Genomics 103 (2014) 31–39

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Characterization of the peripheral blood transcriptome in a repeated measures design using a panel of healthy individuals

Patrick De Boever [a,b,*], Britt Wens [a], Anyiawung Chiara Forcheh [c], Hans Reynders [d], Vera Nelen [e], Jos Kleinjans [f], Nicolas Van Larebeke [g], Geert Verbeke [c], Dirk Valkenborg [a,h], Greet Schoeters [a,i,j]

[a] Flemish Institute for Technological Research, Unit Environmental Risk and Health, Belgium
[b] Hasselt University, Centre for Environmental Sciences, Belgium
[c] Catholic University of Leuven, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Belgium
[d] Flemish Government, Environment, Nature and Energy Department, Belgium
[e] Provincial Institute for Hygiene, Belgium
[f] Maastricht University, Department of Toxicogenomics, The Netherlands
[g] Ghent University, Study Centre for Carcinogenesis and Primary Prevention of Cancer, Belgium
[h] Hasselt University, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Belgium
[i] University of Antwerp, Department of Biomedical Sciences, Belgium
[j] University of Southern Denmark, Department of Environmental Medicine, Denmark

ARTICLE INFO

Article history:
Received 9 January 2013
Accepted 29 November 2013
Available online 7 December 2013

Keywords:
Whole blood
Transcriptome
Gene expression variability
Immune system
Biomonitoring
Public health

ABSTRACT

A repeated measures microarray design with 22 healthy, non-smoking volunteers (aging 32 ± 5 years) was set up to study transcriptome profiles in whole blood samples. The results indicate that repeatable data can be obtained with high within-subject correlation. Probes that could discriminate between individuals are associated with immune and inflammatory functions. When investigating possible time trends in the microarray data, we have found no differential expression within a sampling period (within-season effect). Differential expression was observed between sampling seasons and the data suggest a weak response of genes related to immune system functioning. Finally, a high number of probes showed significant season-specific expression variability within subjects. Expression variability increased in springtime and there was an association of the probe list with immune system functioning. Our study suggests that the blood transcriptome of healthy individuals is reproducible over a time period of several months.

1. Introduction

Genome-wide interrogation of transcript abundance in peripheral blood cells is widely used to explore differences between diseased and healthy individuals, define disease stage, monitor a therapy and pharmacodynamic responses to a drug, or define prognosis [1]. Exposure to environmental stressors and toxicants may also disturb an individual's homeostasis, which may be reflected in gene expression changes in peripheral blood. This principle has been shown to hold true for study populations exposed to persistent organic pollutants [2], heavy metals [3], and complex mixtures such as cigarette smoke [4], diesel exhaust [5] and air pollution [6]. Furthermore, analysis of altered gene expression profiles may identify biological pathways that underlie the development of common diseases [7,8]. It is envisioned that microarray technology can become a useful tool for assessing the health impact of stressors in the context of occupational and environmental exposure to chemicals and physical particles [9].

Technical aspects related to gene expression analysis have been standardized and the contribution to gene expression variability has been characterized [10,11]. However, many factors contributing to biological variability are less well documented. For example, temporal changes in gene expression, within- and between-subject variability and lifestyle characteristics (such as diet and smoking status) can influence the outcome of gene expression studies [12].

In this study we report on the short- and long-term gene expression changes in whole blood samples of non-diseased adult individuals, as well as intra- and inter-individual variation in gene expression. Therefore, we have collected repeated measures over a period of six months using gene expression microarrays.

2. Results

2.1. Population characteristics

The population characteristics are summarized in Table 1. 45% of included participants were men. The population had a mean age of

* Corresponding author at: Flemish Institute for Technological Research (VITO), Unit Environmental Risk and Health, Industriezone Vlasmeer 7, 2400 Mol, Belgium. Fax: +32 14 58 05 23.
E-mail address: patrick.deboever@vito.be (P. De Boever).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.11.006

Table 1
Details on the study participants and the outcome of the questionnaires that were completed in conjunction with each blood sampling. Six blood samples were collected for each person over a period of 6 months, resulting in a total set of 132 samples.

Study participants (n = 22)		
Age (years)	32	±5
Body Mass Index	23.1	±2.5
Male (%)	45	
Outcome of the 132 questionnaires (%)		
Fasting at the moment of the examination	12	
Self-reported health issue (week before the blood drawing)	14	
Exposure to passive smoking during the past 24 h	11	
Alcohol consumption (one glass or less)during the past 24 h	77	
Coffee consumption (3 cups or less) during the past 24 h	71	
Physical exercise during the past 24 h	22	
Time spent in traffic on the day of the examination (min)		
In car	81	
In congested traffic	5	
Biking or walking in traffic	4	

32 ± 5 years and a mean body mass index of 23.1 ± 2.5. All participants had a university or college degree. Participants reported to be healthy and without recent history of clinical disease. Twenty two volunteers completed the study and this resulted in a total of 132 whole blood samples, with 6 samples per individual. In 12% of the samples, individuals reported to be fasting. However, the prandial status was not determined in the study. The questionnaires collected in parallel of the blood sampling revealed 14% of self-reported health problems (mostly cases of allergy) for which the individuals were taking standard medication (antihistamines and bronchodilators). In 15% of the surveys, individuals reported a health issue in the week prior to the blood drawing. Moderate coffee and alcohol consumption and average physical activity was reported 24 h before the blood drawing.

2.2. Environmental conditions

Average outdoor temperature was not significantly different between the two sampling campaigns and the ambient concentrations of chemical air pollutants were also comparable. The spring campaign was characterized by significantly higher pollen concentrations. The results are summarized in Table 2.

2.3. Microarray data generation

All blood samples were submitted to RNA extraction and the average RNA yield per sample was 8.44 ± 2.71 μg (n = 132). The quality of the samples was checked after globin removal and RIN-values of 8.93 ± 0.40 (n = 132) were obtained.

Cy3-labeling resulted in an average specific activity of 16.73 ± 2.75 pmol/μg (n = 132). The samples were hybridized successfully on 4X44K Agilent Whole Human Genome microarrays and all the raw data files passed the default quality control metrics that are implemented in Feature Extraction. The median ± standard error of the coefficient of variation of the within-array repeated probes calculated for the 132 samples was 4.4 ± 1.1 %. This indicates a within-slide technical variability of less than 5%. Data preprocessing resulted in a final set of 31,293 probes that was used for downstream analysis.

Table 2
Outdoor temperature and concentrations of ambient air pollutants that were measured in the 2 sampling periods (season 1 and season 2) using fixed measuring stations in the neighborhood of the study site. Averages were calculated for both 3-week sampling periods. Range of the measurements is also shown.

	Season 1		Season 2	
Temperature (°C)	9.9 ± 2.9	2.4–17.4	10.0 ± 4.9	2.6–28.1
Pollutants (μg/m^3)				
Particulate matter < 10 μm	18	3–62	20	3–81
NO	2	1–49	1	1–21
NO_2	18	2–57	17	5–47
SO_2	2	1–24	1	1–19
O_3	39	1–76	57	1–131
Pollen (counts/m^3)				
Trees	Not detected		127 ± 135	4–774
Plants	Not detected		8 ± 2	1–22

2.4. Descriptive statistics

Pairwise Spearman correlations were high and all values exceeded the threshold of 0.94. Fig. 1 summarizes the results in a heat map. The microarray samples in Fig. 1 are sorted according to study participant in order to group the 6 samples of each individual. The Spearman correlations along the diagonal are the highest and this indicates that the within-subject sample correlations were always higher than the between-subject correlations.

Subsequently, the between/within sum of squares (BW) ratio across the 22 individuals was determined for each microarray probe. The analysis allowed identifying the probes that separate maximally between the individuals. When ranking the BW ratios, a steep decrease in the size of the BW ratios was observed for the first 50 probes and ratios leveled off at 250 probes (data not shown). These 250 probes and their annotation are given in a supplementary Excel file (Supplementary file 1). The latter probe set was also used for hierarchical cluster analysis on the gene expression signals. The results revealed a perfect clustering according to individual (Fig. 2). This confirms the statement about strong within-subject correlations that was based on the Spearman correlations. The discrimination between males and females could be attributed to the expression signals of 8 probes. These probes could be mapped to 7 genes located on the Y chromosome (KDM5D, RPS4Y1, RPS4Y2, DDX3Y, EIF1AY, TMSB4Y and TTTY14).

Gene Ontology (GO) analysis of the 250 probes did not yield significantly enriched biological terms at a BH-FDR of 0.1. The significantly enriched canonical pathways (BH-FDR < 0.1) were autoimmune thyroid disease signaling, glutathione-mediated detoxification and altered T cell and B cell signaling in rheumatoid arthritis. The genes that were contributing to this enrichment were mainly related to antigen processing and presentation (HLA-DQA1, HLA-DQB1, HLA-DRB1, HLA-DRB5 and MAPK8), inflammatory response (CCL3, CCL3L1, IL23A, NCR3, PVRL2, PTPN22, FKBP1A and UTS2) and glutathione metabolism (GSTM1, GSTM3, GSTM4 and GSTT1).

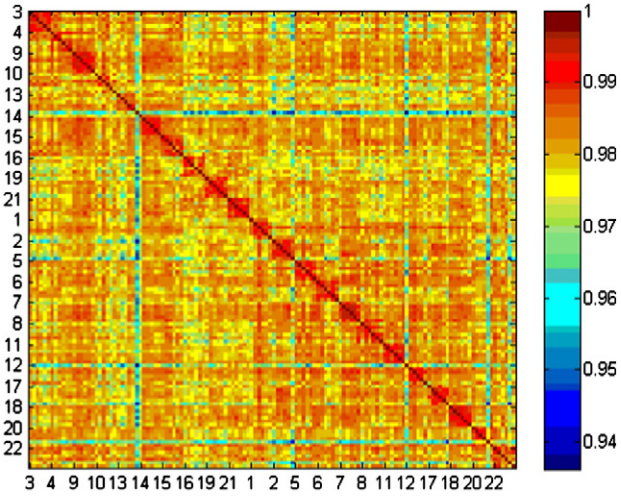

Fig. 1. Heat map of pairwise Spearman correlations. The sample identifiers refer to the 22 study participants. The first 10 numbers (3 to 21) refer to males, whereas the following numbers (1 to 22) refer to females. The samples from one individual are ordered according to the time of sampling. The colors indicate the strength of the correlation, with dark red indicating perfect correlation.