The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool

CHOONG-CHIN LIEW, JUN MA, HONG-CHANG TANG, RUN ZHENG, and ADAM A. DEMPSEY

TORONTO, ONTARIO, CANADA AND BOSTON, MASSACHUSETTS

In our genome-wide survey of gene expression in human peripheral blood cells using both an expressed sequence tag (EST) and a microarray hybridization approach, we identified the expression of a large proportion (approximately 80%) of the genes encoded in the human genome. Comparison of the peripheral blood transcriptome with genes expressed in nine different human tissue types revealed that expression of over 80% was shared with any given tissue. We also sought to determine whether those gene transcripts undetected by these methods were also expressed in peripheral blood cells. Using reverse-transcriptase-polymerase chain reaction, we detected additional tissue-specific gene transcripts including beta-myosin heavy chain (heart specific) and insulin (specific to pancreatic islet beta cells), in circulating blood cells. Arguably, the detection of low levels of tissue-specific transcripts could be considered products of "illegitimate" transcription; however, our study also demonstrates that environmental conditions affect the transcriptional regulation of insulin in the peripheral blood. We thus hypothesize that blood cells can act as sentinels of disease and that we could capitalize on this property of blood for the diagnosis/prognosis of disease (the "Sentinel Principle"). Peripheral blood is an ideal surrogate tissue as it is readily obtainable, provides a large biosensor pool in the form of gene transcripts, and response to changes in the macro- and micro-environments is detectable as alterations in the levels of these gene transcripts. (J Lab Clin Med 2006;147:126–132)

Abbreviations: EST = expressed sequence tag; IRB = Independent Review Board; RT-PCR = reverse-transcriptase-polymerase chain reaction

The human body is nourished by a dynamic circulatory system; the cellular components of which have a relatively rapid turnover rate.[1] Blood is classified as a fluid connective tissue, which can be defined as cells suspended in a fluid matrix functioning to connect the entire biological system at the physiological level. Blood cells also constitute the first line of the immune defense system, using an arsenal of neutrophils, eosinophils, basophils, B cells, T cells, and monocytes to defend against foreign assault and injury. Thus, the blood pervades the entire body, is in a constant state of renewal, and provides a protective barrier between the external and internal environments.

The continuous interactions between blood cells and the entire body gives rise to the possibility that subtle changes occurring in association with injury or disease, within the cells and tissues of the body, may trigger specific changes in gene expression in blood cells reflective of the initiating stimulus. We thus set out to use

From ChondroGene, Inc., Toronto, Ontario, Canada; and Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.

Supported by ChondroGene Inc.

Submitted for publication April 15, 2005; revision submitted October 25, 2005; accepted for publication October 27, 2005.

Reprint requests: Choong-Chin Liew, 800 Petrolia Road, Unit 15, Toronto, Ontario, Canada, M3J 3K4; e-mail: cliew@chondrogene.com.

0022-2143/$ – see front matter

© 2006 Mosby, Inc. All rights reserved.

doi:10.1016/j.lab.2005.10.005

the "Sentinel Principle" to demonstrate that blood could serve diagnostic/prognostic purposes through profiling gene expression in blood cells. We have now demonstrated that monitoring gene expression in blood results in gene expression signatures reflective of over 35 different conditions in human subjects (see the U.S. patent applications: No. 60/115,125, No. 09/477,148, and No. 10/268,730; No. 10/601,518, No.10/802,875, and PCT application No. PCT/US04/020836).[2] Recent studies by others also demonstrate the utility of peripheral blood as a source of significant information showing that gene expression profiles of circulating blood cells were distinctive between persons,[3] and alterations in the expression profiles of blood cells were characteristic in a wide range of diseases, including juvenile arthritis,[4] hypertension,[5–7] cancer,[8,9] chronic fatigue disease[10] and neuronal injuries,[11,12] lupus,[13,14] transplantation,[15] and under various environmental pressures, such as exercise,[16] hexachlorobenzene exposure,[17] arsenic exposure,[18] and smoking.[19] A recent study by our group also demonstrated that psychiatric disorders, specifically schizophrenia and bipolar disorder, could be distinguished through specific peripheral blood gene expression profiles.[20] This rapidly growing body of evidence demonstrates the potential of using peripheral blood as a surrogate tissue for traditional tissue specimens for prognosis and diagnosis. Clearly blood provides significant advantages for this purpose, being readily available in large quantities with minimally invasive techniques.

An ideal surrogate tissue used in gene profiling analysis will be one that expresses many genes, most of which are responsive to physiological or environmental alterations. In this study, we found that genes previously believed to be restricted to non-blood tissues were in fact expressed in peripheral blood cells. Our results also suggest that the expression level of many transcripts in blood cells are responsive to, and thus indicative of, their micro-environment. These discoveries, in addition to the physical characteristics of circulating blood cells, prompted us to hypothesize that circulating blood cells can function as "sentinels" that respond to changes in the macro- or micro-environment in organs, and that blood is an ideal surrogate tissue for diagnostics.

MATERIALS AND METHODS

Isolating RNA from human circulating blood cells. Approximately 10 mL of peripheral blood was collected from each human subject. Research was carried out as required by the principles of the Declaration of Helsinki. All sample collection was approved by the collecting Institute's IRB, and written informed consents were provided in accordance with the requirements of the IRB. All samples were immediately stored on ice until RNA isolation was initiated. All RNA isolation was performed within 4 hours after blood collection. The collected blood was mixed with three volumes of hemolysis buffer (EDTA 0.6 g/L, $KHCO_3$ 1.0 g/L, NH_4Cl 8.2 g/L, pH = 7.4) to lyse red blood cells. The samples were spun at 800 rcf for 10 minutes at 4°C. The resulting pellet was washed with hemolysis buffer several times and treated with TRIZOL reagent (Invitrogen, Carlsbad, Calif) to isolate total RNA following the manufacturer's instructions. Purity and integrity of the RNA were assessed by absorbance at $UV_{260/280}$ and agarose gel electrophoresis. The quality of the RNA isolated for microarray-based expression profiling was further assessed on an Agilent Bioanalyzer 2100 using RNA 6000 Nano Chips (Agilent Technologies, Palo Alto, Calif).

Cataloguing the blood transcriptome using ESTs. The procedures of cDNA library construction and EST generation were described previously.[21] Briefly, RNAs from a pool of five adult peripheral blood and one umbilical cord blood sample were reverse transcribed into double-stranded cDNA followed by end-modification and ligation into a lambda ZAP Express vector; the assembled clones were then packaged into lambda phage, resulting in two cDNA libraries. Phage plaques were randomly picked, the cDNA insert amplified by PCR, and the product sequenced from the 5′-end. A non-redundant list of identified gene transcripts was constructed by performing sequence-based similarity clustering using the TIGR Assembler,[22] in which ESTs with an overlap of a minimum of >95% identity over 40 nt were considered to represent the same transcript and grouped together to form consensus sequences. The resultant EST cluster consensus sequences and unclustered ESTs were annotated by searching the Genbank data repositories with the BLAST algorithm (http://www.ncbi.nlm.nih.gov). Those EST clusters and unclustered ESTs that matched the same gene transcript, based on LocusLink Ids and Genbank Accession numbers, were considered redundant, and a final non-redundant list of gene transcripts was compiled.

Analysis of ESTs. The resulting 44,229 blood-derived ESTs were randomly partitioned into 44 groups, 1000 ESTs per group. A growth curve was plotted using *Sigmaplot* (Sysstat, Richmond, Calif) to estimate the total number of expressed genes in the blood transcriptome. The x-axis on the growth curve represents the total number of ESTs assessed for unique genes in increasing steps of 1000 ESTs (starting at 1000 to a total of 44,229, or 43 steps); the y-axis represents the number of unique genes identified within each group of ESTs. A regression model, $y = ax/(b + x)$, was chosen to fit the curve using *Sigmaplot*. The coefficient "a" in the regression model represents the number of genes contained in the cDNA library studied, which is based on the mathematical concept that as x approaches infinity, $y = a$. The coefficients "a" and "b" were calculated using *Sigmaplot*. The 44 groups of ESTs were randomly shuffled 20 times, resulting in 20 growth curves and 20 sets of coefficients "a" and "b." The average and standard deviation of the coefficients were calculated.

Affymetrix GeneChip profiling and data analysis. Total RNA was extracted from 248 persons. Five micrograms of each total RNA sample was used for hybridization on an

Affymetrix U133Plus2 GeneChip (Affymetrix, Santa Clara, Calif) following the manufacturer's instructions. Genes flagged as "present" or "marginal," as determined by the GeneChip Operating System (GCOS) software (Affymetrix), in at least one hybridization were considered expressed genes. LocusLink IDs were used as gene identifiers to generate a non-redundant list of expressed genes. The gene count was subsequently calculated from the non-redundant list of genes.

Comparing the tissue distribution of the circulating blood cell transcriptome. Genes expressed in nine different human tissue types, including brain, colon, heart, kidney, liver, lung, prostate, spleen, and stomach, were retrieved from the UniGene database (Build 179) at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). UniGene IDs were used to identify corresponding LocusLink IDs. The 15,193 genes found expressed in circulating blood cells by microarray hybridization were compared with those genes identified as expressed in one of the nine tissues using LocusLink IDs.

Detecting tissue-specific transcripts in blood cells by RT-PCR. The Titan One Tube RT-PCR System (Roche Diagnostics, Indianapolis, Ind) was used for all RT-PCRs. The RT-PCR mixture contains 1 μL of total RNA from blood samples (0.1 μg/μL), 4 μL dNTP mixture (2.5mM each), 10 μL 5X RT-PCR buffer, 2.5 μL DTT-solution (100 mM), 1 μL of each primer (20 μM): (1) beta-myosin heavy chain (β-MHC) (NM_000257): Forward ′-*GCTGGAACGTAGAGACTCCCTGCT*-3′ [spans exons 21/22], Reverse 5′-*GGATCCTTCCAGATCATCCACTTG*-3′ [spans exons 24/25]; (2) insulin (INS) (NM_000207): Forward 5′-*GCCCTCTGGGGACCTGAC*-3′ [exon 2], Reverse 5′-*ACCTGCCCCACCTGCAGG*-3′ [spans exons 2/3], and 1 μL enzyme mixture.

RT-PCR was carried out at 55°C for 30 minutes for reverse transcription, followed by 30 cycles at 94°C for 30 seconds, at the appropriate annealing temperature for 20 seconds, and 72°C for 1 minute. A negative PCR control was done using the Expand High Fidelity PCR system (Roche Diagnostics, Indianapolis, Ind) to test whether any amplified product came from genomic DNA contamination in the total RNA preparation. RT-PCR products were analyzed with 1% agarose electrophoresis and purified using GeneClean (Bio101, Vista, Calif). Purified RT-PCR products were sequenced from both ends using BigDye terminator thermocycling chemistry (Applied Biosystems, Foster City, Calif) on an ABI 3770 automated sequencer (Applied Biosystems), and sequences were annotated by searching the Genbank databases at the NCBI using the BLAST program.

To assay for the gene expression of insulin, a drop of blood was collected from each subject by a finger prick. The drop of blood underwent red blood cell hemolysis as described above, and the resulting pellet was re-suspended in 10 μL water, 1 μL of which was used directly as the RNA source for the RT-PCR assays. Samples were collected at two different physiologic states (ie, fasting and non-fasting) from the same four persons for use in this assay.

The amplified insulin gene was quantified using the Quantity One v4.3.1 gel documentation system software (BioRad) and the digitally captured agarose gel image. Briefly, an equal-sized area encompassing each gel band was highlighted on the digital image and the intensity within the selected area was measured. Local background noise was determined by averaging three selected sections of the scanned image, and this value was subtracted from the experimental signal. A *t*-test was used to determine statistical significance between the intensity readings for the fasting (n = 4) and the non-fasting groups (n = 4).

RESULTS

Estimating the number of genes expressed in the blood transcriptome using expressed sequence tags. We generated a total of 44,229 blood-expressed sequence tags from two blood cDNA libraries: an adult peripheral leukocyte cDNA library (15,161 ESTs) and an umbilical cord blood cDNA library (29,068 ESTs). Overall, 10,013 unique transcripts were determined through sequence similarity-based clustering with references to the GenBank, UniGene, and LocusLink databases. The regression model, $y = ax/(b + x)$, was chosen to fit the curve because it provided the highest correlation coefficients and thus provided the best fit (Fig 1). The order of the stepwise addition of the 44 EST groups was randomized 20 times. The estimate of coefficients "a" and "b" based on each growth curve was calculated. The average of coefficient "a" was 16,409 with a standard deviation of 171. Thus, based on this model, we estimate that ~16,400 genes are expressed in human blood cells. Considering there are from 20,000 to 25,000 protein-coding genes in the human genome,[23] this indicates that approximately 66% (16,400/25,000)–82% (16,400/20,000) of the genes encoded in the human genome are expressed in human blood cells.

Estimating the number of genes expressed in the blood transcriptome using Affymetrix GeneChip profiling. For this analysis, only genes with corresponding LocusLink IDs were used to estimate the count of leukocyte-expressed genes. We investigated the approximate 19,924 unique genes with LocusLink IDs present on the *Affymetrix U133Plus2 GeneChips*. A total of 39,204 probe sets were found in at least 1 of the 248 hybridizations, representing 16,304 unique genes with LocusLink IDs. Assuming the *Affymetrix U133Plus2 GeneChip* represents an unbiased sample of the entire genome, we can estimate that approximately 81.8% (16,304/19,924) of the genes encoded in the human genome are expressed in human peripheral leukocytes. Taking into consideration the genes that have not been assigned LocusLink IDs, we can provide an estimate of approximately 16,366 (20,000 × 81.8%)–20,450 (25,000 × 81.8%) genes expressed in human blood cells.

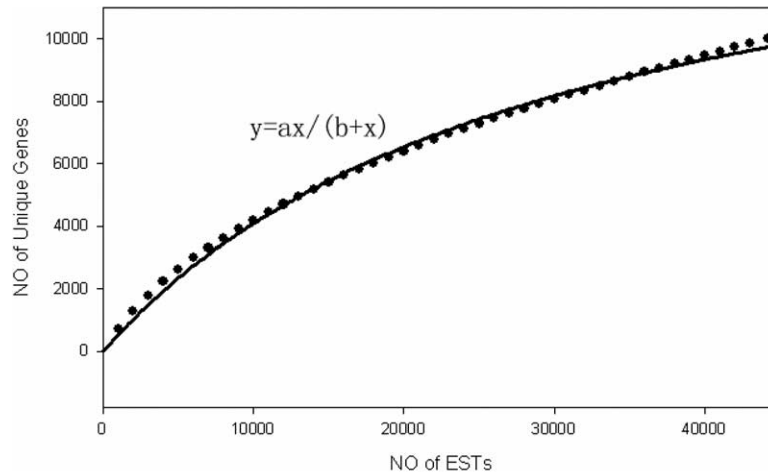

Fig 1. The EST-Gene growth curve. The x-axis represents the number of ESTs in the group used for unique gene assessment, and the y-axis represents the mean number (n = 20 reiterations) of unique genes found within the corresponding EST group. The regression model $y = ax/(b + x)$ provided the best fit to the curve. The coefficient a in the regression model represents the number of genes contained in the cDNA library studied, and b is a coefficient related to the complexity of the cDNA library. The total number of expressed genes was estimated based on the mathematical concept that as x approaches infinity, $y = a$.

Table I. Genes expressed in peripheral blood cells shared with one of nine human tissues

Tissues	Brain	Colon	Heart	Kidney	Liver	Lung	Prostate	Spleen	Stomach
Number of genes/expressed	13961	13767	12440	13428	13840	15202	11706	13224	10898
Number of co-expressed genes in blood	11428	11360	10472	11166	11490	12301	9955	10892	9408
Percentage of co-expressed genes in blood	81.9%	82.5%	84.2%	83.2%	83.0%	80.9%	83.9%	85.0%	86.3%

Comparing genes expressed in circulating blood cells and genes expressed in nine human tissue types. A non-redundant list of genes expressed in each of the nine human tissue types was retrieved by searching the UniGene database: Brain: 13,961; Colon: 13,767; Heart: 12,440; Kidney: 13,429; Liver: 13,840; Lung: 15,202; Prostate: 11,706; Spleen: 13,224; and Stomach: 10,898. The comparison between blood expressed genes (15,193 genes) and genes found in one of the nine tissue types revealed that a large proportion of genes expressed in the nine tissues are also expressed in blood cells (Table I). Approximately 80% of genes expressed in any of the nine tissues were also found expressed in blood cells.

Detecting tissue-specific transcripts in blood cells by RT-PCR. We examined the expression of two "tissue-specific" genes in peripheral blood cells: β-MHC and INS. These genes are normally associated with, and primarily expressed in, the heart and pancreas, respectively. Because their expression in the blood is likely occurring at very low levels, we could not detect their expression in our blood cell EST database or in the microarray hybridization experiment. However, in RT-PCR, we successfully detected the transcripts of these two genes in blood cells (Fig 2). The primers were designed to span exon–exon junctions, and as such, no genomic DNA should be amplified. PCR amplification of the samples without RT did not result in any amplification products, which indicates no genomic DNA contamination. The PCR products were subjected to automated DNA sequencing and confirmed that the products were generated from the targeted genes.

To determine whether environmental changes influenced expression of these low expressed genes, and to remove the possibility that detection was the result of leaky "illegitimate" transcription, we assayed the expression of the insulin gene in the blood from fasting and non-fasting subjects. A significant difference in insulin gene expression in the peripheral blood samples was observed between the fasting and the non-fasting

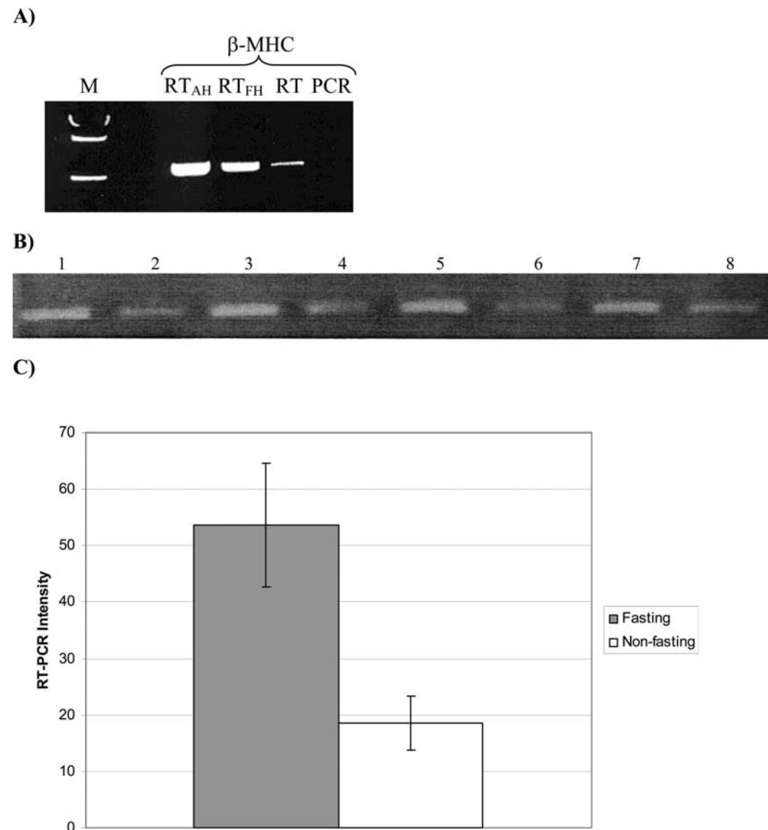

Fig 2. RT-PCR detection of "tissue-specific" gene expression in the peripheral blood. (**A**) RT-PCR was used to assay the expression of β-myosin heavy chain (β-MHC) in the peripheral blood from a single sample. The expression of β-MHC was detected in the peripheral blood (RT), whereas the negative control (PCR) did not result in amplification, which indicates no genomic DNA contamination in the samples used in the assay. RT-PCR products for beta-MHC from both human adult and fetal heart tissue ($RT_{AH}$and RT_{FH}, respectively) are also presented for reference. M indicates a molecular weight marker. (**B**) RT-PCR assay of insulin gene expression in a drop of peripheral blood. Lanes 1, 3, 5, and 7 represent samples from subjects that underwent overnight "fasting," and lanes 2, 4, 6, and 8 represent corresponding samples from subjects that did not undergo fasting or "non-fasting" samples. (**C**) Quantification of insulin RT-PCR in four subjects at fasting and non-fasting states ($P = 0.0033$, n = 4, one-sample *t*-test). The error bars indicate standard deviation from the mean. The y-axis represents signal intensity obtained from the scanned gel image in arbitrary units.

subjects ($P = 0.0033$) (Fig 2, *B* and *C*). Figures 2, *A* and *B* were modified from U.S. patent application No. 60/115,125 with permission.

DISCUSSION

Using two independent methods, we estimated the number of unique genes expressed in the blood transcriptome to be 16,000 to 20,000. The estimate derived from microarray hybridization (81.1%) is similar to the high end of the estimate derived from ESTs (82%). As microarray is more sensitive than the EST approach, we suggest peripheral blood cells express approximately 80% of the genes encoded by the human genome. Comparison of the genes expressed in the blood against a range of different tissues revealed that over 80% of the genes expressed in other tissues overlapped with the blood. Although many of these overlapping genes may be considered "housekeeping" genes, the detection of such a large number of genes shared between blood cells and other tissues cannot be explained by housekeeping functions.

Gene transcription is considered a process under strict control; only genes required by cells or tissues are expressed. However, it has been reported that tissue-specific genes may be expressed in a non-tissue-specific manner.[24–27] Ectopic expression describes "illegitimate" transcripts as products of basal transcription due to the presence of ubiquitous transcription factors and/or a net balance of negative and positive regulatory factors. It has also been suggested that in higher eukaryotes, transcription initiation occurs at a low frequency and the process is regulated in a probabilistic

manner,[28] providing the opportunity for any gene to be expressed, although at varying levels. In our study, we detected the expression of two "tissue-specific" genes, insulin and β-MHC, in peripheral blood. These findings suggest that although the expression level of genes may vary among different tissue or cell types, most genes may be expressed in the blood at a detectable level using conventional methods.

An ideal surrogate tissue used in gene profiling analysis will be one that expresses many genes, many of which are responsive to physiological or environmental alterations. Proving most genes are expressed in blood cells has provided support for the first criteria of being a successful surrogate tissue. Genes, in this context, can be considered "bio-sensors." Many genes provide the potential of being able to detect various signals/stimuli. However, the overall sensitivity of these "bio-sensors" mainly relies on their capability of specific responses to various signals/stimuli. Recent blood gene expression studies have shown that the expression profiles of circulating blood cells do contain specific expression signatures in response to various physiological, pathological, and environmental changes.[4–20] In this study, using insulin as an example, we observed that insulin gene expression in blood seems to be influenced by environmental conditions, specifically fasting and nonfasting states of normal subjects.

These findings suggest that circulating blood cells have unique characteristics that make them a potential new tool for diagnostics: (1) a large proportion of the genes encoded in the human genome have detectable levels of transcripts in circulating blood cells; (2) circulating blood cells come into contact with every cell in the human body and provide an active defense against insult and injury; and (3) macro- and micro-environment changes affect gene expression in blood cells. Therefore, circulating blood cells may provide information as to the health or disease of any particular tissue by the change of gene expression pattern of their transcriptome.

In summary, we suggest that peripheral blood cells express a large proportion of the genes in the human genome, which can respond to changes occurring in the macro- and micro-environment in the body. The continuous interactions between blood cells and the entire body, combined with the fast turnover rate of blood cells, gives rise to the possibility that subtle changes occurring in association with injury or disease within the cells and tissues of the body may trigger specific changes in gene expression at a micro-level within the blood cells. These changes can then be capitalized on as biosensors for diagnostic purposes.

The authors would like to thank K. Wayne Marshall for his productive input and constructive criticism of the manuscript, Eva Cuckerman and Jim Loukides for their technical assistance, and Isolde Prince for her editorial assistance.

REFERENCES

1. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. Blood 1993;81:2844–53.
2. Liew CC. Expressed genome molecular signatures of heart failure. Clin Chem Lab Med. In press.
3. Radich JP, Mao M, Stepaniants S, Biery M, Castle J, Ward T, et al. Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics 2004;83:980–8.
4. Barnes MG, Aronow BJ, Luyrink LK, Moroldo MB, Pavlidis P, Passo MH, et al. Gene expression in juvenile arthritis and spondyloarthropathy: pro-angiogenic ELR+ chemokine genes relate to course of arthritis. Rheumatology (Oxford) 2004;43:973–9.
5. Bull TM, Coldren CD, Moore M, Sotto-Santiago SM, Pham DV, Nana-Sinkam SP, et al. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. Am J Respir Crit Care Med 2004;170:911–9.
6. Chon H, Gaillard CA, van der Meijden BB, Dijstelbloem HM, Kraaijenhagen RJ, van Leenen D, et al. Broadly altered gene expression in blood leukocytes in essential hypertension is absent during treatment. Hypertension 2004;43:947–51.
7. Okuda T, Sumiya T, Mizutani K, Tago N, Miyata T, Tanabe T, et al. Analyses of differential gene expression in genetic hypertensive rats by microarray. Hypertens Res 2002;25:249–55.
8. DePrimo SE, Wong LM, Khatry DB, Nicholas SL, Manning WC, Smolich BD, et al. Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. BMC Cancer 2003;3:3.
9. De Vos J, Thykjaer T, Tarte K, Ensslen M, Raynaud P, Requirand G, et al. Comparison of gene expression profiling between malignant and normal plasma cells with oligonucleotide arrays. Oncogene 2002;21:6848–57.
10. Whistler T, Unger ER, Nisenbaum R, Vernon SD. Integration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome. J Transl Med 2003;1:10.
11. Tang Y, Lu A, Aronow BJ, Sharp FR. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. Ann Neurol 2001;50:699–707.
12. Tang Y, Nee AC, Lu A, Ran R, Sharp FR. Blood genomic expression profile for neuronal injury. J Cereb Blood Flow Metab 2003;23:310–9.
13. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, et al. Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. Clin Immunol 2002;102:283–90.
14. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 2003;197:711–23.
15. Zhang HQ, Lu H, Enosawa S, Takahara S, Sakamoto K, Nakajima T, et al. Microarray analysis of gene expression in peripheral blood mononuclear cells derived from long-surviving renal recipients. Transplantation Proc 2002;34:1757–9.
16. Connolly PH, Caiozzo VJ, Zaldivar F, Nemet D, Larson J, Hung SP, et al. Effects of exercise on gene expression in human peripheral blood mononuclear cells. J Appl Physiol 2004;97:1461–9.
17. Ezendam J, Staedtler F, Pennings J, Vandebriel RJ, Pieters R, Boffetta P, et al. Toxicogenomics of subchronic hexachloroben-

zene exposure in Brown Norway rats. Environ Health Perspect 2004;112:782–91.

18. Wu MM, Chiou HY, Ho IC, Chen CJ, Lee TC. Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. Environ Health Perspect 2003; 111:1429–38.
19. Ryder MI, Hyun W, Loomer P, Haqq C. Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. Oral Microbiol Immunol 2004;19:39–49.
20. Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet 2005;133:1–5.
21. Hwang DM, Dempsey A, Wang RX, Rezvani M, Barrans JD, Dai KS, et al. A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. Circulation 1997;96:4146–203.
22. Sutton G, White O, Adams M, Kerlavage A. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci Technol 1995;1:9–19.
23. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431:931–45.
24. Humphries S, Windass J, Williamson R. Mouse globin gene expression in erythroid and non-erythoid tissues. Cell 1976;7:267–77.
25. Chelly J, Kaplan JC, Maire P, Gantron S, Kahn A. Transcription of the dystrophin gene in human muscle and non-muscle tissues. Nature 1988;333:858–60.
26. Chelly J, Concordet JP, Kaplan JC, Kahn A. Illegitimate transcription: transcription of any gene in any cell type. Proc Natl Acad Sci USA 1989;86:2617–21.
27. Kimoto Y. A single human cell expresses all messenger ribonucleic acids: the arrow of time in a cell. Mol Gen Genet 1998; 258:233–9.
28. Hume DA. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 2000;96:2323–8.