# Comparison of globin RNA processing methods for genome-wide transcriptome analysis from whole blood

*Vanessa Dumeaux[1,2†], Eiliv Lund[1] & Anne-Lise Børresen-Dale[2,3]*

*[†]Author for correspondence*
*[1]University of Tromsø, Institute of Community Medicine, Norway*
*Tel.: +47 7764 4816;*
*Fax: +47 7764 4831;*
*Email: eiliv.lund@ism.uit.no*
*[2]Rikshospitalet-Radiumhospitalet Medical Centre, Department of Genetics, Institute for Cancer Research, Montebello, Norway*
*Tel.: +47 2293 4748;*
*Fax: +47 2293 4440;*
*Email: vanessad@rr-research.no*
*[3]University of Oslo, Faculty of Medicine, Norway*
*Tel.: +47 2293 5677;*
*Fax: +47 2293 4440;*
*Email: a.l.borresen-dale@medisin.uio.no*

**Aims:** Whole blood is likely to become the prime tissue to detect biomarkers using gene expression analyses. In this study, we assessed whether whole-blood or globin-reduced RNA provides the most robust and sensitive results to detect small gene expression changes, for example, in response to hormone therapy (HT) exposure. **Material & methods:** Each sample (n = 12) was processed according to three different protocols: no globin reduction, globin reduction using peptide nucleic acids (PNAs) and globin reduction using magnetic beads in the GlobinClear™ kit from Ambion. **Results:** Both globin reduction approaches using Ambion kit and PNAs were efficient at reducing globin RNA. However, globin reduction using Ambion kit also lowered the remaining cRNA. Samples processed with PNAs gave an intermediary profile closest to the no globin reduction group, with a slight increase in sensitivity of transcript detection and decrease in variability, but a loss of reproducibility. **Conclusions:** Globin RNA processing method in blood transcriptome analyses should be optimized for each microarray platform to be used. In our study, globin reduction was not found to be beneficial using the Applied Biosystems microarray system (AB1700).

Keywords: blood transcriptome, gene expression profile, globin RNA

As a major defense and transport system, blood cells may adjust gene expression activity in response to various clinical, biochemical and pathological conditions. Of note, blood sampling is minimally invasive, easily obtained and at low cost. Previous studies have found that use of peripheral blood cells for transcriptome analysis was valuable to assess disease-associated [1–4] and drug response-related [5] gene signatures. Furthermore, blood-derived gene expression profiles were associated to particular exposures, such as smoking [6], metal fumes [7], benzene [8] or ionizing radiation [9]. In our exploratory research, we have previously studied gene expression profiles from whole blood related to hormone therapy (HT) use in postmenopausal women [10]. Comparing treated with untreated women, very few genes had change in expression greater than 1.5-fold, which urges us to optimize detection sensitivity of low-level gene expression changes in whole blood.

Analysis of whole blood rather than purified peripheral blood mononuclear cells as surrogate specimens minimizes variations due to sample handling and processing. The use of the PAXgene™ blood RNA system (PreAnalytiX GmbH, Hembrechtikon, Switzerland) or similar systems as stabilization reagents offer advantages over nonstabilized samples by direct preservation of RNA. Several reports have shown that total RNA isolated from stabilized whole blood offers high quality RNA suitable for gene expression analyses [11,12]. However, a sensitivity issue is introduced using whole blood due to excess globin mRNA, originating mainly from lysed immature reticulocytes [13]. To address this problem, various protocols have been introduced to reduce globin RNA levels in whole blood cell RNA. While a nuclease-dependent globin reduction protocol was reported to increase the number of present calls, evaluation of this procedure indicated a lack of requisite specificity of RNase H for the RNA:DNA hemoglobin hybrids [14]. Recently, non-nuclease-dependent globin reduction protocols employing either peptide nucleic acid (PNA) oligomers that block hemoglobin-specific reverse transcription or magnetic beads that hybridize specifically into globin RNA have also been introduced [15] and should be further evaluated.

In this study, we assessed whether whole-blood RNA or globin-reduced RNA provides the most robust and sensitive results to detect small gene expression changes in response to an exposure such as HT. We further evaluated different globin RNA reduction methods and their application to the Applied Biosystems Human Genome Survey oligo-microarray platform, AB1700 (Foster City, FL, USA).

## Methods

### *Subjects*

The study population was drawn from women who are participants of the large prospective NOWAC postgenome study [16,101]. Among 500 women who were mailed equipment for

10.2217/17520363.2.1.11 © 2008 Future Medicine Ltd ISSN 1752-0363

blood collection, 445 returned both PAXgene and citrate tubes. Total RNA was isolated from 2.5 ml of whole blood drawn in PAXgene tubes with the PAXGene Blood RNA System.

To assess the effect of globin reduction protocols on gene expression from whole blood, we selected 12 postmenopausal women (six HT users and six non-HT users) who were only using HT at the time of blood sampling and in order to cover a wide body mass index range in both HT and non-HT users. The samples were also selected to contain the highest concentration of total extracted RNA.

### *RNA extraction & cRNA synthesis with & without globin reduction protocols*

Total RNA was isolated from 2.5 ml of whole blood with the PAXgene Blood RNA System.

#### Main protocol without globin reduction

The concentration of all total RNA was measured by the NanoDrop™ Spectrophotometer and the integrity was assessed by the Agilent 2100 Bioanalyzer™. Without globin reduction, 0.9–1 µg of total RNA was incubated with T7-Oligo (dT) Promoter Primer for 5 min at 70°C and then cooled to 4°C. This step is followed by second-strand synthesis and cDNA purification. Next, an IVT-based amplification converts cDNA into digoxigenin (DIG)-labeled cRNA according to instructions of the NanoAmp™ RT-IVT Labeling Kit from Applied Biosystems. DIG-labeled cRNA was then subsequently hybridized to Applied Biosystems Genome Survey Microarrays.

#### Globin reduction protocol using peptide nucleic acids during amplification & labeling of RNA

Similar to the latest Affymetrix–Preanalytix protocol (i.e., GeneChip® Globin-Reduction Kit Handbook), we used PNA oligomers in the Applied Biosystems amplification/labeling protocol to block reverse transcription (RT) of the globin transcripts. To reduce globin mRNA, 0.9–1 µg total RNA and dT Promoter Primer mix was incubated with the globin-specific PNAs for 10 min instead of 5 min at 70°C and then cooled to 4°C. Sequences of the four PNA oligos and preparation of the PNA mix were given in the GeneChip Globin-Reduction Kit Handbook by Affymetrix–Preanalytix. During the subsequent first-strand cDNA synthesis reaction, the PNAs bind to the globin mRNA, blocking reverse transcription of the corresponding transcripts. This step is followed by second-strand synthesis and cDNA purification. An IVT-based amplification then converts cDNA into DIG-labeled cRNA according to instructions of the NanoAmp RT-IVT Labeling Kit.

#### Globin reduction protocol using GlobinClear™ kit from Ambion & amplification/labeling of RNA

α- and β-hemoglobin transcripts are depleted from the blood total RNA samples using the magnetic bead-based GlobinClear Human kit (Ambion, Inc., Texas, USA) according to the manufacturer's instruction. The Ambion protocol is inserted after total RNA extraction and before cDNA synthesis. To remove globin mRNA, 1.4–2.6 µg of total RNA (in 14 µl) from human whole blood is mixed with a biotinylated Capture Oligo Mix in hybridization buffer. Streptavidin Magnetic Beads are then added to bind the biotinylated oligonucleotides, thereby capturing the globin mRNA on the magnetic beads. The Streptavidin Magnetic Beads are then pulled to the side of the tube with a magnet and the RNA, depleted of the globin mRNA, is transferred to a fresh tube. The RNA is further purified adding an RNA Binding Bead suspension to the samples, and using magnetic capture to wash and elute the GLOBINclear-processed RNA. Total RNA integrity after globin reduction with Ambion kit is assessed using Bio-Rad Experion™ Automated Electrophoresis Station. Next, 0.5–0.8 µg RNA is amplified and labeled following the NanoAmp RT-IVT Labeling protocol converting RNA into DIG-labeled cRNA for hybridization to Applied Biosystems Genome Survey Microarrays.

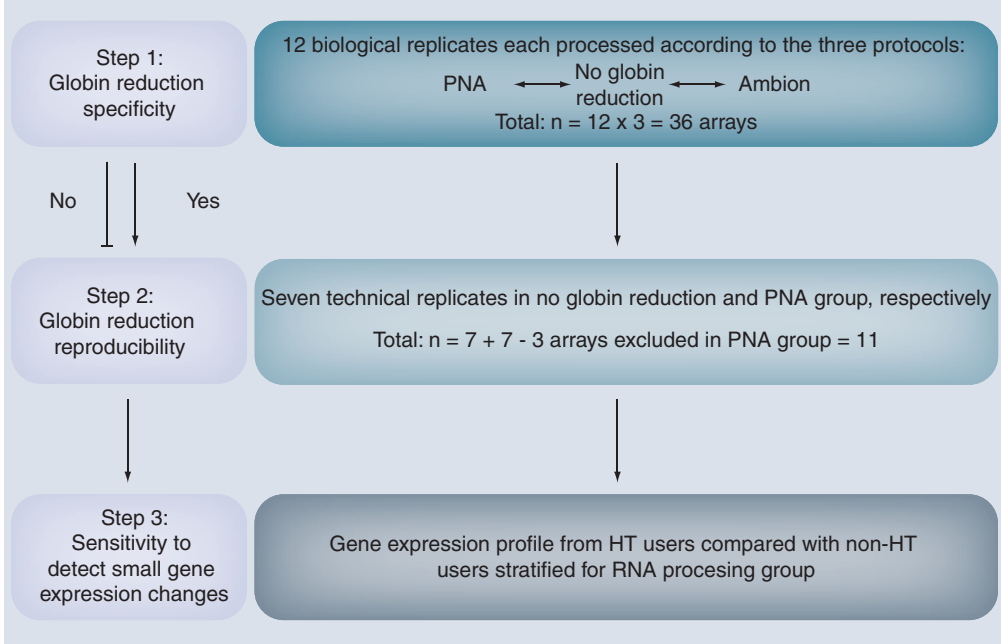
### *Microarray procedure & image analysis*

The concentration of all amplified cRNA was measured using a NanoDrop Spectrophotometer and the integrity was assessed using a Agilent 2100 Bioanalyzer or Bio-Rad Experion Automated Electrophoresis Station.

Array hybridization, chemiluminescence detection, image acquisition and analysis were performed using Applied Biosystems Chemiluminescence Detection Kit and Applied Biosystems 1700 Chemiluminescent Microarray Analyzer following the manufacturer's protocol.

A total of 10 µg of labeled cRNA targets from each sample were first fragmented, mixed with internal control target (24-mer oligo labeled with LIZ® fluorescent dye) and hybridized to a

**Figure 1. Study design.**




HT: Hormone therapy; PNA: Peptide nucleic acid.

prehybridized microarray at 55°C for 16 h. After washing to remove unhybridized DIG-labeled molecules, an alkaline phosphatase–antibody conjugate is added to bind to the DIG-labeled target. The addition of substrate and a chemiluminescence enhancer initiates the chemiluminescent reaction. Eight images are collected for each microarray using the 1700 analyzer, including two 'short' chemiluminescent images (5 s exposure length each) and two 'long' chemiluminescent images (25 s exposure length each) for gene expression analysis, two fluorescent images for feature finding and spot normalization and two QC images for spectrum cross-talk correction. Applied Biosystems Expression System software was used to extract signal intensities, signal to noise ratios (S/N) and flagging values from the microarray images. The Applied Biosystems Human Genome Survey Microarray V2.0 contains 277 control probes and 32,878 probes for the interrogation of 29,098 genes.

Data analysis was performed using R [102], an open-source-interpreted computer language for statistical computation and graphics, and tools from the Bioconductor project [103], adapted to our needs.

Using Applera package in R, we set the filtering criteria so that each gene had a flagging value less than two and a S/N greater than three in at least 50% of the samples. After spot filtration we proceed with quantile normalization. Because the technical replicate pairs are positively correlated, technical replicates were only included in the unsupervised analysis (i.e., principal component analysis [PCA] and hierarchical clustering) and excluded in analyses to identify gene differentially expressed between the three protocols or after HT use. To study HT use effect on gene expression profile, the reduction approach of each globin was analyzed as an independent group (one stratum). Spot filtration and normalization following the same criteria were in this case conducted for each stratum separately.

We conducted a Limma analysis that uses a variant of linear models with an empirically moderated estimate of the standard error [17]. We also used global analysis of covariance (Ancova) [18], which is carried out by comparison of linear models via the extra sum of squares principle. P-values are estimated by permutation of sample numbers and recalculations of the F-value.

Finally, we conducted a gene set enrichment analysis [19], which is a computational method that determines whether an *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g., HT vs non-HT users). We used curated gene sets extracted from the Molecular Signature

**Table 1. RNA concentration and purity.**

| Parameter | No globin reduction (mean [SD]) | PNA (mean [SD]) | Ambion (mean [SD]) |
|---|---|---|---|
| Total RNA concentration (ng/μl) | 136.7 (27.6) | 136.7 (27.6) | 63.5 (13.2) |
| Total RNA 260/280 ratio | 2.06 (0.01) | 2.06 (0.01) | 1.98 (0.05) |
| Total RNA 260/230 ratio | 1.70 (0.37) | 1.70 (0.37) | 1.12 (0.38) |
| Total RNA starting yield (μg) in nine μl maximum | 0.97 (0.06) | 0.97 (0.06) | 0.58 (0.11) |
| cRNA yield (μg) | 74.3 (13.3) | 38.1 (8.1) | 15.6 (5.3) |
| Adjusted cRNA yield (μg)* | 76.6 | 39.3 | 26.9 |
| cRNA 260/280 ratio | 2.20 (0.05) | 2.20 (0.07) | 2.14 (0.11) |
| cRNA 260/230 ratio | 1.77 (0.27) | 1.39 (0.40) | 0.80 (0.28) |

**Adjusted for 1 μg total RNA at start.*
*PNA: Peptide nucleic acid; SD: Standard deviation.*

Database [19], and previously identified as upregulated after HT use with a p-value of less than 0.02 [10].

## Results

### *Study design*

In our study, we processed 12 blood samples (PAXgene tubes from six HT and six non-HT users) according to three different protocols: no globin reduction, globin reduction using PNAs and globin reduction using magnetic beads in the GlobinClear kit (Figure 1). Whenever possible, the same sample that was processed according to the three protocols was amplified and hybridized in the same round. The reproducibility of the globin reduction methods compared with no globin reduction approach was investigated for the PNA method only since this globin reduction method was the most effective and specific (see results below). Seven technical replicates in each group (i.e., no globin reduction and PNA groups) were performed in two amplification rounds (Total: [4 + 3] × 2 = 14; Figure 1). During one amplification round, all three samples with PNAs failed to reverse transcribe long fragments of mRNA due to inhibitory contamination and these arrays were therefore excluded from further analyses. Finally, each RNA processing method was analyzed as an independent group studying blood-derived gene expression changes associated with HT use (Figure 1). Results were then compared in order to determine which protocol offers the highest sensitivity to detect small gene expression changes.

A total of 47 arrays were conducted, including seven and four technical replicates in no globin reduction and PNA group, respectively. Microarray data with accession number GSE7008 are online at the GEO database [104].

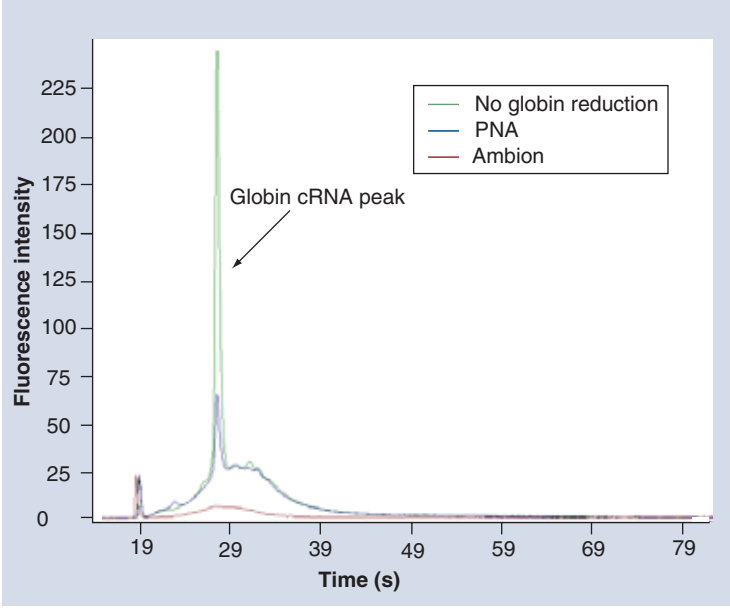
### *Purity & yield of RNA*

After globin reduction using the Ambion kit, 55–100% of total RNA is retrieved. Total RNA purity after Ambion globin reduction is slightly lowered according to the 260/230 ratio (Table 1). Before amplification, total RNA integrity after globin reduction via the Ambion kit was assessed using a Bio-Rad Experion Automated Electrophoresis Station. A slight decrease in 28S peak may suggest a minor RNA degradation after globin reduction using the Ambion kit. However, overall quality of total RNA was estimated sufficient to conduct gene expression analyses (data not shown). Owing to the lower concentration of total RNA after globin reduction with the Ambion kit and the volume limitation of 9 μl by the protocol, the RNA input for amplification is decreased compared with the RNA yield processed without globin reduction or with globin reduction using PNAs (Table 1). After adjustment for the RNA input, the yield of cRNA following globin reduction using the Ambion kit remains the lowest with poor 260/230 purity (Table 1). The equivalent of 50% of cRNA amplified without globin reduction is obtained when conducting globin reduction using PNAs (Table 1).

### *Efficiency & specificity of globin reduction protocols*

Figure 2 shows the results from the Agilent bioanalyzer of one sample processed following the three protocols (i.e., without globin reduction, with globin reduction using PNAs, and with globin reduction using the Ambion kit). This example is representative of results we obtained for all 12 samples (data not shown). Without globin reduction, amplified cRNA from whole-blood samples shows a narrow peak at approximately 29 s that represents amplified globin

**Figure 2. Representative results from Agilent 2100 Bioanalyzer of one sample subjected to three RNA processing protocols.**




Without globin reduction (in green), with globin reduction using PNAs (in blue) and using Ambion kit (in red).
PNA: Peptide nucleic acid.

RNA (Figure 2). Globin-reduced RNA demonstrated a different amplified cRNA profile, in which the distinctive globin peak is either reduced when using PNAs or absent using Ambion kit (Figure 2). Fluorescence intensity at the top of the smooth curve representing the remaining amplified cRNA is lower after globin reduction using the Ambion kit (intensity = 5; Figure 2) than for the two other methods (intensity = 20–25, both without and with globin reduction using PNAs; Figure 2), which raises the question of globin reduction specificity when using the Ambion kit.

The percentage of probes with S/N ratio of three or greater (i.e., present calls) was significantly higher for globin-reduced samples using PNAs (mean = 55.4% ± 3.72) compared with no globin reduction (mean = 52.1% ± 4.17; t-test p-value = 0.02) and the Ambion kit (mean = 51.4% ± 3.45; t-test p-value = 0.007) (Figure 3A). Reproducibility of the PNA-based protocol was slightly lower than in replicates without globin reduction (Figure 3B). In parallel, one sample (sample 34) was twice amplified with PNAs on the same day to study technical reproducibility using the same PNA mix solution. Correlation between these two samples was equal to 0.99. For biological replicates, both globin reduction approaches using either PNAs (coefficient of variation median = 3.4%) or the Ambion kit (coefficient of variation median = 3.6%) offer a slight decrease in variability compared with no globin reduction approach (coefficient of variation median = 4.1%).

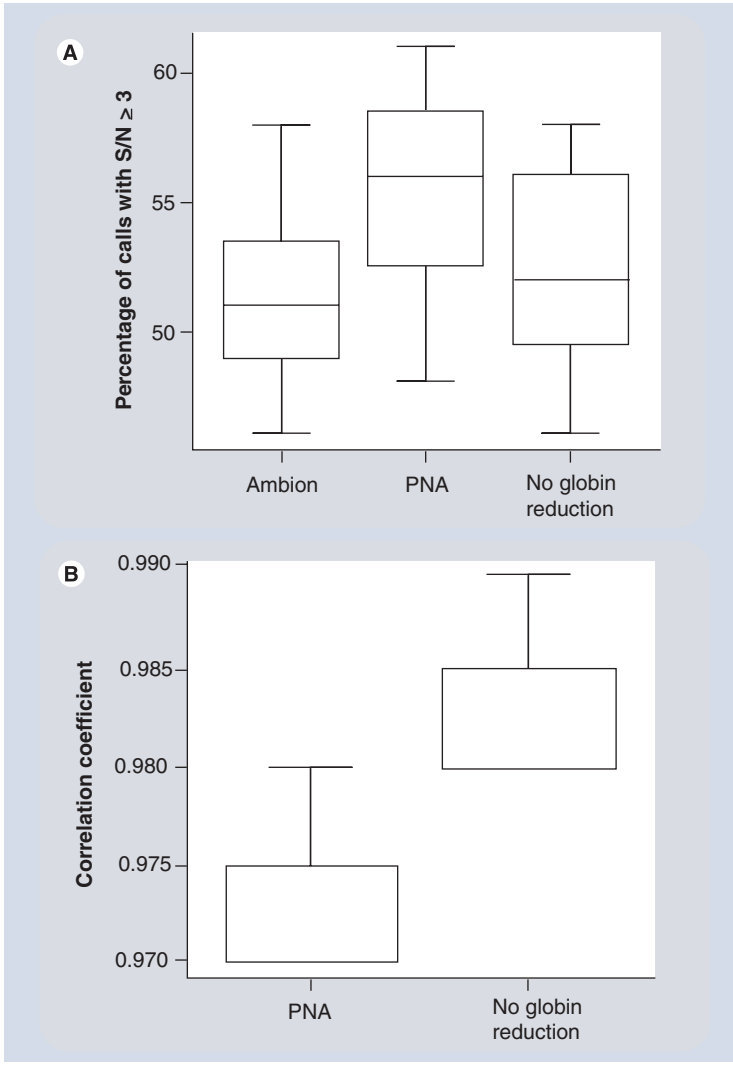
The PCA plot (Figure 4A) shows that samples processed by Ambion kit cluster very distinctively from samples processed without globin reduction and from the globin-reduced samples using PNAs. Samples amplified with PNAs constitute an intermediary group closest to the no globin reduction group (Figure 4A). Average linkage hierarchical clustering analyses using Euclidean distance (Figure 4B) shows that first technical replicates processed without globin reduction and with globin reduction using PNAs cluster well together regardless of the RNA processing method and second that samples processed by Ambion kit cluster together according to the RNA separately.

The Venn diagram in Figure 5A represents the overlap in genes identified as differentially expressed according to RNA processing method with a false discovery rate (FDR) of 0.05 or less by Limma analysis [17] adjusted for multiple comparisons across contrasts. In line with the PCA results, many genes (n = 7253 probes; Figure 5A) are differentially expressed after globin reduction with the Ambion kit with a FDR of 0.05 or less. Signal intensities are decreased in comparison with signal intensities obtained without globin reduction for 36.9% of these probes. This might be attributable to slight degradation of total RNA and/or unspecific binding of globin mRNA using the Ambion protocol. The PNA method changes expression of fewer genes (n = 2834 probes; Figure 5A) and most of them are also found differentially expressed after globin reduction using the Ambion kit. The heatmap in Figure 5B represents the row-scaled signal intensities of the 1616 probes found commonly differentially expressed using the three different protocols.

### *Effect on global gene expression in relation to hormone exposure*

After spot filtration (i.e., filter out spot with flags more than two and S/N <3 in 50% or more of the samples) was carried out independently for all three RNA processing groups, we had a total of 16,331, 17,289 and 16,505 probes in no globin reduction, PNA and Ambion datasets, respectively. Each group included the same 12 women, of whom 50% were using HT at the time of blood sampling.

**Figure 3. Percentage of present calls and technical reproducibility for the different RNA processing methods.**




Boxplot representations of **(A)** percentage of spots with S/N ratio ≥ 3, and **(B)** correlation coefficient between technical replicates according to each RNA processing approach. Technical replicates were run in no globin reduction and PNA groups only.
PNA: Peptide nucleic acid; S/N: Signal to noise.

The ANCOVA global test is a test for the association between expression values for all genes and HT use (see methods). Although all permuted p-values are nonsignificant, the association between HT use and gene expression profile appears stronger without globin reduction (permuted p-value = 0.18) compared with globin reduction approaches (permuted p-value = 0.47 and 0.58 for PNA and Ambion protocol, respectively).

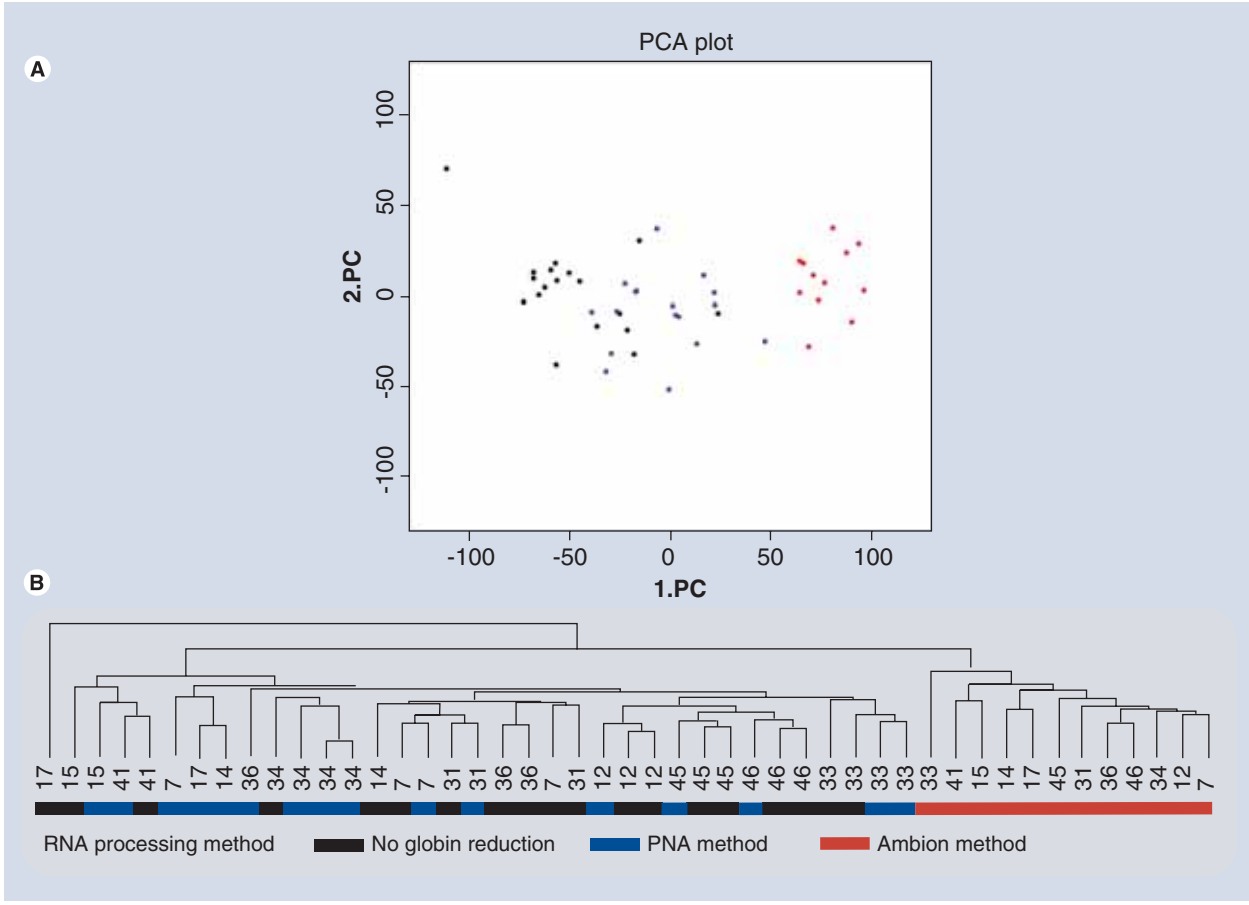
We have conducted enrichment analyses in order to palliate limitations of single gene approaches, especially in small sample size. No gene set from the Molecular Signature Database had a significant enrichment score with FDR of less than 0.25. Figure 6A summarizes the enrichment test in HT users of genes related to estrogen–androgen and steroid metabolism pathways from KEGG [20] and genes upregulated after estradiol exposure identified by Frasor *et al.* [21]. The leading edge subset that can be interpreted as the core of a gene set that accounts for the enrichment signal consists of 11, ten and seven genes in the no globin reduction, PNA and Ambion globin reduction groups, respectively. For all three protocols used, this gene set is not significantly associated with HT exposure, although the gene set appears to be more strongly related to HT use in the no globin reduction approach ($p = 0.35$; FDR = 0.36; Figure 6A). Reassuringly, 50% of the genes found in one leading-edge subset were found in common with one other leading-edge subset studying one other globin RNA processing method (Figure 6A).

In parallel, the set of genes (n = 75 with known gene symbols) previously identified as upregulated after HT use with $p < 0.02$ [10] in our previous dataset are significantly associated with HT use, with a FDR equal to 0.04 for samples processed without globin reduction (Figure 6B). The corresponding FDR was equal to 0.12 and 0.68 when studying globin reduced samples using PNA and Ambion protocol, respectively (Figure 6B). Similarly, in this analysis, 85–92% of the genes in one leading edge subset were found in common with another leading-edge subset for one other globin RNA processing method (Figure 6B).

## Discussion

In our study, both globin reduction protocols using either PNAs or the Ambion kit were mostly efficient at reducing globin RNA from cRNA. However, the Ambion kit significantly lowers the total amount of remaining cRNA given distinct gene expression profiles. The sensitivity of transcript detection is significantly improved using PNAs compared with the other two approaches. Signal variability is slightly decreased after globin reduction using PNAs compared with other approaches; however, reproducibility in technical replicates is lowered compared with no globin reduction. Gene expression profiles obtained after globin reduction using PNAs were closer to profiles obtained without globin reduction, although several genes were significantly differentially expressed between the two groups. Overall, we did not notice any sign of stronger association between gene expression profile and HT exposure after globin reduction.

**Figure 4. Unsupervised analyses.**



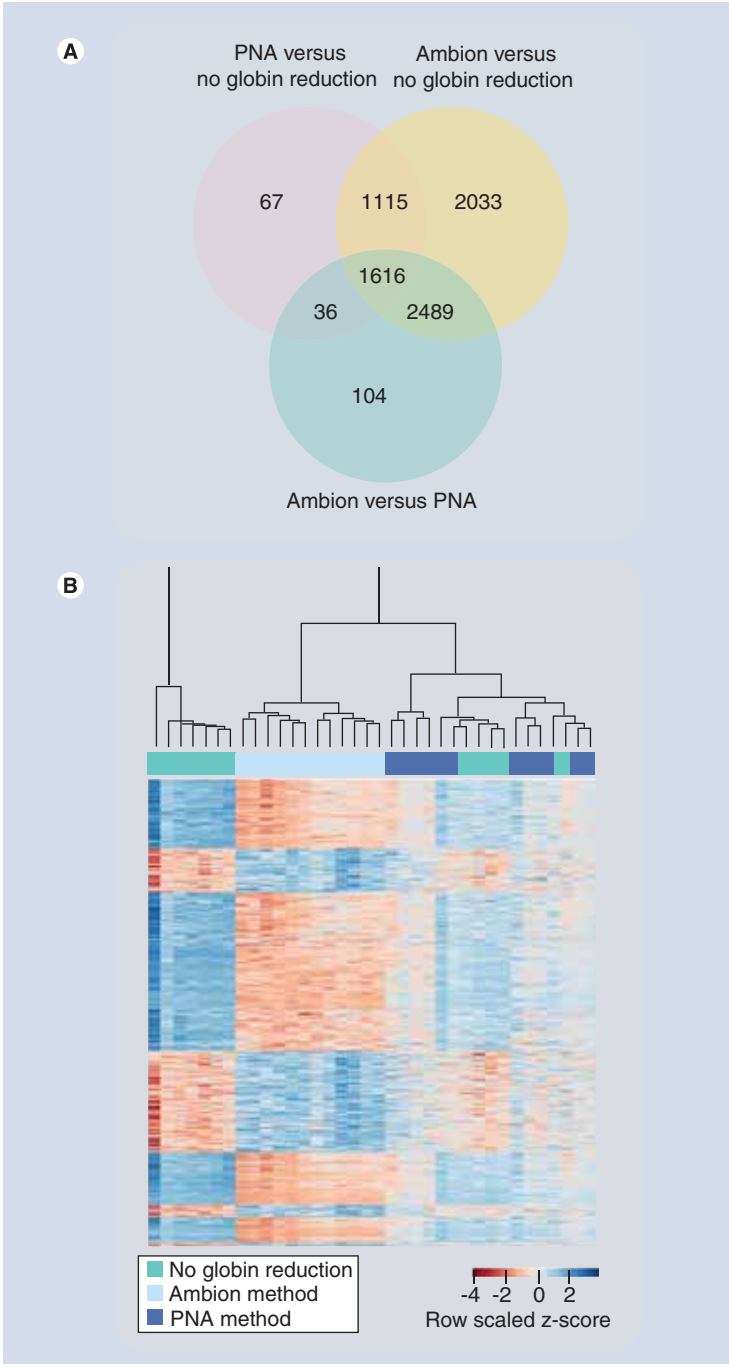

**(A)** Principal component analysis and **(B)** average linkage clustering using Euclidean distance performed on blood gene expression from sample without globin reduction (in black), with globin reduction using PNAs (in blue) and using Ambion kit (in red). Each sample represented by one number in **(B)** has been processed at least once according to the three different protocols.
PCA: Principal component analysis; PNA: Peptide nucleic acid.

A potential issue when employing blood RNA to conduct microarrays in large molecular epidemiology studies is that RNA stabilization [13] and high sensitivity in detection of small gene expression changes [10,14,22] are required. Others have shown that increasing the amounts of globin mRNA in stabilized total RNA resulted in decreasing detection of transcripts [15,23]. Globin reduction with RNase H digestion of globin oligonucleotide/globin mRNA hybrids was criticized for a potential nonspecific deletion of transcripts [14], although this was not observed in a recent study [23]. In response to this criticism, the alternative method using PNAs was introduced (i.e., Affymetrix-Preanalytix protocol, GeneChip® Globin-Reduction Kit Handbook). The Ambion protocol is inserted after total RNA extraction and before cDNA synthesis and is therefore time consuming. In some cases, we had to concentrate the globin-reduced RNA in samples previously selected for containing highest amounts of starting RNA. The PNA-based protocol is performed in the same tube as the cDNA synthesis reaction and is therefore easy to perform. However, PNAs remain expensive and the reproducibility of the method was found to be more difficult compared with the reference, likely owing to challenges when handling PNAs. Measures to prevent their aggregation and precipitation were used, but PNAs remain difficult to dissolve and the exact concentration of each PNA added in the amplification reaction is difficult to assess.

In accordance with previous studies [15,23], but to a lesser extent, we found that globin reduction using PNAs offers a lower variability in signal intensity and a higher percentage of detected transcripts with S/N more than three

compared with no globin reduction. The Ambion kit was more efficient than PNA globin reduction with total removal of the globin peak, but this method also significantly lowered the rest of amplified cRNA. In accordance with previous published data [15], our results demonstrate that globin reduction induces distinct gene expression profiles, but globin reduction using PNAs brought the profile closer to the original profile (i.e., without globin reduction). This was certainly due to the fact that the PNA protocol only partly reduced the presence of globin RNAs in our study.

**Figure 5. Genes differentially expressed between the different RNA processing methods.**

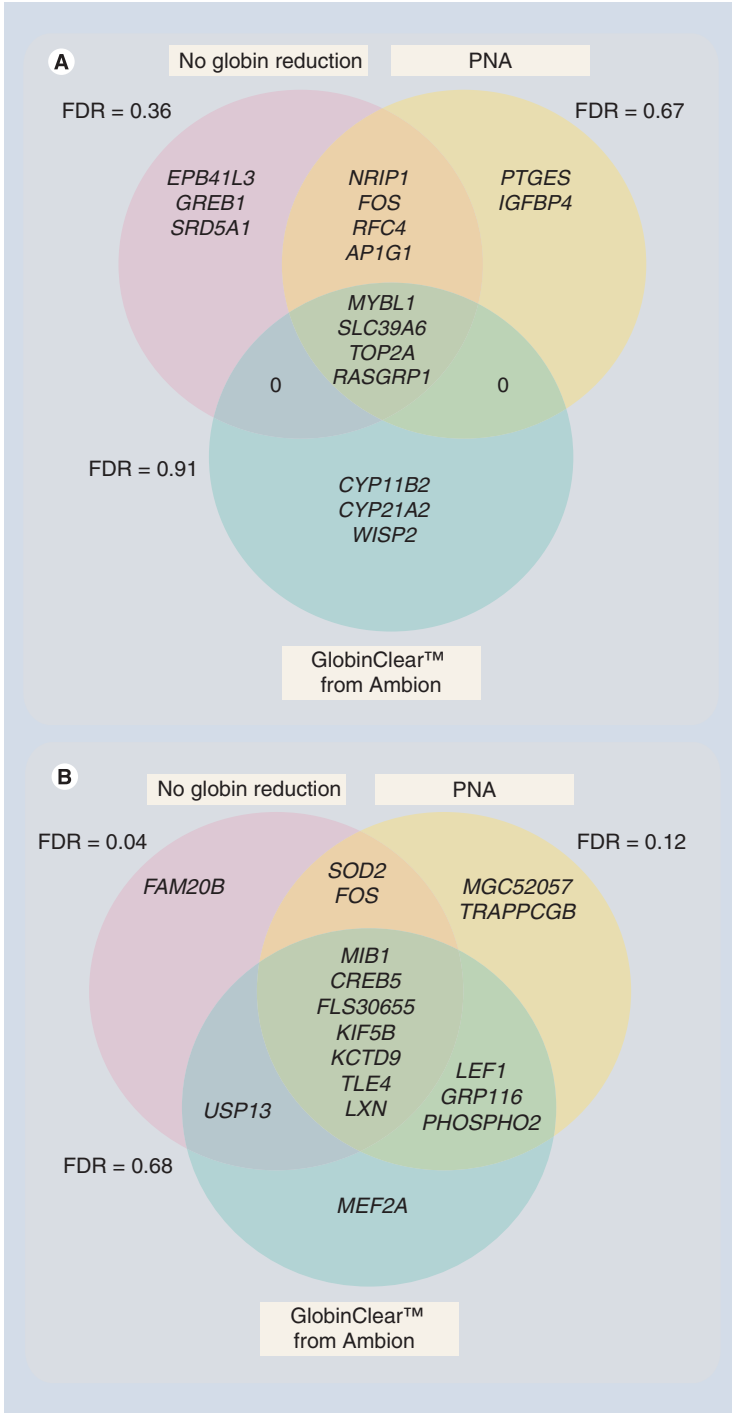



**(A)** Venn diagram showing the overlap of genes differentially expressed between the RNA processing methods with a FDR of 0.05 or less by Limma analysis adjusted for multiple comparisons across contrasts. **(B)** Heatmap representing the row-scaled $\log_2$ signal intensities of the 1616 probes found commonly differentially expressed in all three comparisons (cf venn diagram).
FDR: False discovery rate; PNA: Peptide nucleic acid.

Although gene expression profile after globin reduction differs significantly from the profile without globin reduction, the overall goal is to provide potential higher sensitivity to detect small gene expression changes. We hypothesized that microarrays could be used to identify changes in blood gene expression followed by hormone exposure in women [10]. In our study, the small sample size in each RNA processing group (n = 12) made results unstable and conclusions concerning gene expression changes after hormone use were difficult to draw. White blood cells have been defined as the most transcriptionnally active cells in blood and may give the most sensitive gene expression profiles in response to defined factors [24]. However, this pilot study aimed to determine the most sensitive protocol for conducting gene expression analyses from whole blood using Applied Biosystems platform rather than to draw conclusions about gene expression changes after HT use. We observed that the amplification date had a significant effect on the gene expression profile especially in the PNA approach with lowered reproducibility, but reassuringly most genes counted in the gene set enrichment results for one globin RNA processing method were found in common with one other globin RNA processing method [25]. Furthermore, the set of genes previously identified as upregulated after HT use with a p-value of less than 0.02 [10] is significantly associated with HT user with a FDR equal to 0.04 for samples processed without globin reduction. As expected, this gene set is most strongly related to HT exposure in the no globin reduction dataset since the previous study was, similarly, conducted on whole blood without globin reduction. Overall, no indication of a stronger association between gene expression profile and hormone exposure after globin reduction was noted.

These conclusions differ somewhat from previous results including 14 healthy individuals showing an increase in accuracy and confidence

**Figure 6. Venn diagram showing the overlap of genes contributing to the gene set enrichment score in hormone therapy users.**




The gene set includes **(A)** genes related to steroid metabolism and estrogen receptor responders* and **(B)** genes previously identified as upregulated after hormone therapy use with $p < 0.02$ [10]. *Information was extracted from Kyoto Encyclopedia of Genes and Genomes [20] and from the Molecular Signature Database [19], which contains curated gene sets [21].
FDR: False discovery rate; PNA: Peptide nucleic acid.

of prediction of gender differences after globin reduction [23]. The reported discrepancy is most likely due to the difference in various microarray platforms. To our knowledge, all studies investigating the gene expression profile before and after globin reduction were conducted on the Affymetrix platform [15,23,26–28]. The percentage of detected transcripts after globin reduction on Affymetrix arrays was more significantly improved compared with our results on Applied Biosystems arrays. These two microarray platforms are different in many aspects, for example probe design and type of signal measured (chemiluminescence vs fluorescence). Therefore, the appropriate RNA processing method should be determined according to the microarray platform to be used. However, this will pose a problem regarding standardization of the methods since globin-reduced and no globin-reduced approaches produce distinct gene expression profiles that are not comparable. This finding has important implications for the interpretation of results from clinical studies employing different RNA stabilization and isolation strategies for the evaluation of gene expression patterns in blood.

## Conclusion

In conclusion, globin RNA processing method in blood transcriptome analyses should be optimized for each microarray platform. In our case, globin reduction was not beneficial using the Applied Biosystems microarray. According to these results, another larger study including around 300 blood samples is ongoing without globin reduction using Applied Biosystems platform to identify changes in gene expression changes followed by hormone exposure in women. Our results also emphasize the benefit of working with gene sets rather than single gene approaches, giving more stable results across studies and urging us to better understand blood gene expression profile in order to work with the most accurate gene set(s).

## Future perspective

Such studies are an essential requisite for establishing blood collection and RNA isolation protocols in large-scale studies for the investigation of gene expression patterns in whole blood. Uncertainty remains about the use of gene expression profiles from whole blood as surrogate tissue to biomonitor-defined exposure (i.e., hormone) and its association to disease risk (i.e., breast cancer).

However, we believe that gene expression profiling from blood in human populations could assist us in:

- Measuring exposition and defining outcome
- Understanding mode of action
- Understanding the etiology of environmentally induced disease
- Improving risk assessment methods and models

### Financial & competing interests disclosure

*This work was in part supported by grants from the Norwegian Research Council. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.*

*No writing assistance was utilized in the production of this manuscript.*

### Ethical conduct of research

*The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.*

**Executive summary**

- In our study, globin reduction does not offer higher sensitivity to detect biomarkers of hormone exposure in blood transcriptome.
- Both globin reduction protocols using either peptide nucleic acids (PNAs) or magnetic beads were mostly efficient at reducing globin RNA from cRNA.
- The Ambion kit using magnetic beads significantly lowers the total amount of remaining cRNA, which is an indication of low specificity.
- The sensitivity of transcript detection is significantly improved using PNAs compared with the other two approaches.
- Signal variability is slightly decreased after globin reduction using PNAs compared with other approaches; however, reproducibility in technical replicates was lowered compared with no globin reduction.
- Globin RNA processing methods in blood transcriptome analyses should be optimized for each microarray platform. In our case, globin reduction was not found to be beneficial using the Applied Biosystems microarray.
- Analysis with gene sets rather than single-gene approaches gives more stable results across studies.

### Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Twine NC, Stover JA, Marshall B *et al.*: Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. *Cancer Res.* 63, 6069–6075 (2003).
2. Tang Y, Nee AC, Lu A, Ran R, Sharp FR: Blood genomic expression profile for neuronal injury. *J. Cereb. Blood Flow Metab.* 23, 310–319 (2003).
3. Gladkevich A, Kauffman HF, Korf J: Lymphocytes as a neural probe: potential for studying psychiatric disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 28, 559–576 (2004).
4. Achiron A, Gurevich M, Friedman N, Kaminski N, Mandel M: Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. *Ann. Neurol.* 55, 410–417 (2004).
5. Burczynski ME, Twine NC, Dukart G *et al.*: Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma. *Clin. Cancer Res.* 11, 1181–1189 (2005).
6. Lampe JW, Stepaniants SB, Mao M *et al.*: Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev.* 13, 445–453 (2004).
7. Wang Z, Neuburg D, Li C *et al.*: Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ. Health Perspect.* 113, 233–241 (2005).
8. Forrest MS, Lan Q, Hubbard AE *et al.*: Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. *Environ. Health Perspect.* 113, 801–807 (2005).
9. Amundson SA, Do KT, Shahab S *et al.*: Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat. Res.* 154, 342–346 (2000).
10. Dumeaux V, Johansen J, Borresen-Dale AL, Lund E: Gene expression profiling of whole-blood samples from women exposed to hormone replacement therapy. *Mol. Cancer Ther.* 5, 868–876 (2006).

•• First report investigating whole blood gene expression profile associated with hormone exposure in postmenopausal women.

11. Thach DC, Lin B, Walter E *et al.*: Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *J. Immunol. Methods* 283, 269–279 (2003).
12. Feezor RJ, Baker HV, Mindrinos M *et al.*: Whole blood and leukocyte RNA isolation for gene expression analyses. *Physiol. Genomics* 19, 247–254 (2004).

•• Good report investigating the reliability of RNA from whole blood or its leukocyte compartment by examining their reproducibility, variance and signal-to-noise ratios.

13. Debey S, Schoenbeck U, Hellmich M *et al.*: Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. *Pharmacogenomics J.* 4, 193–207 (2004).
14. Burczynski ME, Dorner AJ: Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. *Pharmacogenomics* 7, 187–202 (2006).
15. Liu J, Walter E, Stenger D, Thach D: Effects of globin mRNA reduction methods on gene expression profiles from whole blood. *J. Mol. Diagn.* 8, 551–558 (2006).
• First paper to compare the effect of two different globin reduction protocols on gene expression profile measurements.
16. Lund E, Dumeaux V, Braaten T *et al.*: Cohort Profile: The Norwegian Women and Cancer Study NOWAC Kvinner og kreft. *Int. J. Epidemiol.* DOI 10.1093/ije/dym137 (2007).
17. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article 3 (2004).
• Linear analysis of microarray data.
18. Mansmann U, Meister R: Testing differential gene expression in functional groups. Goeman's global test versus an ANCOVA approach. *Methods Inf. Med.* 44, 449–453 (2005).
19. Subramanian A, Tamayo P, Mootha VK *et al.*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* 102, 15545–15550 (2005).
•• Computational approach method that determines whether an *a priori* defined set of genes shows statistically significance. Palliate limitations of single-gene approaches.
20. Kanehisa M, Goto S: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30 (2000).
21. Frasor J, Stossi F, Danes JM, Komm B, Lyttle CR, Katzenellenbogen BS: Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer Res.* 64, 1522–1533 (2004).
22. Rockett JC, Kavlock RJ, Lambright CR *et al.*: DNA arrays to monitor gene expression in rat blood and uterus following 17β-estradiol exposure: biomonitoring environmental effects using surrogate tissues. *Toxicol. Sci.* 69, 49–59 (2002).
• First report comparing blood gene expression profile to tissue gene expression profile (i.e., uterine tissue).
23. Debey S, Zander T, Brors B, Popov A, Eils R, Schultze JL: A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials. *Genomics* 87, 653–664 (2006).
24. Li L, Ying L, Naesens M *et al.*: Interference of globin genes with biomarker discovery for allograft rejection in peripheral blood samples. *Physiol. Genomics* (2007).
•• Recent report investigating effect of globin reduction to detect biomarker for allograph rejection in whole blood gene expression. Only report that relates this globin reduced profile to gene expression profile from mononuclear blood cells.
25. Manoli T, Gretz N, Grone HJ, Kenzelmann M, Eils R, Brors B: Group testing for pathway analysis improves comparability of different microarray datasets. *Bioinformatics* 22, 2500–2506 (2006).
26. Affymetrix Technical Note: Globin reduction protocol: a method for processing whole blood RNA samples for improved array results. (2003).
27. Ambion TechNotes 12(3): Improved Microarray Sensitivity Using Whole Blood RNA Samples (2005).
28. Expression analysis Technical Note: Globin RNA Reduction in Blood Samples. (2005).

## Websites

101. The Norwegian Women and Cancer Study (2007) http://uit.no/kk/NOWAC/
102. The Comprehensive R Archive Network http://cran.r-project.org
103. Bioconductor – open source software for bioinformatics http://www.bioconductor.org
104. Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo