElute cleanup system (Qiagen) according to the manufacturer's recommendations. Differences include changing the recommended washing buffer before labelling the cDNA to 80% ethanol and the elution buffer to sterile deionised water. Moreover, the neutralisation step was carried out using 5 μL 1 M HCl.

4.3. Hybridisation. The arrays used were provided by the KTH Microarray Center (<http://www.ktharray.se/>), and consisted of the Human Genome Oligo Set version 3.0 (Operon) printed in 30% DMSO onto UltraGAPS slides (Corning). After printing, the slides were UV cross-linked at 150 mJ/cm². The slides were prehybridised for 30 minutes at 42°C in a prehybridisation solution consisting of 5x SSC, 0.1% SDS (Sigma-Aldrich) and 1% BSA (Sigma-Aldrich) to avoid unspecific hybridisation to the glass surface. The slides were subsequently washed in water and isopropanol (Sigma-Aldrich) and dried using a slide centrifuge. The two samples were dye-balanced to make sure that equal amounts of dye were hybridised in each channel, then they were pooled and denatured (3 minutes at 95°C) in a hybridisation mixture containing 50% formamide (Sigma-Aldrich), 5x SSC and 0.1% SDS (Sigma-Aldrich), 20 μg human Cot-1 DNA (Invitrogen), and 20 μg Yeast tRNA (Invitrogen). The 65- μL hybridisation mixture was then cooled on ice for 1 minute and applied under a cover slip (Erie Scientific Company), placed on top of the printed array; it was then hybridised for 24 hours at 42°C in a water bath. Following hybridisation, the slides were washed with increasing stringency using 2x SSC and 0.1% SDS at 42°C, followed by 0.1x SSC and 0.1% SDS at room temperature and finally by five repeated washes with 0.1x SSC at room temperature.

4.4. Image and Statistical Analysis. The arrays were scanned at 10- μm resolution using an Agilent G2565BA scanner (Agilent Technologies, USA), with the photo multiplier tube set to 100% for each laser. The images thus acquired were analysed using the irregular gridding algorithm in Genepix Pro 5.1 (Axon) and the resulting data were imported into the R environment for statistical computing and visualisation [17]. The raw data were extracted from the median foreground intensity for both channels and subsequently filtered on the basis of flags (features either not found by the image software, or marked as bad spots), a low signal compared to the local background and saturated signals. Data normalisation was carried out using a block-wise Lowess approach, included in the aroma package [18] and a per-feature mean log₂-ratio (*M*-value) across all slides was calculated using the limma software package [19]. The data was compared with relevant parts of the MAQC data set, using clustering and correlation analysis tools in the KTH software package [17]. Mean fold change and *P*-values for each platform were calculated, and all contrasts set up between all platforms using the limma software package. The

files containing raw data were made publicly available in the ArrayExpress repository [20], with experiment number E-TABM-749. For intraplatform comparisons, a subset of platforms with the same probe set (Operon human probe set version 3) and experimental design from MAQC was used. Features were mapped using the oligonucleotide ID, which is a unique identifier for each oligonucleotide sequence. For interplatform comparisons, the M -values for 12091 genes common to the platforms were extracted, as well as a subset of 906 genes with available TaqMan real-time PCR-data; these were compared to all the microarray platforms within MAQC. Genes for which the interquartile range was lower than 0.5 were removed from the analysis, in order to reduce background noise from genes with stable expressions levels across all tested platforms.

4.5. Analysis of the Bead Binding Capacity. In addition to hybridising the labelled cDNA samples to microarrays, the binding capacity of the carboxylic acid-coated paramagnetic beads was measured. The number of washes after binding the cDNA was optimised, after which elution was carried out and cDNA and fluorophore concentrations were determined using a Nanodrop ND-1000 instrument (NanoDrop).

Acknowledgments

The authors wish to thank Annelie Waldén, Anna Westring, Max Källér, and Peter Nilsson for help with printing and quality control of the slides, Valtteri Wirta for critical reading of the manuscript and Anna Stenius, Erik Vernet, and Emma Lundberg for kindly providing the cells used for optimisation purposes. This work was supported through a support from the Swedish Research Council and Knut and Alice Wallenberg Foundation.

References

- [1] R. D. Canales, Y. Luo, J. C. Willey, et al., "Evaluation of DNA microarray results with quantitative gene expression platforms," *Nature Biotechnology*, vol. 24, no. 9, pp. 1115–1122, 2006.
- [2] T. A. Patterson, E. K. Lobenhofer, S. B. Fulmer-Smentek, et al., "Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project," *Nature Biotechnology*, vol. 24, no. 9, pp. 1140–1150, 2006.
- [3] L. Shi, L. H. Reid, W. D. Jones, et al., "The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements," *Nature Biotechnology*, vol. 24, no. 9, pp. 1151–1161, 2006.
- [4] G. Mcgall, J. Labadie, P. Brock, G. Wallraff, T. Nguyen, and W. Hinsberg, "Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 24, pp. 13555–13560, 1996.
- [5] M. Lee, J. M. Trent, and M. L. Bittner, "Optimization of oligonucleotide microarray fabricated by spotting 65-mer," *Analytical Biochemistry*, vol. 368, no. 1, pp. 61–69, 2007.
- [6] K. Kuhn, S. C. Baker, E. Chudin, et al., "A novel, high-performance random array platform for quantitative gene expression profiling," *Genome Research*, vol. 14, no. 11, pp. 2347–2356, 2004.
- [7] D. R. Dorris, R. Ramakrishnan, D. Trakas, et al., "A highly reproducible, linear, and automated sample preparation method for DNA microarrays," *Genome Research*, vol. 12, no. 6, pp. 976–984, 2002.
- [8] F. Raymond, S. Metairon, R. Borner, M. Hofmann, and M. Kussmann, "Automated target preparation for microarray-based gene expression analysis," *Analytical Chemistry*, vol. 78, no. 18, pp. 6299–6305, 2006.
- [9] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, vol. 5, no. 7, pp. 621–628, 2008.
- [10] N. Cloonan, A. R. R. Forrest, G. Kolle, et al., "Stem cell transcriptome profiling via massive-scale mRNA sequencing," *Nature Methods*, vol. 5, no. 7, pp. 613–619, 2008.
- [11] U. Nagalakshmi, Z. Wang, K. Waern, et al., "The transcriptional landscape of the yeast genome defined by RNA sequencing," *Science*, vol. 320, no. 5881, pp. 1344–1349, 2008.
- [12] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Research*, vol. 35, supplement 1, pp. D61–D65, 2007.
- [13] D. Thierry-Mieg and J. Thierry-Mieg, "AceView: a comprehensive cDNA-supported gene and transcripts annotation," *Genome Biology*, vol. 7, supplement 1, S12 pages, 2006.
- [14] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: gene-centered information at NCBI," *Nucleic Acids Research*, vol. 35, supplement 1, pp. D26–D31, 2007.
- [15] V. Wirta, A. Holmberg, M. Lukacs, et al., "Assembly of a gene sequence tag microarray by reversible biotin-streptavidin capture for transcript analysis of *Arabidopsis thaliana*," *BMC Biotechnology*, vol. 5, article 5, 2005.
- [16] A. Holmberg, A. Blomstergren, O. Nord, M. Lukacs, J. Lundberg, and M. Uhlén, "The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures," *Electrophoresis*, vol. 26, no. 3, pp. 501–510, 2005.
- [17] V. Wirta, M. Gry, J. Lindberg, and D. Klevebring, "kth: an R package for analysis of microarray data," 2007.
- [18] H. Bengtsson, "aroma—an R object-oriented microarray analysis environment," Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden, 2004.
- [19] G. K. Smyth, *Limma: Linear Models for Microarray Data*, Springer, New York, NY, USA, 2005.
- [20] H. Parkinson, M. Kapushesky, M. Shojatalab, et al., "Array-Express—a public database of microarray experiments and gene expression profiles," *Nucleic Acids Research*, vol. 35, supplement 1, pp. D747–D750, 2007.