

Short communication

Gene expression measurements in the context of epidemiological studies

Background: Gene expression measurements became an attractive tool to assess biological responses in epidemiological studies. However, collection of blood samples poses various technical problems. We used gene expression data from two epidemiological studies to evaluate differences between sampling methods, comparability of two methods for measuring RNA levels and stability of RNA samples over time.

Methods: For the PARSIFAL study, PBLC of 1155 children were collected using EDTA tubes in two countries. In the PASTURE study, tubes containing RNA-stabilizing solutions (PAXgene® Blood RNA Tubes; PreAnalytiX) were used to collect cord blood leucocytes of 982 children in five countries. Real-time PCR (conventional single tube assay and high-throughput low density arrays) was used to quantify expression of various innate immunity genes. In 77 PARSIFAL samples, gene expression was measured repeatedly during prolonged storage.

Results: In PARSIFAL (EDTA tubes) the median RNA yield after extraction significantly differed between the two centres (70 and 34 ng/µl). Collecting blood into an RNA-stabilizing solution markedly reduced differences in RNA yield in PASTURE (range of medians 91–107 ng/µl). The agreement [Spearman rank correlation (r)] between repeated measurements of gene expression decreased with increasing storage time [e.g., for CD14: r (first/second measurement) = 0.35; r (first/third measurement) = 0.03]. RNA levels measured with either the conventional method or low-density arrays were comparable ($r > 0.9$).

Conclusion: Collecting blood samples into tubes containing an RNA-stabilizing solution increases RNA yield and reduces its variability. Long-term storage of samples may lead to RNA degradation, requiring special attention in longitudinal studies.

C. Bieli¹, R. Frei², V. Schickinger², J. Steinle², C. Bommer², S. Loeliger², C. Braun-Fahrländer¹, E. von Mutius³, G. Pershagen⁴, R. Lauener²

¹Institute of Social and Preventive Medicine, University of Basel, Basel, Switzerland; ²University of Zurich, Children's Hospital, Zurich, Switzerland; ³University Children's Hospital, Munich, Germany; ⁴Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

Key words: blood sampling; epidemiology; gene expression; innate immunity; PAXgene; RNA degradation

Christian Bieli, MD
Institute of Social and Preventive Medicine
University of Basel
Steinengraben 49
4051 Basel
Switzerland

C. Bieli and R. Frei contributed equally

Accepted for publication 10 March 2008

High-throughput genomic technologies allow to measure large-scale gene expression profiles in a time-efficient way. These methodological advances have spread the range of potential applications and allow gene expression analyses to be used in the context of epidemiological studies (1, 2).

However, in epidemiological study settings, other sources of measurement error have to be considered than in clinical or laboratory research: distance between location of blood collection and further processing,

different fieldworkers involved in the studies, and, particularly in international studies, different laboratories and laboratory staff are likely to affect quality of the obtained gene expression data.

In the present study, we had the possibility to analyse gene expression data of two epidemiological studies [PARSIFAL (3) and PASTURE (4)]. We compared centre-associated variations in RNA quantities obtained using two sampling methods. Furthermore, we tested the comparability of low-density arrays, a high-throughput method, with the conventional single tube assay to assess gene expression. Finally, the PARSIFAL study allowed us to follow up RNA quality, as multiple measures of the same samples at different time points were available. The aim was to evaluate sources of measurement error in gene expression data in order to avoid these factors in future investigations.

Abbreviations: CBLC, cord blood leucocytes; EDTA, ethylene diamine tetraacetic acid; LDA, low density array; PARSIFAL, prevention of allergy risk factors for sensitization in children related to farming and anthroposophic lifestyle; PASTURE, protection against allergy: study in rural environments; PBLC, peripheral blood leucocytes; rtPCR, real-time polymerase chain reaction; STA, (conventional) single tube assay.

Methods

Population

We analysed data from farmers' and nonfarmers' children enrolled in the cross-sectional study PARSIFAL (3) (Prevention of Allergy Risk factors for Sensitization in children related to Farming and Anthroposophic Lifestyle) and in the longitudinal study PASTURE (4) (Protection Against Allergy: Study in Rural Environments).

In PARSIFAL, parents from five countries completed a self-administered questionnaire to assess asthma and allergies and exposures associated with farming. Participants were also asked for consent to further investigations, including venous blood sampling. In two countries (A and B), a random sample of 1155 farmers' and nonfarmers' children was drawn for gene expression analyses of innate immunity genes.

PASTURE is a longitudinal study conducted in five European countries (A, B, C, D and E). Each country built a birth cohort of about 100 farmers' and 100 nonfarmers' children. Several detailed questionnaires were completed and cord blood were collected for the measurement of the child's immunological status. Gene expression of innate immunity genes was measured in the total population of 982 children.

Sample collection and RNA extraction

For the PARSIFAL study, PBLC were collected in EDTA blood in 2002. Of the 1155 selected samples, 1140 (98.7%) were available for total RNA isolation using the QIAmp RNA Blood Mini Kit (Qiagen, Hilden, Germany) supplemented with RNase-free DNase (Qiagen). The RNA was solved in water and stored at -80°C .

For the PASTURE study cord blood leucocytes from the 982 children were collected in tubes containing an RNA-stabilizing solution (PAXgene® Blood RNA Tubes; PreAnalytiX, Qiagen, Hilden, Germany) in 2005 and isolated according to the manufacturer's instructions. The RNA was treated with DNase and solved in *Tris*-Buffer and stored at -80°C .

Reverse transcription and gene expression analyses

Reverse transcription was performed with reagents from Applied Biosystems according to the manufacturer's instructions. The RNA concentration had to reach a cut-off concentration of 21 ng/ μl to be used for the reverse transcription.

Two different methods of the quantitative real-time polymerase chain reaction (rtPCR) were used to measure mRNA levels:

1. rtPCR with plates (conventional single tube assay, STA-rtPCR) were performed in a final volume of 30 μl containing the adequate amounts of cDNA, primers, probe and TaqMan Universal MasterMix (Applied Biosystems, Foster City, CA, USA) and were analyzed on an ABI Prism 7700 Sequence Detection System™ (Applied Biosystems). All oligonucleotides were synthesized by Microsynth.
2. rtPCR with low-density arrays (LDA-rtPCR) allows to measure the expression of 48 genes with a small amount of cDNA. The LDA-rtPCR was performed in a final volume of 1 μl and analyzed on an ABI Prism 7900 Sequence Detection System (Applied Biosystems).

Gene expression of CD14, TLR2 and TLR4 was analysed in PARSIFAL samples from countries A and B in March 2002 with STA-rtPCR. The high-throughput method LDA-rtPCR was used to measure the expression of the same genes in a subsample of the PARSIFAL population from country A in November 2004. In

September 2005, gene expression measurement of the genes was repeated in all PARSIFAL samples from country A. For CD14, STA-rtPCR analyses were repeated in December 2005 to test the reproducibility of the different methods.

Statistical analyses

To test for equal proportions of nominal variables in different groups, the Chi-squared (χ^2) test was used. Agreement between numeric variables (gene expression measurements) was measured by the spearman rank correlation coefficient.

All analyses were carried out using R 2.5.0 (R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria).

Results

RNA amounts extracted using two methods of blood collection

Using EDTA tubes for blood sampling in the PARSIFAL study, the median of extracted RNA was twofold higher in country A than in country B (Table 1). Consequently, the proportion of samples below the cut-off value for further gene expression analyses (21 ng/ μl) differed greatly between the two countries: whereas 98.8% of the samples of country A yielded sufficient RNA for gene expression analyses, this was only true for 68.9% of the samples of country B.

In the PASTURE study, blood was sampled in tubes containing RNA-stabilizing solutions; only slight differences in the amount of extracted RNA were observed between the participating countries (Table 1). The country-specific medians were in a narrow range between 91 and 107 ng/ μl . The proportion of samples with sufficient RNA amount for gene expression ranged between 91.1% and 99.4%.

Reproducibility of high-throughput LDA-rtPCR measurements

We measured agreement of rtPCR results of CD14 mRNA levels between the conventional (single tube assay) and the high-throughput method (low-density arrays) by calculating rank correlation coefficients. To exclude interfering effects of the reverse transcription, we

Table 1. RNA yield after RNA extraction using two methods of blood sampling

Country	EDTA tubes		<i>P</i> *	RNA-stabilizing tubes		<i>P</i> *
	Median amount (ng/ μl)	Below 21 ng/ μl [n/N (%)]		Median amount (ng/ μl)	Below 21 ng/ μl [n/N (%)]	
A	70	5/417 (1.2)	<0.0001	91	17/191 (8.9)	0.006
B	34	225/723 (31.1)		100	10/214 (4.7)	
C	—	—		99	1/180 (0.6)	
D	—	—		99	10/200 (5.0)	
E	—	—		107	9/197 (4.6)	

* χ^2 test of equal proportions.

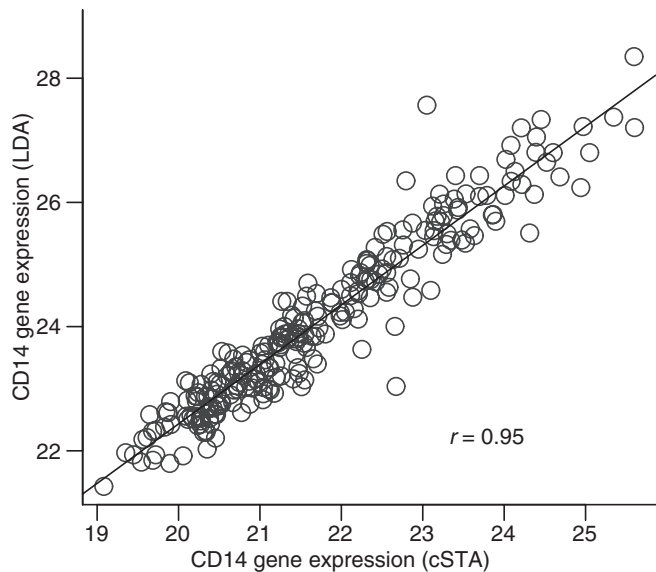


Figure 1. Correlation between CD14 gene expression values resulting from conventional single tube assay cSTA and from low-density arrays LDA using the same CD14 primers and probe set. (r , spearman rank correlation coefficient; straight line, least squares regression line).

measured the expression of the CD14 gene in the same samples either with the conventional STA or with the high-throughput LDA using the same primers and probe set. Correlation of STA-rtPCR and LDA-rtPCR results was high ($r > 0.9$, Fig. 1).

Consistency of rtPCR CD14 gene expression in repeated measurements over time

In 77 samples of the PARSIFAL population of country A gene expression of CD14, TLR2 and TLR4 were analysed repeatedly in 2002, 2004 and 2005. Figure 2 shows agreement (Spearman rank correlation) of gene expression measurements between 2002 and 2004 and between 2002 and 2005 for CD14, TLR2 and TLR4. For all three genes, the agreement of measurements between 2002 and 2004 is moderate (0.35–0.56) and low between 2002 and 2005 (0.03–0.32).

Discussion

We observed that the amount of extracted RNA depends on the method used for blood sampling and that agreement between rtPCR results decreases over storage time. These results suggest that preanalytical procedures and length of storage time are crucial factors for sample quality.

The median amounts of the total extracted RNA amount markedly differed between the two countries studied in PARSIFAL. In this study, blood was collected in EDTA tubes and subsequently transported

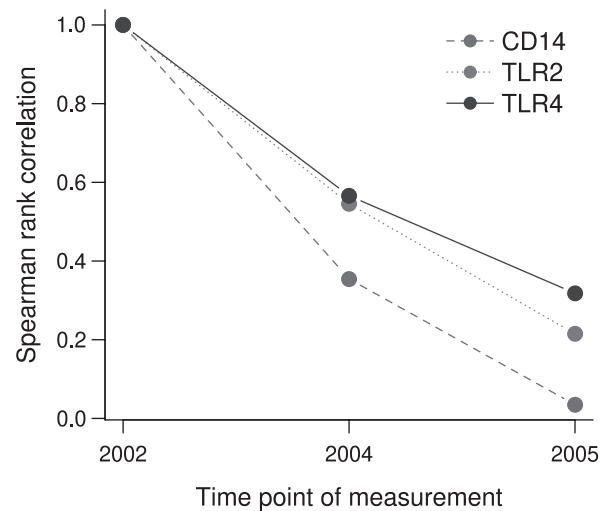


Figure 2. Agreement between repeated measurement of gene expression of CD14, TLR2 and TLR4 during long-time storage. (y-axis, Spearman rank correlation between CD14 gene expression measured in 2002 and subsequent measurements (2004 and 2005); x-axis, time points of measurements).

to labs in the two study centres, where RNA has been extracted. Although maximal attention was given to control temperature during transport, to minimize transportation time and time until RNA extraction in the labs, it seems probable that differences in these parameters translated into different RNA yields between centres. Based on this experience, we then used in the PASTURE study a blood collection method suitable to stabilize RNA directly in the field. This method resulted in considerably higher RNA yields with lower variability between different centres. It therefore helped to minimize preanalytical errors and proved to yield useful amounts of RNA in the different countries involved.

Figure 2 shows that the agreement between the results of repeated measurements of the expression of the same genes in the same samples at various time points during long-time storage decreased with each additional measurement. Total RNA of these samples was extracted in 2002, thereafter stored in water at -80°C and only thawed up for the two subsequent analyses. Water as solvent for long-term storage of RNA may not be optimal, since unbuffered solutions enhance RNA degradation. This can be avoided with pH-buffered solvents such as *Tris*-buffer, storage of the RNA in precipitated form or reverse transcription of all RNA into cDNA being more stable during long-term storage. RNA degradation caused by duration of storage most likely occurs at random, resulting in a nondifferential error in statistical analyses. Weak associations between gene expression and environmental factors may therefore not be detectable anymore in older samples.

In this study, we assessed gene expression with two different rtPCR methods, the conventional single tube

assay and the high throughput low-density arrays. We found a high comparability between these two methods.

Biological data, such as hosts' immune responses to environmental exposures gathered in the context of epidemiological studies, are most important in order to better understand the biological mechanisms triggered by gene-environment interactions. However, collection of samples and initial work-up in epidemiological studies are made in the field or nonspecialized laboratories rather than in well-controlled lab-settings. Stringent quality control mechanisms are therefore mandatory when performing biological analyses in the context of epidemiological studies.

References

1. Ege MJ, Bieli C, Frei R, van Strien RT, Riedler J, Ublagger E et al. Prenatal farm exposure is related to the expression of receptors of the innate immunity and to atopic sensitization in school-age children. *J Allergy Clin Immunol* 2006;**117**:817–823.
2. Lauener RP, Birchler T, Adamski J, Braun-Fahrlander C, Bufe A, Herz U et al. Expression of CD14 and Toll-like receptor 2 in farmers' and non-farmers' children. *Lancet* 2002;**360**:465–466.
3. Alfven T, Braun-Fahrlander C, Brunekreef B, von Mutius E, Riedler J, Scheynius A et al. Allergic diseases and atopic sensitization in children related to farming and anthroposophic lifestyle – the PARSIFAL study. *Allergy* 2006;**61**:414–421.
4. von Mutius E, Schmid S. The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy* 2006;**61**:407–413.

Acknowledgments

PARSIFAL was supported by a research grant from the European Union QLRT 1999-01391 and by funding from the Swedish Foundation for Health Care Science and Allergy Research. The PASTURE study was funded within the European Union 'Quality of Life and Management of Living Resources' program, key action "Environment and Health", project no. QLRT-2001-00250. The Zurich group was supported by grant Nr. 32-100324 from the Swiss National Foundation and by the Kühne-Foundation. The present work was funded by GA²LEN (Global Allergy and Asthma European Network) and carried out as part of the GA²LEN workpackage 2.2. GA²LEN is a European Commission funded network of excellence dedicated to allergy and asthma.