The effects of globin on microarray-based gene expression analysis of mouse blood

Mary E. Winn · Matthew A. Zapala ·
Iiris Hovatta · Victoria B. Risbrough ·
Elizabeth Lillie · Nicholas J. Schork

Abstract The use of mouse blood as a model for human blood is often considered in the development of clinically relevant, gene expression-based disease biomarkers. However, the ability to derive biologically meaningful insights from microarray-based gene expression patterns in mouse whole blood, as in human whole blood, is hindered by high levels of globin mRNA. In order to characterize the effects of globin reduction on gene expression of peripheral mouse blood, we performed gene set enrichment analysis on genes identified as expressed in blood via microarray-based genome-wide transcriptome analysis. Depletion of globin mRNA enhanced the quality of microarray data as shown by improved gene expression detection and increased sensitivity. Compared to genes expressed in whole blood, genes detected as expressed in blood following globin reduction were enriched for low abundance transcripts implicated in many biological pathways, including development, g-protein signaling, and immune response. Broadly, globin reduction resulted in improved detection of expressed genes that serve as molecular binding proteins and enzymes in cellular metabolism, intracellular transport/localization, transcription, and translation, as well as genes that potentially could act as biomarkers for diseases such as schizophrenia. These significantly enriched pathways overlap considerably with those identified in globin-reduced human blood suggesting that globin-reduced mouse blood gene expression studies may be useful for identifying genes relevant to human disease. Overall, the results of this investigation provide a better understanding of the impact of reducing globin transcripts in mouse blood and highlight the

M. E. Winn
Graduate Program in Biomedical Sciences, Department of
Medicine, University of California at San Diego,
La Jolla, CA 92093, USA
e-mail: mewinn@ucsd.edu

M. E. Winn · M. A. Zapala · E. Lillie · N. J. Schork (✉)
Scripps Genomic Medicine, Department of Molecular &
Experimental Medicine, and The Scripps Translational Science
Institute, The Scripps Research Institute,
La Jolla, CA 92037, USA
e-mail: nschork@scripps.edu

M. A. Zapala
Department of Radiology, University of California at San Diego,
La Jolla, CA 92093, USA

I. Hovatta
Research Program of Molecular Neurology and Department
of Medical Genetics, University of Helsinki, Helsinki, Finland

I. Hovatta
Department of Mental Health and Substance Abuse Services,
National Institute for Health and Welfare, Helsinki, Finland

V. B. Risbrough
Department of Psychiatry, University of California at San Diego,
La Jolla, CA 92093, USA

V. B. Risbrough
Center of Excellence for Stress and Mental Health,
Veteran's Affairs, La Jolla, CA 92093, USA

potential of microarray-based, globin-reduced, mouse blood gene expression studies in biomarker development.

Introduction

Microarray-based gene expression profiling is a frequently used and powerful tool in the search for molecular "fingerprints" of specific diseases and thus the development of clinically relevant biomarkers for those diseases. For example, seminal investigations have considered the use of large-scale gene expression analyses to classify disease states (Alizadeh et al. 2000; Dyrskjot et al. 2003; Golub et al. 1999), develop diagnostic and prognostic gene profiles (Mirnics et al. 2000; van't Veer et al. 2002; Welsh et al. 2001), and characterize transcriptomic fingerprints of drug efficacy and toxicity (Gunther et al. 2003). While many of these landmark studies relied heavily on primary tissue samples, recent investigations have focused on peripheral blood, a more accessible tissue (Chao et al. 2008; Coppola et al. 2008; Glatt et al. 2005; Le-Niculescu et al. 2009; Miller et al. 2007; Wang et al. 2005). Aside from being relatively easy to obtain, many of the physiological characteristics of blood cells suggest that peripheral blood gene expression is a reasonable surrogate for specific primary tissue gene expression and hence can be used in the development of clinically meaningful expression-based biomarkers for diseases whose molecular "lesions" are associated with particular nonblood primary tissues (Fan and Hegde 2005; Liew et al. 2006; Mohr and Liew 2007).

One limitation of current studies designed to correlate blood gene expression patterns to primary tissue gene expression patterns is that blood and tissue samples are often not collected from the same set of individuals (Glatt et al. 2005; Liew et al. 2006; Solmi et al. 2006; Sullivan et al. 2006). The use of independent sources of blood and nonblood tissue gene expression information limits interpretability and generalizability of relevant studies and calls into question any putative blood-based gene expression biomarker panel for the diagnosis and treatment of a disease whose primary lesions are not in blood. Nonetheless, it is often too difficult or even impossible to obtain blood and primary tissue samples from the same living individuals. The mouse provides one possible solution to this dilemma.

Unlike human subjects, blood and primary tissue samples can be easily collected from the same living mouse under highly controlled conditions. Intra- and interindividual variation introduced by gender, age, time of day, genetic variation, and environment can be reduced in such studies (Cobb et al. 2005; Leonardson et al. 2010; Radich et al. 2004; Whitney et al. 2003), while clinically acceptable and highly standardized protocols for blood collection, RNA isolation, and globin reduction can be employed to lessen technical, assay-induced variation (Cobb et al. 2005; Debey et al. 2004, 2006). In addition, the fact that many different isogenic strains of mice exist suggests that given the clone-like nature of the mice within such strains, it is possible to sample expression patterns in different tissues from different individuals within particular strains and test the expression patterns for consistencies as though they were obtained from the same individuals.

It is known that high levels of globin transcripts in the blood can confound the accurate assessment of the expression levels of genes in the blood (Wu et al. 2007), as globin mRNA represents up to 70% of the total expressed transcripts and consequently limits the ability to accurately detect genes expressed at low levels in the blood. Thus, globin reduction is often considered a necessary step in the evaluation of whole-blood gene expression profiles via microarrays. GLOBINclear^{TM}, a commercially available (Applied Biosystems/Ambion, Austin, TX) globin reduction protocol, has been shown to improve gene expression detection sensitivity, remove up to 95% of α- and β-globin mRNA, and diminish globin-specific expression patterns in human whole-blood samples (Field et al. 2007; Liu et al. 2006; Whitley et al. 2005; Wright et al. 2008). GLOBINclear^{TM} is also advantageous for mouse studies in that it has been specifically developed for the mouse (Whitley et al. 2007).

In this article we describe a study designed to (1) evaluate and characterize the effects of globin reduction on whole-blood gene expression in different mouse strains, (2) determine which pathways are enriched for genes that appear to be heavily influenced by the confounding or masking effects of globin in the blood, and (3) assess the utility of globin-reduced mouse whole blood in the identification of potential biomarkers of human disease.

Materials and methods

Sample collection

All animal procedures were performed according to protocols approved by the University of California San Diego Institutional Animal Care and Use Committee. Seven-week-old male mice were purchased from The Jackson Laboratory (Bar Harbor, ME) (129S1/SvImJ, A/J, C57BL/6 J, C3H/HeJ, DBA/2 J, and FVB/NJ) and individually housed for 1 week prior to blood collection. All mice were anesthetized using isoflurane in a fume hood and whole blood was collected via cardiac puncture. The blood was transferred to an EDTA tube and then TRIzol® LS reagent (Invitrogen, Carlsbad, CA) was immediately added (3:1

TRIzol:blood) creating a solution in a 15-ml tube that was stored at -80°C for no more than 2 weeks.

Processing and globin reduction

The extraction of total RNA from the blood was performed using the TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Total RNA quantity and quality was assessed by spectrophotometer and the RNA 6000 Nano Chip kit (Agilent, Santa Clara, CA). α- and β-globin mRNA were reduced from a portion of the total RNA samples using the GLOBINclear^{TM} Mouse/Rat kit (Ambion) according to the manufacturer's instructions, with the recommended start quantity of 10 μg of total RNA.

Sample amplification and microarray analysis

Gene expression analysis was performed on all whole-blood RNA and globin-depleted samples using Mouse 430 2.0 arrays (Affymetrix, Santa Clara, CA) containing 45,101 probe sets. Sample labeling, hybridization, and scanning were performed as previously described (Zapala et al. 2005). Three biological replicate samples from independent mice were prepared for each strain for a total of 18 mice. All raw data is available on the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo, GSE19282).

Data analysis

Data processing and analysis were performed using R (http://www.R-project.org) and Bioconductor (http://www.bioconductor.org) (Gentleman et al. 2004). Array images were visually scanned for artifacts while quality control reports (affyQCReport) (Parman and Halling 2008) were assessed and determined to be acceptable under Affymetrix guidelines (Affymetrix GeneChip Expression Analysis Data Analysis Fundamentals). MAS5 detection calls were obtained using the Bioconductor *affy* package (Gautier et al. 2004). MAS5 detection calls are used to determine whether a particular probe set is detected above background. Only probe sets called present (detection $P < 0.05$) were utilized for analysis. The associated false discovery rate (Benjamini and Hochberg 1995) for a detection P value of 0.05 was $28.9 \pm 11.3\%$ over all arrays. To accommodate for false discoveries, probe sets were filtered for those present across all whole-blood or globin-reduced samples prior to gene set enrichment analysis (McClintick and Edenberg 2006).

Gene set enrichment analysis

Probe sets that were present in all 18 samples were assigned to two categories (whole-blood RNA or globin-reduced RNA) (Supplementary Tables 1 and 2) and imported into MetaCore (http://www.genego.com) for enrichment analysis in GeneGo Pathway Maps, GeneGo Diseases by Biomarkers (GeneGo, St. Joseph, MI), GO Processes, and GO Molecular Functions. GeneGo Pathway Maps represent a set of genes participating in a consecutive set of metabolic signals, or metabolic transformations, confirmed as a whole by experimental data or by inferred relationships. GeneGo Diseases by Biomarkers are groups of genes implicated in certain diseases based on classifications in Medical Subject Headings (http://www.nlm.nih.gov/mesh/). The list of genes represented on the Affymetrix Mouse 430 2.0 array was used as a base gene list when calculating P values in the MetaCore enrichment procedures. MetaCore (GeneGo) uses a hypergeometric model to determine the significance of enrichment (Falcon and Gentleman 2007).

Comparison to human whole-blood gene expression and tissue gene expression

For mouse and human whole-blood gene expression comparisons, raw human whole-blood gene expression data were downloaded from the Gene Expression Omnibus (GSE2888, GSE 16728). MAS5 calls were obtained as outlined above and filtered for orthologous probe sets present in 80% or more of globin-reduced samples for each study individually. Orthologous genes and their associated Affymetrix probe sets were identified using Ensembl Biomart (http://www.www.ensembl.org/biomart). Filtered probe set lists (Supplementary Table 3) were imported into MetaCore and tested for enrichment in GeneGo Pathway Maps. The list of genes represented on the Affymetrix HG U133A array was used as a base gene list when calculating P values in the MetaCore enrichment procedures for GSE2888 and GSE16728.
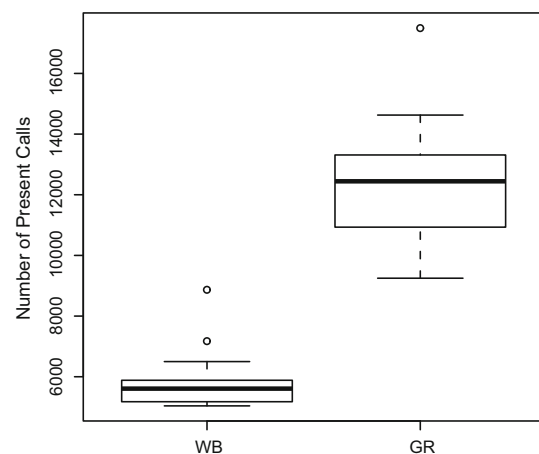
Fig. 1 Boxplots of present calls in whole-blood RNA and globin-reduced blood RNA samples. The *boxes* represent the lower quartile through the upper quartile, while the whiskers extend to 1.5 times the interquartile range. *Open circles* denote outliers. A *bold line* denotes the median. *WB* whole-blood RNA, *GR* globin-reduced blood RNA

Results

Probe detection and microarray sensitivity following globin reduction

Consistent with previously published study results, reduction of globin mRNA in mouse whole blood resulted in a consistent increase in the number of probe sets detected and improved microarray sensitivity, particularly for low-abundance genes. The average number of present calls in globin-reduced samples was 12,411 ± 1904 compared to 5840 ± 944 in untreated samples (Fig. 1), while 5383 probe sets were present across all samples following globin reduction in contrast to 1791 present probe sets in whole-blood RNA. Of the probe sets present across all arrays ($n = 5400$), 3609 probe sets were unique to globin-reduced RNA, 17 to whole-blood RNA, and 1774 were common to both whole-blood and globin-reduced RNA. Greater than

Table 1 Distribution of probe sets detected in (1) globin-reduced samples only, (2) whole-blood samples only, or (3) both whole-blood and globin-reduced samples across all detectable probes ($n = 5400$)

Percentile	Globin-reduced only (total = 3609)		Whole-blood only (total = 17)		Common (total = 1774)	
	Whole-blood RNA	Globin-reduced RNA	Whole-blood RNA	Globin-reduced RNA	Whole-blood RNA	Globin-reduced RNA
<5%	264	256	3	9	3	5
5–25%	1063	1050	2	4	15	26
25–50%	1269	1242	2	2	79	106
50-75%	878	899	4	2	468	449
75–95%	135	162	5	0	940	918
>95%	0	0	1	0	269	270

Detectable probes were ranked according to their average, normalized expression intensities in both whole-blood and globin-reduced RNA
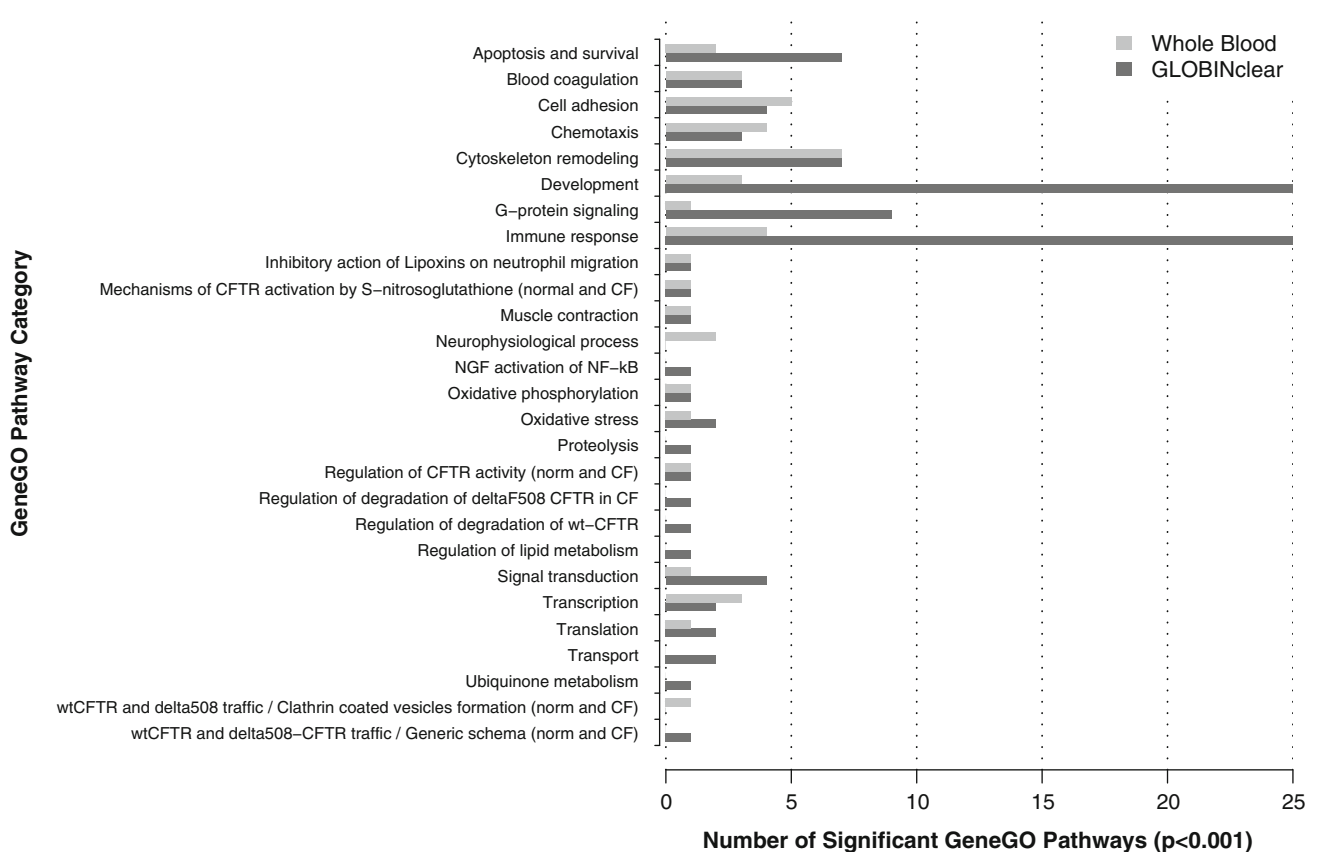
Fig. 2 Number of significantly enriched GeneGO pathways maps at a threshold of $P \leq 0.001$ in mouse whole-blood RNA (*gray*) and globin-reduced RNA (*black*). The associated false discovery rate is less than 0.01 for a P value threshold less than or equal to 0.001

one third of all probe sets present only in globin-reduced RNA were ranked among the bottom 25% of all detectable probe sets, while less than 4% were among the top 25% (Table 1). In contrast, two thirds of the probe sets detected in both whole-blood and globin-reduced blood RNA were among the highest 25%. Again, this suggests that globin reduction has a greater influence on the ability to detect genes expressed at low levels in whole blood and supports the idea that high levels of globin mRNA decrease detection sensitivity.

Gene set enrichment analysis of consistently present probe sets in whole and globin-reduced RNA

In order to evaluate the potential of expression profiles generated from globin-reduced mouse blood compared to whole mouse blood, probe sets detected as present across all samples in whole-blood RNA ($n = 1791$; Supplementary

Table 1) and globin-reduced RNA ($n = 5383$; Supplementary Table 2) were imported into MetaCore for gene set enrichment analysis. Globin reduction increased the ability to detect genes in peripheral whole blood involved in a variety of different biological pathways, most notably development, g-protein signaling, and immune response (Fig. 2). In total, the number of significantly enriched ($P < 0.001$) GeneGo Pathway Maps increased from 43 in whole-blood samples to 107 in globin-reduced samples. Globin-reduced samples were similarly enriched for GO biological processes and molecular functions (Supplementary Figs. 1 and 2), including processes and molecular functions previously described as enriched in globin-reduced human blood. This indicates that the globin expression signal significantly weakens the ability to detect the expression levels of many genes interrogated on microarrays, not necessarily due to some biological connection with globin's involvement in particular biological processes, but

Table 2 Top 25 statistically significant GeneGO disease categories in globin-reduced mouse blood RNA as compared to whole-blood RNA

GeneGO disease (by Biomarker)	Globin-reduced RNA		Whole-blood RNA		McNemar test
	Network objects	P value[a]	Network objects	P value[a]	
Hemic and lymphatic diseases	597/1927	1.02×10^{-11}	251/1927	5.75×10^{-09}	$<2.2 \times 10^{-16}$
Hematologic diseases	424/1316	1.04×10^{-10}	195/1316	1.21×10^{-11}	$<2.2 \times 10^{-16}$
Neoplasms	1571/5827	4.24×10^{-10}	626/5827	3.45×10^{-09}	$<2.2 \times 10^{-16}$
Anemia	135/344	2.25×10^{-09}	73/344	2.24×10^{-11}	2.13×10^{-12}
Bone marrow diseases	256/748	2.60×10^{-09}	114/748	1.25×10^{-07}	$<2.2 \times 10^{-16}$
Spherocytosis, hereditary	13/13	1.50×10^{-08}	9/13	3.16×10^{-07}	NS
Tay-Sachs disease	11/11	2.40×10^{-07}	8/11	8.45×10^{-07}	NS
Gangliosidoses	12/13	5.99×10^{-07}	8/13	5.53×10^{-06}	NS
Anemia, hemolytic, congenital	66/153	6.44×10^{-07}	40/153	1.81×10^{-09}	9.44×10^{-07}
Myelodysplastic syndromes	210/630	7.85×10^{-07}	89/630	7.19×10^{-05}	$<2.2 \times 10^{-16}$
Immunoproliferative disorders	459/1538	1.34×10^{-06}	182/1538	5.51×10^{-04}	$<2.2 \times 10^{-16}$
Lymphoproliferative disorders	458/1537	1.68×10^{-06}	181/1537	7.34×10^{-04}	$<2.2 \times 10^{-16}$
Neoplastic processes	280/887	2.21×10^{-06}	144/887	1.75×10^{-11}	$<2.2 \times 10^{-16}$
Gangliosidoses GM2	11/12	2.22×10^{-06}	8/12	2.32×10^{-06}	NS
Neoplasms by site	1091/3996	3.36×10^{-06}	430/3996	1.65×10^{-04}	$<2.2 \times 10^{-16}$
Breast neoplasms	539/1852	3.88×10^{-06}	212/1852	1.09×10^{-03}	$<2.2 \times 10^{-16}$
Anemia, hemolytic	77/194	4.03×10^{-06}	47/194	1.08×10^{-09}	1.19×10^{-07}
Breast diseases	539/1853	4.17×10^{-06}	212/1853	1.12×10^{-03}	$<2.2 \times 10^{-16}$
Lymphatic diseases	464/1572	4.60×10^{-06}	182/1572	1.63×10^{-03}	$<2.2 \times 10^{-16}$
Leukemia, myeloid	193/586	5.52×10^{-06}	80/586	5.11×10^{-04}	$<2.2 \times 10^{-16}$
Leukemia, nonlymphocytic, acute	155/455	6.20×10^{-06}	64/455	8.63×10^{-04}	$<2.2 \times 10^{-16}$
Leukemia	339/1115	8.07×10^{-06}	128/1115	1.17×10^{-02}	$<2.2 \times 10^{-16}$
Thalassemia	29/55	9.49×10^{-06}	18/55	1.71×10^{-06}	2.57×10^{-03}
Neoplasms by histologic type	1092/4027	1.59×10^{-05}	442/4027	1.00×10^{-05}	$<2.2 \times 10^{-16}$
RNA virus infections	195/603	1.80×10^{-05}	77/603	4.21×10^{-03}	$<2.2 \times 10^{-16}$

Network objects represent the proportion of disease-associated genes (biomarkers) identified per disease category, with the denominator representing the number of objects assayed by the Affymetrix Mouse 430 2.0 array

[a] GeneGO hypergeometric model

rather by virtue of their globin-relative signal strength detectable via chip-based multiprobe hybridization. The genes that happen to be affected by this phenomenon collectively participate in a number of biologically meaningful functions and processes. Thus, studies that do not reduce globin yet investigate mouse blood gene expression as a way of understanding disease processes are likely to fail to implicate many important genes since their expression levels are masked by globin.

The ability to detect genes previously implicated in disease was also significantly improved following globin reduction. In general, both whole-blood and globin-reduced blood RNA were significantly enriched for genes associated with diseases involving all cell types found in peripheral blood, including thrombocytes/platelets (thrombocytopenia), erythrocytes (anemia), myelocytes (myeloid leukemia), and lymphocytes (lymphoma) (Table 2, Supplementary Tables 4 and 5). Most importantly, the average potential to identify expression levels of genes involved in a given disease or biological network more than doubled following globin reduction as seen by the proportion of disease biomarkers or disease-associated genes identified as present. To test whether the removal of globin transcripts has a statistically significant effect on the proportion of genes identified, significantly enriched GeneGo Diseases ($P \leq 0.05$) in either whole-blood RNA or globin-reduced RNA were assessed using the McNemar test. For the majority of significantly enriched disease networks, the reduction of globin transcripts significantly improved the ability to identify biomarkers of disease (data not shown), including a variety of nonhematologic-based diseases such as neuromuscular diseases, neurodegenerative diseases, and chromosome aberrations (Table 3).

Comparisons to human whole blood

In order to evaluate the relevance of mouse blood gene expression to human blood gene expression profiles, the results of our mouse expression study were compared to two human studies by gene set enrichment analysis (Lu et al. 2009) (GSE2888; GSE16728) (Table 4). Of the 325 statistically significantly enriched GeneGo Pathway Maps ($P < 0.001$), 233 were significantly enriched in our mouse study and the two human studies (Fig. 3) and include 97 of the 107 pathways found to be significant after globin reduction of mouse whole-blood RNA, including the development and immune response pathways. Although these results are not completely definitive due to differences in protocols used across each study, including different globin reduction methods, the number of samples, and microarray designs, the overlap in significantly enriched pathway categories suggests that genes expressed in mouse peripheral blood reflect those expressed in human blood.

Discussion

Our analysis of the effects of globin reduction on mouse whole-blood-derived total RNA confirms the previously observed increase in expression detection sensitivity and overall detection rate in both humans and mice (Field et al. 2007; Whitley et al. 2005, 2007) and further emphasizes the importance of globin reduction in evaluating biologically significant pathways and disease processes in mouse models. Gene set enrichment analysis also indicates globin-reduced mouse blood RNA is a reasonable and practical

Table 3 Nonhematologic-based diseases with a significantly improved proportion of gene expression levels detected in globin-reduced mouse blood RNA as compared to whole mouse blood RNA

GeneGO disease (by biomarkers)	Whole-blood network objects	Globin-reduced network objects	P value[a]
Alzheimer's disease	87/721	194/721	$<2.20 \times 10^{-16}$
Amyotrophic lateral sclerosis	19/167	54/167	9.08×10^{-9}
Cerebellar ataxia	3/32	13/32	4.43×10^{-3}
Dementia	97/829	226/829	$<2.20 \times 10^{-16}$
Down syndrome	21/133	41/133	2.15×10^{-5}
Muscle hypertonia	5/48	21/48	1.77×10^{-4}
Neurodegenerative diseases	161/1289	354/1289	$<2.20 \times 10^{-16}$
Ophthalmoplegia	1/34	12/34	2.57×10^{-3}
Parkinson's disease	38/283	64/283	9.44×10^{-7}
Schizophrenia	106/752	208/752	$<2.20 \times 10^{-16}$

All diseases are significantly enriched with a GeneGO hypergeometric $P \leq 0.05$ in whole or globin-reduced blood RNA samples. Network objects represent the proportion of disease-associated genes identified per disease category, with the denominator representing the number of objects assayed by the Affymetrix Mouse 430 2.0 array

[a] McNemar test comparing the proportion of network objects identified before and after globin reduction

Table 4 Data set characteristics for human-mouse gene set enrichment analysis

Study	Microarray	Globin reduction method	No. of globin-reduced samples	Probe sets present in ≥80% of samples	Significant GeneGO pathways ($P < 0.001$)
Mouse	Affymetrix Mouse 430 2.0	GLOBINclear	18	5428	267
GSE2888	Affymetrix HG U133A	Affymetrix protocol	14	6048	279
GSE16728	Affymetrix HG U133 Plus 2.0	GLOBINclear	10	9363	296
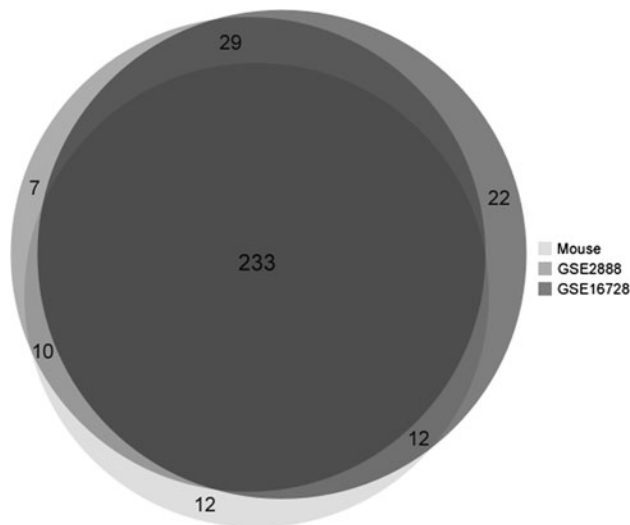
Fig. 3 Venn diagram comparing significant GeneGO pathway maps ($P \leq 0.001$) enriched in globin-reduced mouse RNA and globin-reduced human (GSE2888 and GSE16728) RNA. The associated false discovery rate is approximately 0.01 for a P value threshold less than or equal to 0.001

model for the study of blood-based gene expression correlates of human disease, as biological pathways significantly enriched in globin-reduced mouse and human blood overlap considerably.

However, globin reduction may not always be beneficial or necessary in certain disease contexts. Our results suggest that globin reduction appears to have little effect on the ability to detect certain classes of biological pathways (Fig. 2), while a small number of genes significantly decrease in expression or fall below the limits of detection following globin reduction, as noted in previous studies (Field et al. 2007). A BLAST search did not identify significant homology between the 17 probe sets whose expression level-based presence was unique to whole-blood RNA samples and the globin gene family, suggesting that these probe sets are not specifically removed during globin reduction but rather decrease in expression due to a slight decline in RNA quality (Vartanian et al. 2009) or other nonspecific effects. Nonetheless, these findings indicate that globin-reduced, peripheral blood-based gene expression profiling of relevant mouse models may reveal unique patterns of gene expression relevant to human

disease and aid in the discovery of clinically significant biomarkers.

References

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS et al (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Chao A, Wang TH, Lee YS, Hong JH, Tsai CN et al (2008) Analysis of functional groups of differentially expressed genes in the peripheral blood of patients with cervical cancer undergoing concurrent chemoradiation treatment. Radiat Res 169:76–86

Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV et al (2005) Application of genome-wide expression analysis to human health and disease. Proc Natl Acad Sci USA 102:4801–4806

Coppola G, Karydas A, Rademakers R, Wang Q, Baker M et al (2008) Gene expression study on peripheral blood identifies progranulin mutations. Ann Neurol 64:92–96

Debey S, Schoenbeck U, Hellmich M, Gathof BS, Pillai R et al (2004) 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

Debey S, Zander T, Brors B, Popov A, Eils R et al (2006) 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

Dyrskjot L, Thykjaer T, Kruhoffer M, Jensen JL, Marcussen N et al (2003) Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet 33:90–96

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258

Fan H, Hegde PS (2005) The transcriptome in blood: challenges and solutions for robust expression profiling. Curr Mol Med 5:3–10

Field LA, Jordan RM, Hadix JA, Dunn MA, Shriver CD et al (2007) Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. Clin Biochem 40:499–502

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307–315

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R et al (2005) Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci USA 102:15533–15538

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537

Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP (2003) Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. Proc Natl Acad Sci USA 100:9608–9613

Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD et al (2009) Identifying blood biomarkers for mood disorders using convergent functional genomics. Mol Psychiatry 14:156–174

Leonardson AS, Zhu J, Chen Y, Wang K, Lamb JR et al (2010) The effect of food intake on gene expression in human peripheral blood. Hum Mol Genet 19:159–169

Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA (2006) The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med 147:126–132

Liu J, Walter E, Stenger D, Thach D (2006) Effects of globin mRNA reduction methods on gene expression profiles from whole blood. J Mol Diagn 8:551–558

Lu Y, Huggins P, Bar-Joseph Z (2009) Cross species analysis of microarray expression data. Bioinformatics 25:1476–1483

McClintick JN, Edenberg HJ (2006) Effects of filtering by Present call on analysis of microarray experiments. BMC Bioinformatics 7:49

Miller TE, You L, Myerburg RJ, Benke PJ, Bishopric NH (2007) Whole blood RNA offers a rapid, comprehensive approach to genetic diagnosis of cardiovascular diseases. Genet Med 9:23–33

Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P (2000) Molecular characterization of schizophrenia viewed by micro-array analysis of gene expression in prefrontal cortex. Neuron 28:53–67

Mohr S, Liew CC (2007) The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends Mol Med 13:422–432

Parman C, Halling C (2008) affyQCReport: a package to generate QC reports for Affymetrix array data. R package version 1.18.0. Available at: http://prs.ism.ac.jp/bioc/2.2/bioc/vignettes/affyQCReport/inst/doc/affyQCReport.pdf

Radich JP, Mao M, Stepaniants S, Biery M, Castle J et al (2004) Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics 83:980–988

Solmi R, Ugolini G, Rosati G, Zanotti S, Lauriola M et al (2006) Microarray-based identification and RT-PCR test screening for epithelial-specific mRNAs in peripheral blood of patients with colon cancer. BMC Cancer 6:250

Sullivan PF, Fan C, Perou CM (2006) Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet 141B:261–268

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536

Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR et al (2009) Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. BMC Genomics 10:2

Wang Z, Neuburg D, Li C, Su L, Kim JY et al (2005) Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environ Health Perspect 113:233–241

Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA et al (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci USA 98:1176–1181

Whitley P, Moturi S, Santiago J, Johnson C, Setterquist R (2005) Improved microarray sensitivity using whole blood RNA samples. Ambion TechNotes 12:20–23

Whitley P, Gonzales J, Goldrick M (2007) Improved gene expression profiling with mouse blood samples. Ambion TechNotes 13:27–28

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC et al (2003) Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA 100:1896–1901

Wright C, Bergstrom D, Dai H, Marton M, Morris M et al (2008) Characterization of globin RNA interference in gene expression profiling of whole-blood samples. Clin Chem 54:396–405

Wu K, Miyada G, Martin J, Finkelstein D (2007) Globin reduction protocol: a method for processing whole blood RNA samples for improved array results. Affymetrix Technical Note. Available at: http://media.affymetrix.com:80/support/technical/technotes/blood2_technote.pdf

Zapala MA, Hovatta I, Ellison JA, Wodicka L, Del Rio JA et al (2005) Adult mouse brain gene expression patterns bear an embryologic imprint. Proc Natl Acad Sci USA 102:10357–10362