UNIVERSITY OF CALIFORNIA, SAN DIEGO

Measuring and Correlating Blood and Brain Gene Expression Levels: Assays, Inbred Mouse Strain Comparisons, and Applications to Human Disease Assessment

A dissertation submitted in partial satisfaction of the requirements
for the degree of Doctor of Philosophy

in

Biomedical Sciences

by

Mary Elizabeth Winn

Committee in charge:

Professor Nicholas J Schork, Chair
Professor Gene Yeo, Co-Chair
Professor Eric Courchesne
Professor Ron Kuczenski
Professor Sanford Shattil

2011

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

The dissertation of Mary Elizabeth Winn is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2011

DEDICATION

To my parents,
Dennis E. Winn II and Ann M. Winn,
to my siblings,
Jessica A. Winn and Stephen J. Winn,
and
to all who have supported me throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

ACKNOWLEDGEMENTS

I would like to acknowledge the members of my committee for their time and guidance in this work. First and foremost, I would like to thank my committee chair, Nik Schork. Not only did Nik guide me through the development and pursuit of the research presented here, he also helped put me in a place to succeed by encouraging collaboration and by supporting my broad intellectual and career interests. I would also like to thank Gene Yeo, Eric Courchesne, Ron Kuczenski, and Sanford Shattil for serving on my committee and for their guidance and feedback from proposal to defense. In particular, I would like to thank Eric Courchesne for the use of his rich data set, allowing me to expand and pursue my research interests beyond model organism.

I owe a deep debt of gratitude to those that provided access to or helped collect the data that form the basis of this work. I thank Nik Schork, Matt Zapala, and Betsy Lillie for access to and information about the mouse gene expression data, Iiris Hovatta and Vickie Risborough for access to the mouse behavioral phenotypes, and Eric Courchesne for access to the data collected by the San Diego Autism Center of Excellence. I thank Sarah Murray, Marian Shaw, and all those at Scripps Genomic Medicine and the Scripps Translational Science Institute who collected and processed samples utilized in this work as well as Jian-Bing Fan, Craig April, and Brandy Klotzle for processing the Illumina DASL assays.

I also consider myself fortunate to have the support of my family and friends for which I am grateful for their encouragement and understanding. I want to extend a special thanks to my parents, Dennis and Ann Winn, my sister Jessica, and my

brother Stephen, for their unconditional support throughout my life and the pursuit of my graduate degree. I could not have made it to where I am today without their love and support. I would also like to acknowledge my many friends and colleagues in the Schork and Courchesne labs, the Scripps Translational Science Institute, and the Biomedical Sciences Graduate Program at UCSD for their camaraderie and support. Thank you to my mentors, David Ginsburg and Randy Westrick, for their guidance, support, and engaging discussions. Finally, I would like to thank my friends across the nation for keeping me grounded and providing balance along my journey.

Chapter 2, in full, has been published as: Mary E Winn, Matthew A Zapala, Iiris Hovatta, Victoria B Risbrough, Elizabeth Lillie, Nicholas J Schork (2010) The effects of globin on microarray-based gene expression analysis of mouse blood. Mammalian Genome 21, 268-275. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication. Mary E Winn, Matthew A Zapala, Iiris Hovatta, Victoria B Risbrough, Elizabeth Lillie, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been published as: Mary E Winn, Marian Shaw, Craig April, Brandy Klotzle, Jian-Bing Fan, Sarah S Murray, Nicholas J Schork (2011) Gene Expression Profiling of Human Whole Blood Samples with the Illumina WG-DASL Assay. BMC Genomics 12, 412-412. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, is currently being prepared for submission for publication.

Mary E Winn, Tiziano Pramparo, Maggie L Chow, Anthony Wynshaw-Boris, Karen Pierce, Eric Courchesne, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

VITA

EDUCATION

2005 Bachelor of Science, Cellular and Molecular Biology, University of Michigan, Ann Arbor

2011 Doctor of Philosophy, Biomedical Sciences, University of California, San Diego

PROFESSIONAL EXPERIENCE

2007 – 2011 Graduate Student Researcher, Scripps Translational Science Institute, Scripps Genomic Medicine, and University of California, San Diego, Department of Medicine, Nicholas J Schork.

2005 – 2007 Research Lab Technician Associate, University of Michigan, Howard Hughes Medical Institute, David Ginsburg.

PUBLICATIONS

2011

Thomas GS, Voros S, McPherson JA, Lansky AJ, Weiland FL, Cheng SC, Bloom SA, Salha HH, Winn ME, Bateman TM, Elashoff MR, Lieu HD, Johnson AM, Daniels SE, Rosenberg S. Gene Expression Testing For Obstructive Coronary Artery Disease in Symptomatic Non-Diabetic Patients Referred For Myocardial Perfusion Imaging: The Compass Trial. *Submitted*

Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, Fan JB, April C, Fu XD, Wynshaw-Boris A, Schork NJ, Courchesne E. Age Dependent Brain Gene Expression and Copy Number Anomalies in Autism Suggest Distinct Age-Dependent Pathological Processes. *Submitted*

Chow ML, Winn ME, Li HR, April C, Wynshaw-Boris A, Fan JB, Fu XD, Courchesne E, Schork NJ. Preprocessing and Quality Control Strategies for Illumina DASL Assay-Based Brain Gene Expression Studies with Semi-Degraded Samples. *Submitted*

Chow ML, Li HR, Winn ME, April C, Barnes CC, Wynshaw-Boris A, Fan JB, Fu XD, Courchesne E, Schork NJ (2011) Genome-wide Expression Assay Comparison Across Frozen and Fixed Postmortem Brain Tissue Samples. BMC Genomics 12, 449

Winn ME, Shaw M, April C, Klotzle B, Fan JB, Murray SS, Schork NJ (2011) Gene Expression Profiling of Human Whole Blood Samples with the Illumina WG-DASL Assay. BMC Genomics 12, 412-412

2010

Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT, Sehnert AJ, Yau M, Kraus WE, Newby LK, Schwartz RS, Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A, Winn ME, Schork NJ, Topol EJ (2010) Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. Annals of Internal Medicine 153, 425-434

Westrick RJ, Mohlke KL, Korepta LM, Yang AY, Zhu G, Manning SL, Winn ME, Dougherty KM, Ginsburg D (2010) Spontaneous Irs1 passenger mutation linked to a gene-targeted SerpinB2 allele. Proceedings of the National Academy of Sciences of the United States of America 107, 16904-16909

Winn ME, Zapala MA, Hovatta I, Risbrough VB, Lillie E, Schork NJ (2010) The effects of globin on microarray-based gene expression analysis of mouse blood. Mammalian Genome 21, 268-275

2007

Westrick RJ, Winn ME, Eitzman DT (2007) Murine models of vascular thrombosis (Eitzman series). Arteriosclerosis, Thrombosis, and Vascular Biology 27, 2079-2093

ABSTRACTS

2011

Winn ME, Schork NJ, Pierce KL, Courchesne E (2011) Blood-Based Gene Expression In Infants and Toddlers with ASD. International Meeting for Autism Research

Chow ML, Pramparo T, Winn ME, Stoner R, Boyle MP, Lein E, Roy S, Li HR, Fan JB, April C, Fu XD, Colamarino S, Mouton P, Weiss L, Schork NJ, Wynshaw-Boris A, Courchesne E (2011) Aberrant Proliferative and Organizational Pathways with Disrupted Cortical Lamination In Young Autistic Males. International Meeting for Autism Research

Glatt SJ, Winn ME, Roe C, Wong T, Ahrens-Barbeau C, Chandler S, Collins M, Lopez L, Tsuang M, Pierce KL, Schork NJ, Courchesne E (2011) Blood-Based Transcriptomic Biomarker Profiles of Autistic Spectrum and Other Developmental Disorders. International Meeting for Autism Research

2009

Bloss CS, Winn ME, Shaw M, Cary M, Topol EJ, Schork NJ, Houser MK, Murray SS (2009) Gene Expression Profiling to Develop Blood-Based Biomarkers of Parkinson's Disease Dementia: A Pilot Study. Scripps Clinical Research Open House

Winn ME, Shaw M, April C, Klotzle B, Fan JB, Bloss CS, Murray SS, Topol EJ, Schork NJ (2009) Gene Expression Profiling of Human Whole Blood Samples with the Illumina DASL Assay. 59th American Society of Human Genetics Meeting

2006

Westrick RJ, Manning SL, Winn ME, Dobies SL, Stotz GM, Siemieniak DR, Sanford E, Ginsburg D (2006) A Sensitized ENU Mouse Mutagenesis Screen for Dominant Genetic Modifiers of Thrombosis in the Factor V Leiden Mouse. 20th International Mouse Genome Conference

2005

Westrick RJ, Manning SL, Dobies SL, Winn ME, Siemieniak DR, Korepta LM, and Ginsburg D (2005) A Sensitized ENU Mouse Mutagenesis Screen to Identify Genetic Modifiers of Thrombosis in the Factor V Leiden Mouse. 19th International Mouse Genome Conference

ABSTRACT OF THE DISSERTATION

Measuring and Correlating Blood and Brain Gene Expression Levels: Assays, Inbred Mouse Strain Comparisons, and Applications to Human Disease Assessment

by

Mary Elizabeth Winn

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2011

Professor Nicholas J Schork, Chair

Professor Gene Yeo, Co-Chair

Microarray-based gene expression profiling is a frequently utilized tool in the search for disease-specific molecular patterns and the development of clinically relevant panels of biomarkers. Although advances in high-throughput gene expression technology make for more reliable and interpretable studies, investigations of living humans are often limited by tissue accessibility. This is especially true for neural-based illnesses, where studies rely heavily on post-mortem brain tissue. As a result, medical researchers have focused on blood, a more easily

accessible and clinically obtainable tissue. In this work I explore: 1.) the technical aspects associated with assessing peripheral whole blood gene expression via microarray; and 2.) the biological significance of blood-based gene expression patterns with respect to brain-based gene expression patterns and behavioral phenotypes in mice and humans. I describe the effects of globin reduction on blood-based gene expression in mice by comparing gene expression patterns before and after globin reduction of mouse whole blood (Chapter 2). Globin reduction was found to improve the ability to detect low abundance, biologically relevant genes. I also evaluated globin reduction in the context of human blood and two Illumina gene expression assays: (i) the IVT-based direct hybridization assay; and (ii) the WG-DASL assay (Chapter 4). As in mice, I was able to recapitulate the known benefits of globin reduction in both assays, while WG-DASL appeared to be more sensitive compared to IVT. Lastly, I characterized the correlations between blood gene expression levels and behavioral phenotypes and compared blood gene expression-trait correlations with brain gene expression-trait correlations in respect to neuropsychiatric phenotypes in mice (Chapter 3) and autism in humans (Chapter 5). In both mice and humans, blood was only able to capture a small portion of the associations identified in the brain on an individual gene level. At a pathway level, blood was able to capture a larger portion of the associated brain pathways in humans as compared to mice. I conclude blood gene expression, although it may capture a small portion of the expression patterns associated with 'primary' neural insults, is more likely to capture variation due to 'secondary' perturbations or other biological and environmental insults.

CHAPTER 1

Introduction and Background

INTRODUCTION

Translational Genomics, Genome-wide Expression Analysis, and Biomarker Discovery

Since the completion of the Human Genome Project in 2003 (International Human Genome Sequence Consortium 2004), the Phase1 and Phase 2 HapMap Projects in 2005 (International HapMap Consortium 2005) and 2007 (Frazer et al. 2007), and the pilot phase of the Encyclopedia of DNA Elements (ENCODE) project in 2007 (Birney et al. 2007), there has been an explosion in the development of resources and studies aimed at identifying the fundamental causes of complex human disease (Topol et al. 2007). In order to determine the specific genetic factors mediating disease susceptibility, researchers have utilized a variety of strategies including direct DNA sequencing, single-nucleotide polymorphism (SNP) genotyping-based genome-wide association studies, global gene expression, proteomic and metabolomic studies, and *in silico* and computational model analysis of gene and sequence function. Much of the focus of these research efforts has been to identify genetic factors contributing to disease – either as stable markers of disease susceptibility or as 'biomarkers' whose elevations and de-elevations are indicative of disease pathogenesis – so that the resulting insights can be 'translated' into viable medical practices and there by usher in an era of genetics-based "personalized medicine" (Feero et al. 2008).

Biomarker discovery thus plays a key role in the process of translating scientific breakthroughs to the bedside. As a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention, a biomarker

can be used to classify disease states, develop diagnostic and prognostic tools, or illustrate drug efficacy and toxicity (Atkinson et al. 2001). Classic biomarkers have included physiological measurements (blood pressure, cholesterol levels), imaging (x-ray), and protein molecules (human chorionic gonadotropin (hCG), C-reactive protein (CRP)). Advances in genomics and associated molecular biology and pathology technologies have served to revolutionize the field, giving rise to hundreds (if not thousands) of potential genetic and molecular markers of disease, such as SNP genotypes and gene expression profiles. However, many of these candidate biomarkers have failed to be effective when tested in clinical settings. As of August 1, 2011, the US Food and Drug Administration (FDA) had approved 81 genetic biomarkers from the hundreds that have been implicated in pharmacogenetic studies (**Table 1-1**) (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm).

Gene expression is a highly regulated process by which genomic information is converted to a functional gene product such as RNA or protein. Individual differences in gene expression that arise from genetic, epigenetic, or environmental variation are likely an underlying cause of complex human disease. Microarray-based gene expression profiling has become a frequently employed tool in the search for disease-specific molecular patterns (i.e., disease 'fingerprints') and the development of clinically relevant panels of biomarkers. Since the advent of microarray technology in the mid-1990s, over 30,000 articles have been published on microarray gene expression with over fifty percent published in the last four years (**Figure 1-1**). Investigations utilizing microarrays have yielded insights into disease classifications (Golub et al. 1999, Alizadeh et al. 2000, Dyrskjot et al. 2003),

diagnostic and prognostic gene profiles (Mirnics et al. 2000, Welsh et al. 2001, van 't Veer et al. 2002), and drug efficacy and toxicity profiles (Gunther et al. 2003).

Neuropsychiatric Diseases, Tissue Accessibility and Blood-based Gene Expression

Although advances in gene expression technologies probably made studies of gene expression more reliable and interpretable than they were 20 years ago, gene expression studies on living humans are still limited by tissue accessibility. This is especially true for neuropsychiatric illnesses, where it is nearly impossible to collect brain samples from living individuals. As a result, the majority of studies of neuropsychiatric diseases rely on post-mortem brain tissue. Alternatively, medical researchers have begun to focus on peripheral blood (e.g., leukoyctes, lymphocytes, lymphoblastoid cell lines (LCLs), peripheral blood mononuclear cells (PBMCs), whole blood), a more easily accessible and clinically obtainable tissue.

Aside from accessibility, the physiological characteristics of blood cells suggest peripheral blood is an ideal surrogate for primary tissue (Liew et al. 2006). Importantly, the natural variation, heritability, and processing-induced variation of blood-based gene expression have been examined (**Table 1-2**), while the advantages and disadvantages of studying blood gene expression patterns have been carefully reviewed (Fan et al. 2005, Mohr et al. 2007). Nonetheless, the biological relevance of gene expression levels in the blood to human neuropsychiatric disease remains relatively unknown, given that brain tissue and blood would be difficult to collect from the same living individual (Sullivan et al. 2006). Given the significant amount of intra- and inter- individual variation in blood gene expression, such an approach limits

interpretability and brings in to question a study's ability to identify reliable and replicable biomarkers for the diagnosis and treatment of human disease.

Blood-based gene expression correlations with neuropsychiatric conditions and behaviors are quite likely to reflect 'secondary' if not 'primary' molecular perturbations in the diseased brain, such as the presence of a tumor or an immunological insult. As a result, the secondary patterns of, or changes in, gene expression identified in the blood may be seen as biomarkers of a disease state rather than potential targets for pharmacological intervention. On the other hand, in certain neuropsychiatric states, there could be an actual over- or under- deposition of, e.g., neurohormones, into the bloodstream that reflects the primary (or one of the *many* primary) etiological defects contributing to these states. Thus, ultimately, blood-based gene expression patterns associated with certain neuropsychiatric states may indeed reflect a combination of primary and secondary effects.

Mouse Models of Human Disease

Unlike human subjects, blood and primary tissue samples can be easily collected from the same living mouse under highly controlled conditions. Intra- and inter-individual sources of variation introduced by gender, age, time of day, genetic variation, and environment can be studied and accommodated (Whitney et al. 2003, Radich et al. 2004, Cobb et al. 2005), while clinically acceptable and highly standardized protocols for blood collection, RNA isolation, and globin reduction can be employed to lessen technical-induced variation (Debey et al. 2004, Cobb et al. 2005, Debey et al. 2006). Mice have also been widely used to mimic and provide insight into the genetic basis of human disease since the early 1900s (Rosenthal and

Brown 2007) with an assortment of valuable online resources now available (**Table 1-3**)(Peters et al. 2007).

Studies of human disease have benefited from the vast pool of genetic resources developed and utilized in mice, such as: 1. quantitative trait locus (QTL) mapping to identify genomic locations harboring mutations that influence a relevant phenotype's expression; 2. the creation of chromosome substitution, recombinant-inbred, and congenic mouse strains to explore the impact of specific chromosomes and genetic locations on phenotype expression; 3. *in silico* mapping studies which compare known polymorphic sites across different mouse strains to phenotypic differences exhibited by those strains; 4. complex "genetical genomics" analyses mixing mouse strain crosses, QTL mapping strategies, and gene expression studies to uncover complex gene expression regulatory networks behind phenotype expression; 5. mutagenesis strategies designed to correlate specific mutations with specific phenotypes; and 6. knockout and transgenic studies which consider the effects of specific genes on the expression of a particular phenotype (Peters et al. 2007). Putting such efforts and studies into context requires understanding the genetic and phenotypic backgrounds of the strains used in these studies. In this light, characterizing correlations between blood and brain gene expression levels and gene-phenotype relationships across commonly used inbred mouse strains, although not trivial, may contribute to our understanding of previous mouse-based studies, yield compelling candidate genes for human neuropsychiatric diseases, and provide insight into the potential of blood-based gene expression patterns as biomarkers of neuropsychiatric conditions.

As noted, however, mouse models of human disease are not without their

drawbacks (i.e. genetic background) (Rivera et al. 2008). This is especially true for models of neuropsychiatric disorders and behavioral phenotypes, which are hampered not only by a variety of genetic and environmental factors but also by a lack of clear disease and behavioral definitions for mice (Bucan and Abel 2002, Cryan and Mombereau 2004). Furthermore, gene-phenotype studies in humans that are based on gene expression studies in mice are complicated by a lack of insight as to how the genes in question fit into larger species and tissue-specific regulatory networks that influence phenotypic expression, with no firm understanding of the issues and potential effects these forces may have had in shaping the function or impact of a gene of interest and its associated regulatory network in the human biochemical and physiologic milieu. Thus, it is important to fully characterize the genetic, regulatory, biochemical, and phenotypic backgrounds of the mouse strains used in studies of neuro-cognitive and behavioral phenotypes in an effort to not only put previous studies into context, but also direct future studies.

Microarray Gene Expression Profiling and Globin Reduction

Since blood is high in alpha- and beta- globin, which tend to confound detection of expression levels of non-globin genes, microarray-based gene expression profiles of peripheral whole blood suffer from poor sensitivity and high variability, hampering their utility as reliable and reproducible clinical biomarkers of disease. A variety of globin reduction methods were developed to address the needs associated with microarray-based evaluation of peripheral whole blood, including: peptide nucleic acids (Debey et al. 2006) and magnetic beads (Whitley et al. 2005). These globin reduction methods have been tested using a variety of microarray

platforms (Tian et al. 2009; Vartanian et al. 2009; Dumeaux et al. 2008; Debey et al. 2006) and often validate the importance of globin reduction in the assessment of human peripheral whole blood. These studies also indicate globin reduction may not always be beneficial or necessary. Globin reduction of human whole blood samples is known to require a relatively large amount of input RNA (Vartanian et al. 2009), may induce its own unique expression profile (Liu et al. 2006), or fail to improve reproducibility (Vartanian et al. 2009; Dumeaux et al. 2008). It is therefore necessary to continue to consider the effects of globin reduction as new microarray platforms are developed and different organisms (i.e. *Mus musculus* or *Rattus norvegicus*) are used as the source of whole blood.

Although globin reduction methods have been developed for mouse whole blood (Whitley et al. 2007), little has been done to characterize the effects of globin reduction of microarray gene expression profiling of mouse whole blood. Globin reduction of human peripheral whole blood improves detection levels of genes in various biological processes (Field et al. 2007) and genes relevant to disease (Raghavachari et al. 2009). Mouse whole blood, on the other hand, has a higher ratio of reticulocytes to lymphocytes (32:1) than human whole blood (9:1)(Fan et al. 2005). Globin reduction of mouse whole blood may not be able to overcome the significantly higher level of globin. If researchers are to use the mouse as a model of human disease it will be important to understand the limits of whole blood gene expression profiling in mice.

Researchers must also understand the need for globin reduction with various microarray platforms. Globin reduction has been shown to be effective using Affymetrix GeneChips (Vartanian et al. 2009) and the standard Illumina *in vitro*

transcription assay with (Tian et al. 2009) while globin reduction provided no benefit when used in conjunction with the Applied Biosystems AB1700 microarray system (Dumeaux et al. 2008). As microarray gene expression platforms continue to develop and become more accurate, so will the need to assess effects of globin reduction on whole blood gene expression profiling. For example, Tian et al. assessed globin reduction in the context of *in vitro* transcription amplification hybridized to the Illumina Sentrix HumanRef-6 BeadChip. The HumanRef-6 BeadChip has since been replaced by the HumanRef-8, and more recently by the HumanRef-12. Each BeadChip assesses a different number of probes with the potential to affect the background noise due to high levels of globin in whole blood. Other technologies created for the profiling of highly degraded samples, such as the Illumina Whole-Genome DASL (cDNA-mediated Annealing, Selection, extension and Ligation) (April et al. 2009), may also be potentially beneficial in the microarray gene expression profiling of peripheral whole blood.

Finding an Accessible Surrogate Tissue for Neural Tissue

Linkage and genome-wide association studies focusing on the identification of susceptibility genes for neural-based disorders have proven to be difficult (Gershon et al. 2011; Altshuler et al. 2008; Hovatta and Barlow 2008), most likely due to the heterogeneous and complex nature of neurodevelopmental, neuropsychiatric, and neurodegenerative diseases. One approach to disentangling these complex diseases has been the development of mouse models that mimic certain aspects of anxiety (Belzung and Griebel 2001; Crawley and Goodwin 1980), depression (Pollack et al. 2010), schizophrenia (Braff and Geyer 1990; Geyer et al. 1990), Parkinson's disease

(Taylor et al. 2010), autism (Moy et al. 2007; Moy et al. 2004), and other neural-based diseases. These models lend themselves well to the characterization of blood gene expression, or any other tissue source, as a surrogate for brain gene expression.

Taking advantage of the natural variation manifested by behavioral phenotypes across well-characterized inbred mouse strains, previous studies combining mouse strain analysis and behavioral testing with microarray gene expression profiling have identified genes whose expression levels are associated with behavior in mice (de Jong et al. 2010; Nadler et al. 2006; Hovatta et al. 2005). In some cases, the associated genes have also been shown to be associated with disease in human populations (Donner et al. 2008), highlighting the potential of mouse models in the identification of genes relevant to human disease. By expanding this approach to blood-based gene expression, it may be possible to discover disease susceptibility genes in an easily accessible, clinically relevant tissue.

The use of mouse models enables researchers to overcome many of the difficulties in analyzing whether or not blood is a viable surrogate for brain gene expression in identifying disease susceptibility genes, particularly small sample sizes and sample degradation, as well as the inability to collect blood and brain from the same living individual. Analysis of mouse brain and spleen suggest blood gene expression is capable of acting as a surrogate for brain tissue for a subset of genes (Davies et al. 2009), while gene expression experiments in vervet monkeys also exhibit the potential of blood gene expression to act as a surrogate for brain gene expression (Jasinska et al. 2009). Nonetheless, conflicting results from the assessment of human tissues display little overlap between human blood and brain

gene expression (Cai et al. 2010). While Cai et al. took advantage of multiple large data sets, their study suffers from several pitfalls: 1.) brain and blood samples used were processed using different microarray platforms (Affymetrix and Illumina, respectively); and 2.) brain and blood samples were collected from different individuals. Thus, the results from mouse studies will need to be extended to human studies in order to assess the true clinical validity of blood-based gene expression biomarkers.

Genetic Background Effect Analysis

As researchers continue to study the behavior of inbred mouse strains, they must be aware of strain differences that may affect the results of any study if not accounted for properly. For example, when assessing physiological differences between strains using an activity involving sight, variation across strains with respect to visual acuity and blindness will affect behavioral test performance. In addition, genetic variation between strains could influence phenotypic expression across the strains or also contribute to assay failure. Thus, the genetic background of the strain used to develop a knockout or gene transfer investigation can have enormous effects on the study aims and hypotheses (Austin et al. 2004; Accili 2004), and traditional introgression studies have showed varying effects of a target gene's activity as a function of genetic background (Letts et al. 1995). The same holds true for microarray gene expression analyses. Sequence variation has been shown to affect hybridization and lead to an increase in false associations (de Jong S et al. 2010; Peirce et al. 2006, Radcliffe et al. 2006). Methods capable of correcting for strain-based hybridization, such as GeSNP (Greenhall et al. 2007), should be considered.

For example, the algorithm employed by GeSNP was used to identify sequence differences between three rare strains of inbred mice (Carter et al. 2005), to improve the reliability of gene expression data by masking probe pairs that cover regions with sequence differences between humans and chimpanzees (Cáceres et al. 2003), and was also applied in the expression QTL (eQTL) study described above (Hovatta et al. 2007). The results demonstrate that the GeSNP algorithm can identify sequence differences using array-based gene expression data.

SPECIFIC AIMS

The research pursued in this dissertation focuses on the comparison of blood and brain microarray-based gene expression in mice and humans. Overall, the research focuses on: 1. the effects of globin reduction on blood gene expression levels and patterns; 2. the correlation of blood and brain gene expression to various behavioral phenotypes; and 3. the utility of blood gene expression in the development of genomic biomarkers for neural-based diseases using two existing data sets: (i) a data consisting of five brain tissues collected from six mouse strains (Hovatta et al. 2005); and (ii) data on post-mortem brain samples from the San Diego Autism Center of Excellence (Chow et al. 2011, Submitted). The specific aims are as follows:

1. The assessment of the effects of globin reduction on blood-based gene expression in mice by comparing gene expression patterns before and after globin reduction of mouse whole blood.

2. The characterization of the correlations between blood gene expression levels and behavioral phenotypes and how blood gene expression-trait correlations compare

with brain gene expression-trait correlations in order to assess the utility of blood gene expression levels for identifying neuropsychiatric and behavioral phenotypes.

3. The evaluation of gene expression profiling of human whole blood samples with two Illumina gene expression assays: (i) the *in vitro* transcription (IVT) assay; and (ii) the whole-genome cDNA-mediated Annealing, Selection, Extension, and Ligation (WG-DASL) assay.

4. The application of knowledge gained from specific aims I-III to the comparison of gene expression from brain and lymphocytes collected from individuals diagnosed with autism.

ENUMERATION OF CHAPTERS

Whole blood is widely recognized as an acceptable tissue source in clinically applicable gene expression studies. Nonetheless, blood-based gene expression profiling is not without challenges. Researchers must take in to account: 1. multiple cell types in the blood with varying expression patterns and cell counts; 2. intra- and inter-individual variation in cell composition of the blood; and 3. sample collection and processing-induced variation and alterations in gene expression patterns (Fan and Hegde 2005). One such challenge is the level of globin mRNA transcripts in whole blood; it is this challenge Chapter 2 serves to address in the context of mouse whole blood. Globin reduction in mouse whole blood is found to be important to improving the ability to detect genes involved in various human diseases, particularly neural-based diseases.

Chapter 3 focuses on mouse blood and mouse brain gene expression-behavioral trait correlations with an underlying emphasis on the effects of cell type

and strain-specific genetic background on blood- and brain- specific gene expression using Multivariate Distance Matrix Regression (MDMR) (Zapala et al. 2006) and Bi-variate Correlated Errors Analysis (Akritas and Bershady 1996). Strain effects on blood-specific gene expression were assessed using six commonly utilized inbred mouse strains (129S1/SvImJ, A/J, C3H/HeJ, C57BL/6J, DBA/2J, and FVB/NJ); blood cell count effects using data from the Mouse Phenome Database (http://phenome.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home); and brain tissue effects using regions (bed of nucleus striatum, hippocampus, hypothalamus, periaqueductal gray, and pituitary) known to play a role in neuropsychiatric phenotypes: pre-pulse inhibition, fear potentiated startle, and other anxiety phenotypes.

Chapters 4 and 5 present the results pertaining to human samples. The challenges related to globin reduction and microarray platform using human peripheral blood are addressed in Chapter 4. The fifth chapter extends the analyses pursued in specific aim 2 and the microarray platforms studied in specific aim 3 to an autism data set of lymphocyte gene expression. We find that human lymphocyte gene expression reflects similar pathways as those identified in brain gene expression. We also conclude that lymphocyte gene expression profiles are capable of discriminating between autism cases and controls.

TABLES

Table 1-1. 81 genetic biomarkers approved by the US Food and Drug Administration (FDA) as of August 1, 2011.

Drug	Therapeutic Area	Biomarker	Label Sections
Abacavir	Antivirals	HLA-B*5701	Boxed Warning, Contradindications, Warnings and Precautions, Patient Counseling Information
Aripiprazole	Psychiatry	CYP2D6	Clinical Pharmacology
Arsenic Trioxide	Oncology	PML/RARα	Boxed Warning, Clinical Pharmacology, Indications and Usage, Warnings
Atomoxetine	Psychiatry	CYP2D6	Dosage and Administration, Warnings and Precautions, Drug Interactions, Clinical Pharmacology
Atorvastatin	Metabolic and Endocrinology	LDL receptor	Indications and Usage, Dosage and Administration, Warnings and Precautions, Clinical Pharmacology, Clinical Studies
Azathioprine	Rheumatology	TPMT	Dosage and Administration, Warnings and Precautions, Drug Interactions, Adverse Reactions, Clinical Pharmacology
Boceprevir	Antivirals	IL28B	Clinical Pharmacology
Busulfan	Oncology	Ph Chromosome	Clinical Studies
Capecitabine	Oncology	DPD	Contraindications, Precautions, Patient Information
Carbamazepine	Neurology	HLA-B*1502	Boxed Warning, Warnings and Precautions
Carvedilol	Cardiovascular	CYP2D6	Drug Interactions, Clinical Pharmacology
Celecoxib	Analgesics	CYP2C9	Dosage and Administration, Drug Interactions, Use in Specific Populations, Clinical Pharmacology

Table 1-1. Continued.

Drug	Therapeutic Area	Biomarker	Label Sections
Cetuximab (1)	Oncology	EGFR	Indications and Usage, Warnings and Precautions, Description, Clinical Pharmacology, Clinical Studies
Cetuximab (2)	Oncology	KRAS	Indications and Usage, Clinical Pharmacology, Clinical Studies
Cevimeline	Dermatology and Dental	CYP2D6	Drug Interactions
Chloroquine	Antiinfectives	G6PD	Precautions
Clopidogrel	Cardiovascular	CYP2C19	Boxed Warning, Dosage and Administration, Warnings and Precautions, Drug Interactions, Clinical Pharmacology
Clozapine	Psychiatry	CYP2D6	Drug Interactions, Clinical Pharmacology
Codeine	Analgesics	CYP2D6	Warnings and Precautions, Use in Specific Populations, Clinical Pharmacology
Dapsone	Dermatology and Dental	G6PD	Indications and Usage, Precautions, Adverse Reactions, Patient Counseling Information
Dasatinib	Oncology	Ph Chromosome	Indications and Usage, Clinical Studies, Patient Counseling Information
Dexlansoprazole	Gastroenterology	CYP2C19	Clinical Pharmacology
Dextromethorphan and Quinidine	Neurology	CYP2D6	Clinical Pharmacology, Warnings and Precautions
Diazepam	Psychiatry	CYP2C19	Drug Interactions, Clinical Pharmacology
Doxepin	Psychiatry	CYP2D6	Precautions
Drospirenone and Ethinyl Estradiol	Reproductive	CYP2C19	Precautions, Drug Interactions
Erlotinib	Oncology	EGFR	Clinical Pharmacology
Esomeprazole	Gastroenterology	CYP2C19	Drug Interactions, Clinical Pharmacology

Table 1-1. Continued.

Drug	Therapeutic Area	Biomarker	Label Sections
Fluorouracil	Dermatology and Dental	DPD	Contraindications, Warnings
Fluoxetine	Psychiatry	CYP2D6	Warnings, Precautions, Clinical Pharmacology
Fluoxetine and Olanzapine	Psychiatry	CYP2D6	Drug Interactions, Clinical Pharmacology
Flurbiprofen	Rheumatology	CYP2C9	Clinical Pharmacology, Special Populations
Fulvestrant	Oncology	ER receptor	Indications and Usage, Patient Counseling Information
Gefitinib	Oncology	EGFR	Clinical Pharmacology
Imatinib (1)	Oncology	C-Kit	Indications and Usage, Dosage and Administration Clinical Pharmacology, Clinical Studies
Imatinib (2)	Oncology	Ph Chromosome	Indications and Usage, Dosage and Administration, Clinical Pharmacology, Clinical Studies
Imatinib (3)	Oncology	PDGFR	Indications and Usage, Dosage and Administration, Clincal Studies
Imatinib (4)	Oncology	FIP1L1-PDGFRα	Indications and Usage, Dosage and Administration, Clinical Studies
Irinotecan	Oncology	UGT1A1	Dosage and Administration, Warnings, Clinical Pharmacology
Isosorbide and Hydralazine	Cardiovascular	NAT1; NAT2	Clinical Pharmacology
Lapatinib	Oncology	Her2/neu	Indications and Usage, Clinical Pharmacology, Patient Counseling Information
Lenalidomide	Hematology	5q Chromosome	Boxed Warning, Indications and Usage

Table 1-1. Continued.

Drug	Therapeutic Area	Biomarker	Label Sections
Maraviroc	Antivirals	CCR5	Indications and Usage, Warnings and Precautions, Clinical Pharmacology, Clinical Studies, Patient Counseling Information
Mercaptopurine	Oncology	TPMT	Dosage and Administration, Contraindications, Precautions, Adverse Reactions, Clinical Pharmacology
Metoprolol	Cardiovascular	CYP2D6	Precautions, Clinical Pharmacology
Nelfinavir	Antivirals	CYP2C19	Drug Interactions, Clinical Pharmacology
Nilotinib (1)	Oncology	Ph Chromosome	Indications and Usage, Patient Counseling Information
Nilotinib (2)	Oncology	UGT1A1	Warnings and Precautions, Clinical Pharmacology
Panitumumab (1)	Oncology	EGFR	Indications and Usage, Warnings and Precautions, Clinical Pharmacology, Clinical Studies
Panitumumab (2)	Oncology	KRAS	Indications and Usage, Clinical Pharmacology, Clinical Studies
Peginterferon alfa-2b	Antivirals	IL28B	Clinical Pharmacology
Prasugrel	Cardiovascular	CYP2C19	Use in Specific Populations, Clinical Pharmacology, Clinical Studies
Propafenone	Cardiovascular	CYP2D6	Clinical Pharmacology
Propranolol	Cardiovascular	CYP2D6	Precautions, Drug Interactions, Clinical Pharmacology
Protriptyline	Psychiatry	CYP2D6	Precautions
Quinidine	Antimalarials/Antiarrhythmics	CYP2D6	Precautions
Rabeprazole	Gastroenterology	CYP2C19	Drug Interactions, Clinical Pharmacology
Rasburicase	Oncology	G6PD	Boxed Warning, Contraindications

Table 1-1. Continued.

Drug	Therapeutic Area	Biomarker	Label Sections
Rifampin, Isoniazid, and Pyrazinamide	Antiinfectives	NAT1; NAT2	Adverse Reactions, Clinical Pharmacology
Risperidone	Psychiatry	CYP2D6	Drug Interactions, Clinical Pharmacology
Sodium Phenylacetate and Sodium Benzoate	Gastroenterology	UCD (NAGS; CPS; ASS; OTC; ASL; ARG)	Indications and Usage, Description, Clinical Pharmacology
Sodium Phenylbutyrate	Gastroenterology	UCD (NAGS; CPS; ASS; OTC; ASL; ARG)	Indications and Usage, Dosage and Administration, Nutritional Management
Tamoxifen	Oncology	ER receptor	Indications and Usage, Precautions, Medication Guide
Telaprevir	Antivirals	IL28B	Clinical Pharmacology
Terbinafine	Antifungals	CYP2D6	Drug Interactions
Tetrabenazine	Neurology	CYP2D6	Dosage and Administration, Warnings, Clinical Pharmacology
Thioguanine	Oncology	TPMT	Dosage and Administration, Precautions, Warnings
Thioridazine	Psychiatry	CYP2D6	Precautions, Warnings, Contraindications
Ticagrelor	Cardiovascular	CYP2C19	Clinical Studies
Timolol	Opththalmology	CYP2D6	Clinical Pharmacology
Tiotropium	Pulmonary	CYP2D6	Clinical Pharmacology
Tolterodine	Reproductive and Urologic	CYP2D6	Clinical Pharmacology, Drug Interactions, Warnings and Precautions
Tositumomab	Oncology	CD20 antigen	Indications and Usage, Clinical Pharmacology
Tramadol and Acetaminophen	Analgesics	CYP2D6	Clinical Pharmacology
Trastuzumab	Oncology	Her2/neu	Indications and Usage, Precautions, Clinical Pharmacology
Tretinoin	Dermatology and Dental	PML/RARα	Boxed Warning, Dosage and Administration, Precautions

Table 1-1. Continued.

Drug	Therapeutic Area	Biomarker	Label Sections
Valproic Acid	Psychiatry	UCD (NAGS; CPS; ASS; OTC; ASL; ARG)	Contraindications, Precautions, Adverse Reactions
Venlafaxine	Psychiatry	CYP2D6	Drug Interactions
Voriconazole	Antifungals	CYP2C19	Clinical Pharmacology, Drug Interactions
Warfarin (1)	Hematology	CYP2C9	Dosage and Administration, Precautions, Clinical Pharmacology
Warfarin (2)	Hematology	VKORC1	Dosage and Administration, Precautions, Clinical Pharmacology

Table 1-2. Example studies investigating natural variation, heritability, processing-induced variation of blood-based gene expression.

Reference	Cell Type	Sample Size	Technology	Comments
Cheung et al. 2003	LCLs	90 subjects	cDNA	Heritability
Schadt et al. 2003	LCLs	56 subjects	Affymetrix	Heritability
Whitney et al. 2003	Whole Blood PBMCs	75 subjects	cDNA	Natural Variation
Debey et al. 2004	PBMCs	29 healthy subjects	Affymetrix	Processing-induced
Morley et al. 2004	LCLs	14 CEPH Families	Affymetrix	Heritability Natural Variation
Nicholson et al. 2004	PBMCs	12 subjects	cDNA	Natural Variation
Radich et al. 2004	Leukocytes	32 subjects	cDNA	Natural Variation
Cobb et al. 2005	Whole Blood Leukocytes	23 healthy 34 trauma/burn	Affymetrix	Natural, Heritability, Processing-induced
Eady et al. 2005	PBMCs	18 healthy subjects	cDNA	Natural Variation
Palmer et al. 2006	Leukocytes	7 healthy subjects	cDNA	Cell-type Specific Variation
Kim et al. 2007	Whole Blood	42 subjects	Affymetrix	Natural, Processing-induced
Karlovich et al. 2009	Whole Blood	20 healthy subjects	Affymetrix	Longitudinal
Meaburn et al. 2009	Whole Blood	10 subjects	Affymetrix	Natural, Heritability, Processing-induced
Dumeaux et al. 2010	Whole Blood	286 subjects	Applied Biosystems	Natural Variation
Min et al. 2010	Whole Blood PBMCs LCLs	6 subjects	Illumina	Cell-type Specific, Processing-induced

Table 1-3. Online resources for the mouse (Peter et al. 2007).

Mouse strain information and resources (inbred and mutant)	
International Mouse Strain Resources	http://www.informatics.jax.org/imsr/index.jsp
JAX Mice	http://jaxmice.jax.org/index.html
Federation of International Mouse Resources	http://www.fimre.org
Mouse Mutant Resource	http://www.jax.org/mmr/index.html
Mouse Mutant Regional Resource Centers	http://www.mmrrc.org
Riken Bioresource Center	http://www.brc.riken.jp/lab/animal/en
The European Mouse Archive	http://www.emmanet.org
Mouse Models of Human Cancer Consortium	http://http://mouse.ncifcrf.gov
Canadian Mouse Mutant Repository	http://www.cmmr.ca/index.html

Knockout and transgenic mice	
International Gene Trap Consortium*	http://www.genetrap.org
Mouse Genome Infomatics Deltagen/Lexicon	http://www.informatics.jax.org
Induced Mutant Resource, Jackson Laboratory	http://www.jax.org/imr/index.html
Samuel Lunenfeld Research Institute‡	http://www.mshri.on.ca/nagy
Mouse Mutant Regional Resource Centers	http://www.mmrrc.org
Micer	http://www.sanger.ac.uk/PostGenomics/mousegenomics

Sequence/phenotype databases	
Ensembl	http://www.ensembl.org/Mus_musculus/index.html
Map Viewer at NCBI	http://www.ncbi.nlm.nih.gov/mapview
Genome Browser, UCSC	http://genome.ucsc.edu/cgi-bin/hgGateway
Mouse Genome Informatics Database	http://www.informatics.jax.org
Vertebrate Genome Annotation	http://vega.sanger.ac.uk/index.html
Panther	http://pantherdb.org
Mouse Phenome Database	http://www.jax.org/phenome
Eumorphia	http://www.eumorphia.org
Mouse Tumor Biology Database	http://tumor.informatics.jax.org
German Mouse Clinic	http://www.gsf.de/ieg/gmc

Pathways analysis	
Ingenuity	http://www.ingenuity.com
GenMAPP	http://www.genmapp.org
KEGG Pathway Database	http://www.genome.jp/kegg/pathway.html

SNP databases	
Roche	http://mousesnp.roche.com
GNF	http://snp.gnf.org

Table 1-3. Continued.

SNP databases (continued)	
NCBI	http://www.ncbi.nlm.nih.gov/SNP
Mouse Phenome Database	http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=snps/door
Mouse Genome Informatics Database	http://www.informatics.jax.org/menus/strain_menu.shtml
Perlegen	http://mouse.perlegen.com/mouse
Wellcome Trust Centre for Human Genetics	http://www.well.ox.ac.uk/mouse/INBREDS
Broad Institute	http://www.broad.mit.edu/personal/claire/MouseHapMap/Inbred.htm
Expression databases	
GNF SymAtlas	http://symatlas.gnf.org
Institute for Genomic Research	http://pga.tigr.org
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo
The Jackson Laboratory	http://www.informatics.jax.org/menus/expression_menu.shtml
Brain Atlas	http://www.brainatlas.org
GenSat	http://www.gensat.org/index.html
EMAGE	http://genex.hgu.mrc.ac.uk/Emage/database/emageIntro.html
Comparative genomics	
VISTA	http://genome.lbl.gov/vista/index.shtml
Mouse Genome Informatics Database	http://www.informatics.jax.org/menus/homology_menu.shtml
Rat Genome Database	http://www.rgd.mcw.edu/VCMAP/mapview.shtml
Quantitative traits analysis	
The Jackson Laboratory, Churchill Laboratory	http://www.jax.org/staff/churchill/labsite
R/qtl	http://www.biostat.jhsph.edu/~kbroman/qtl
Web/QTL	http://www.genenetwork.org/home.html
The Jackson Laboratory PGA	http://pga.jax.org/resources/index.html
The Complex Trait Consortium	http:://www.complextrait.org

* Members are: BayGenomics (USA), Centre for Modelling Human Disease (Toronto, Canada), Embryonic Stem Cell Database (University of Manitoba, Canada), Exchangeable Gene Trap Clones (Kumamoto University, Japan), German Gene Trap Consortium (Germany), Sanger Institute Gene Trap Resource (Cambridge, UK), Soriano Lab Gene Trap Database (Fred Hutchinson Cancer Research Center, Seattle, USA), TIGEM-IRBM Gene Trap (Naples, Italy). ‡ Database of *Cre*-expressing strains. GNF, Genomics Institute of the Novartis Research Foundation; NCBI, National Center for Biotechnology information; UCSC, University of California at Santa Cruz.

FIGURES

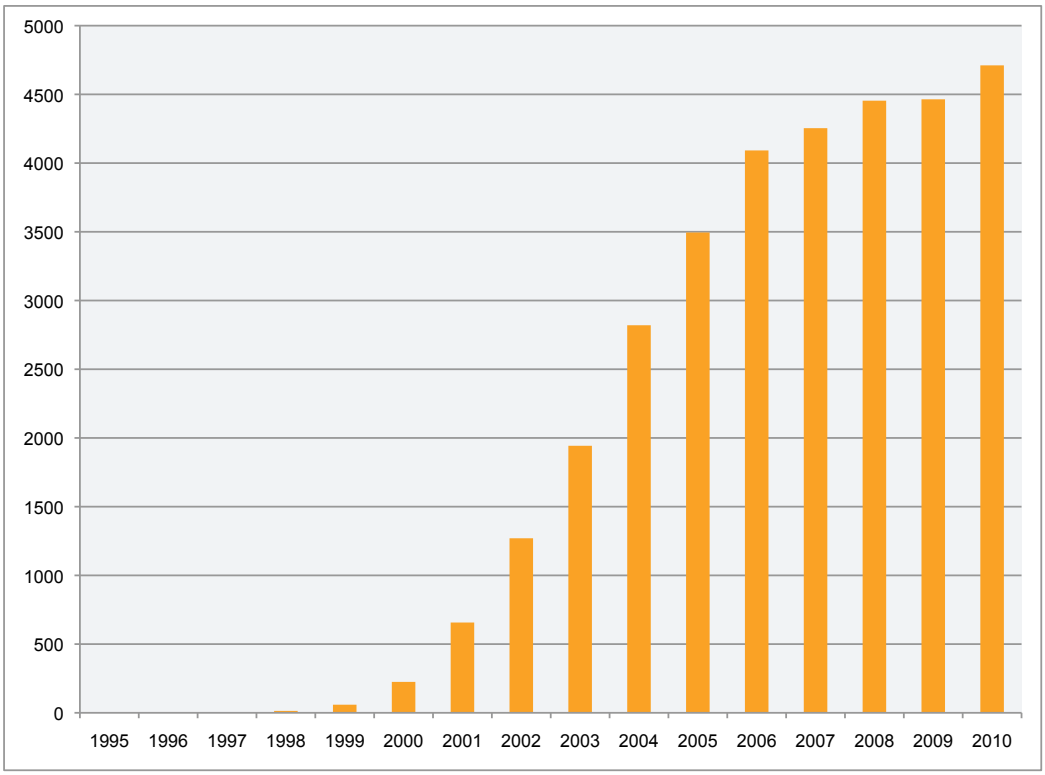

Figure 1-1. Number of microarray publications per year from 1995 to 2010 based on the PubMed Search, “microarray AND gene AND expression”.

REFERENCES

Accili D (2004) A note of caution on the Knockout Mouse Project. Nature Genetics 36, 1132-1132

Akritas MG, Bershady MA (1996) Linear Regression for Astronomical Data with Measurement Errors and Intrinsic Scatter. The Astrophysical Journal 470, 706-706

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403, 503-511

Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science (New York, N.Y.) 322, 881-888

April C, Klotzle B, Royce T, Wickham-Garcia E, Boyaniwsky T, Izzo J, Cox D, Jones W, Rubio R, Holton K, Matulonis U, Quackenbush J, Fan J-B (2009) Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. PLoS ONE 4, e8162-e8162

Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, Oates JA, Peck CC, Schooley RT, Spilker BA, Woodcock J, Zeger SL (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework*. Clin Pharmacol Ther 69, 89-95

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, Grieder FB, Heintz N, Hicks G, Insel TR, Joyner A, Koller BH, Lloyd KCK, Magnuson T, Moore MW, Nagy A, Pollock JD, Roses AD, Sands AT, Seed B, Skarnes WC, Snoddy J, Soriano P, Stewart DJ, Stewart F, Stillman B, Varmus H, Varticovski L, Verma IM, Vogt TF, von Melchner H, Witkowski J, Woychik RP, Wurst W, Yancopoulos GD, Young SG, Zambrowicz B (2004) The knockout mouse project. Nature Genetics 36, 921-924

Belzung C, Griebel G (2001) Measuring normal and pathological anxiety-like behaviour in mice: a review. Behavioural Brain Research 125, 141-149

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SCJ, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P,

Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung W-K, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei C-L, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CWH, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JNS, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PIW, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VVB, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816

Braff DL, Geyer MA (1990) Sensorimotor gating and schizophrenia. Human and animal model studies. Archives of General Psychiatry 47, 181-188

Bućan M, Abel T (2002) The mouse: genetics meets behaviour. Nature Reviews. Genetics 3, 114-123

Cáceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proceedings of the National Academy of Sciences of the United States of America 100, 13030-13035

Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R, Horvath S (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11, 589-589

Carter TA, Del Rio JA, Greenhall JA, Latronica ML, Lockhart DJ, Barlow C (2001) Chipping away at complex behavior: transcriptome/phenotype correlations in the mouse brain. Physiology & Behavior 73, 849-857

Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen K-Y, Morley M, Spielman RS (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genetics 33, 422-425

Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, Herndon DN, Lowry SF, Maier RV, Schoenfeld DA, Moldawer LL, Davis RW, Tompkins RG, Bankey P, Billiar T, Camp D, Chaudry I, Freeman B, Gamelli R, Gibran N, Harbrecht B, Heagy W, Heimbach D, Horton J, Hunt J, Lederer J, Mannick J, McKinley B, Minei J, Moore E, Moore F, Munford R, Nathens A, O'Keefe G, Purdue G, Rahme L, Remick D, Sailors M, Shapiro M, Silver G, Smith R, Stephanopoulos G, Stormo G, Toner M, Warren S, West M, Wolfe S, Young V (2005) Application of genome-wide expression analysis to human health and disease. Proc Natl Acad Sci U S A 102, 4801-4806

Crawley J, Goodwin FK (1980) Preliminary report of a simple animal behavior model for the anxiolytic effects of benzodiazepines. Pharmacology, Biochemistry, and Behavior 13, 167-170

Cryan JF, Mombereau C (2004) In search of a depressed mouse: utility of models for studying depression-related behavior in genetically modified mice. Molecular Psychiatry 9, 326-357

Davies MN, Lawn S, Whatley S, Fernandes C, Williams RW, Schalkwyk LC (2009) To What Extent is Blood a Reasonable Surrogate for Brain in Gene Expression Studies: Estimation from Mouse Hippocampus and Spleen. Front Neurosci 3

de Jong S, Fuller TF, Janson E, Strengman E, Horvath S, Kas MJH, Ophoff RA (2010) Gene expression profiling in C57BL/6J and A/J mouse inbred strains reveals gene networks specific for brain regions independent of genetic background. BMC Genomics 11, 20-20

Debey S, Schoenbeck U, Hellmich M, Gathof BS, Pillai R, Zander T, Schultze JL (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, Schultze JL (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

Donner J, Pirkola S, Silander K, Kananen L, Terwilliger JD, Lönnqvist J, Peltonen L, Hovatta I (2008) An Association Analysis of Murine Anxiety Genes in Humans Implicates Novel Candidate Genes for Anxiety Disorders. Biological Psychiatry 64, 672-680

Dumeaux V, Lund E, Børresen-Dale A-L (2008) Comparison of globin RNA processing methods for genome-wide transcriptome analysis from whole blood. Biomarkers in Medicine 2, 11-21

Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Børresen-Dale A-L, Lund E (2010) Deciphering Normal Blood Gene Expression Variation, The NOWAC Postgenome Study. PLoS Genet 6

Dyrskjøt L, Thykjaer T, Kruhøffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, Wolf H, Orntoft TF (2003) Identifying distinct classes of bladder carcinoma using microarrays. Nature Genetics 33, 90-96

Eady JJ, Wortley GM, Wormstone YM, Hughes JC, Astley SnB, Foxall RJ, Doleman JF, Elliott RM (2005) Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. Physiological Genomics 22, 402-411

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

Feero WG, Guttmacher AE, Collins FS (2008) The Genome Gets Personal, Almost. JAMA: The Journal of the American Medical Association 299, 1351-1352

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

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JT-F, Galver LM, Fan J-B, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier J-F, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok P-Y, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui L-C, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R,

Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-861

Gershon ES, Alliey-Rodriguez N, Liu C (2011) After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. The American Journal of Psychiatry 168, 253-256

Geyer MA, Swerdlow NR, Mansbach RS, Braff DL (1990) Startle response models of sensorimotor gating and habituation deficits in schizophrenia. Brain Research Bulletin 25, 485-498

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531-537

Greenhall JA, Zapala MA, Caceres M, Libiger O, Barlow C, Schork NJ, Lockhart DJ (2007) Detecting genetic variation in microarray expression data. Genome Res 17, 1228-1235

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 U S A 100, 9608-9613

Hovatta I, Tennant RS, Helton R, Marr RA, Singer O, Redwine JM, Ellison JA, Schadt EE, Verma IM, Lockhart DJ, Barlow C (2005) Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice. Nature 438, 662-666

Hovatta I, Zapala MA, Broide RS, Schadt EE, Libiger O, Schork NJ, Lockhart DJ, Barlow C (2007) DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. Genome Biology 8, R25-R25

Hovatta I, Barlow C (2008) Molecular genetics of anxiety in mice and men. Annals of Medicine 40, 92-109

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437, 1299-1320

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431, 931-945

Jasinska AJ, Service S, Choi O-w, DeYoung J, Grujic O, Kong S-y, Jorgensen MJ, Bailey J, Breidenthal S, Fairbanks LA, Woods RP, Jentsch JD, Freimer NB (2009) Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits. Human Molecular Genetics 18, 4415-4427

Karlovich C, Duchateau-Nguyen G, Johnson A, McLoughlin P, Navarro M, Fleurbaey C, Steiner L, Tessier M, Nguyen T, Wilhelm-Seiler M, Caulfield J (2009) A longitudinal study of gene expression in healthy individuals. BMC Medical Genomics 2, 33-33

Kim SJ, Dix DJ, Thompson KE, Murrell RN, Schmid JE, Gallagher JE, Rockett JC (2007) Effects of Storage, RNA Extraction, Genechip Type, and Donor Sex on Gene Expression Profiling of Human Whole Blood. Clin Chem 53, 1038-1045

Letts VA, Schork NJ, Copp AJ, Bernfield M, Frankel WN (1995) A curly-tail modifier locus, mct1, on mouse chromosome 17. Genomics 29, 719-724

Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA (2006) The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. The Journal of Laboratory and Clinical Medicine 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

Meaburn EL, Fernandes C, Craig IW, Plomin R, Schalkwyk LC (2009) Assessing individual differences in genome-wide gene expression in human whole blood: reliability over four hours and stability over 10 months. Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies 12, 372-380

Min J, Barrett A, Watts T, Pettersson F, Lockstone H, Lindgren C, Taylor J, Allen M, Zondervan K, McCarthy M (2010) Variability of gene expression profiles in human blood and lymphoblastoid cell lines. BMC Genomics 11, 96-96

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

Mohr S, Liew C-C (2007) The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends in Molecular Medicine 13, 422-432

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430, 743-747

Moy SS, Nadler JJ, Perez A, Barbaro RP, Johns JM, Magnuson TR, Piven J, Crawley JN (2004) Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. Genes, Brain, and Behavior 3, 287-302

Moy SS, Nadler JJ, Poe MD, Nonneman RJ, Young NB, Koller BH, Crawley JN, Duncan GE, Bodfish JW (2008) Development of a mouse test for repetitive, restricted behaviors: relevance to autism. Behavioural Brain Research 188, 178-194

Nadler JJ, Zou F, Huang H, Moy SS, Lauder J, Crawley JN, Threadgill DW, Wright FA, Magnuson TR (2006) Large-scale gene expression differences across brain regions and inbred strains correlate with a behavioral phenotype. Genetics 174, 1229-1236

Nicholson AC, Unger ER, Mangalathu R, Ojaniemi H, Vernon SD (2004) Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells. Brain Research. Molecular Brain Research 129, 193-197

Palmer C, Diehn M, Alizadeh AA, Brown PO (2006) Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 7, 115-115

Peirce JL, Li H, Wang J, Manly KF, Hitzemann RJ, Belknap JK, Rosen GD, Goodwin S, Sutter TR, Williams RW, Lu L (2006) How replicable are mRNA expression QTL? Mammalian Genome: Official Journal of the International Mammalian Genome Society 17, 643-656

Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL (2007) The mouse as a model for human biology: a resource guide for complex trait analysis. Nature Reviews. Genetics 8, 58-69

Pollak DD, Rey CE, Monje FJ (2010) Rodent models in depression research: classical strategies and new directions. Annals of Medicine 42, 252-264

Radcliffe RA, Lee MJ, Williams RW (2006) Prediction of cis-QTLs in a pair of inbred mouse strains with the use of expression and haplotype data from public databases. Mammalian Genome: Official Journal of the International Mammalian Genome Society 17, 629-642

Radich JP, Mao M, Stepaniants S, Biery M, Castle J, Ward T, Schimmack G, Kobayashi S, Carleton M, Lampe J, Linsley PS (2004) Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics 83, 980-988

Raghavachari N, Xu X, Munson PJ, Gladwin MT (2009) Characterization of whole blood gene expression profiles as a sequel to globin mRNA reduction in patients with sickle cell disease. PLoS ONE 4, e6484-e6484

Rivera J, Tessarollo L (2008) Genetic background and the dilemma of translating mouse studies to humans. Immunity 28, 1-4

Rosenthal N, Brown S (2007) The mouse ascending: perspectives for human-disease models. Nature Cell Biology 9, 993-999

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422, 297-302

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

Taylor TN, Greene JG, Miller GW (2010) Behavioral phenotyping of mouse models of Parkinson's disease. Behavioural Brain Research 211, 1-10

Tian Z, Palmer N, Schmid P, Yao H, Galdzicki M, Berger B, Wu E, Kohane IS (2009) A Practical Platform for Blood Biomarker Study by Using Global Gene Expression Profiling of Peripheral Whole Blood. PLoS ONE 4, e5157-e5157

Topol EJ, Murray SS, Frazer KA (2007) The Genomics Gold Rush. JAMA: The Journal of the American Medical Association 298, 218-221

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530-536

Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA (2009) Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. BMC Genomics 10, 2

Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM (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 U S A 98, 1176-1181

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO (2003) Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 100, 1896-1901

Zapala MA, Schork NJ (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proceedings of the National Academy of Sciences of the United States of America 103, 19430-19435

CHAPTER 2

The Effects of Globin on Microarray-based Gene Expression Analysis of Mouse Blood

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 could potentially 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 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. 2008; 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 non-blood primary tissues (Liew et al. 2006; Fan and Hegde 2005; 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 non-blood 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 inter-individual variation introduced by gender, age, time of day, genetic variation, and environment can be reduced in such studies (Leonardson et al. 2010; Cobb et al. 2005; 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; Debey et al. 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™, a commercially available globin reduction protocol, has been shown to improve gene expression detection sensitivity, remove upwards of 95% of alpha- and beta-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™ is also advantageous for mouse studies in that it has been specifically developed for the mouse (Whitley et al. 2007).

In this paper 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, USA) (129S1/SvImJ, A/J, C57BL/6J, C3H/HeJ, DBA/2J, 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 collected via cardiac puncture. The blood was transferred to an EDTA tube and then Trizol LS reagent immediately added (3:1 Trizol:blood) creating a solution in a 15 ml tube that was stored at -80°C for no more than two weeks.

Processing and Globin Reduction

The extraction of total RNA from the blood was performed using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Total RNA quantity and quality was assessed by spectrophotometer and the Agilent nano RNA chip. Alpha and beta globin mRNA were reduced from a portion of the total RNA samples using the GLOBINclear™ Mouse/Rat kit (Ambion, Austin, TX, USA) 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, USA) 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 was 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). 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 ± 11.3 percent 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 Endenberg 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) (See Supplementary Materials Table 2-S1 and 2-S2) and imported into MetaCore (http://www.genego.com) for enrichment analysis in GeneGO Pathway Maps, GeneGO Diseases (by Biomarkers), 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 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 percent 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 (See Supplementary Materials Table 2-S3) 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.

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 12411 ± 1904 compared to 5840 ± 944 in untreated samples (**Figure 2-1**), 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 and globin-reduced RNA. Greater than 1/3 of all probe sets present only in globin-reduced RNA were ranked among the bottom 25 percent of all detectable probe sets, while less than 4 percent were among the top 25 percent (**Table 2-1**). In contrast, 2/3 of the probe sets detected in both whole blood and globin-reduced blood RNA were among the highest 25 percent. Again, this suggests 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 as compared to whole mouse blood, probe sets detected as present across all samples in whole blood RNA (n=1791; See Supplementary Materials Table 2-S1) and globin-reduced RNA (n=5383; See Supplementary Materials Table 2-S2) 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 (**Figure 2-2**). 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 (See Supplementary Materials Figure 2-S1 and 2-S2) 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 rather by virtue of their globin-relative signal strength detectable via chip-based multi-probe 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 2-2**, See Supplementary Materials Table 2-S4 & 2-S5). 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 by 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 non-hematologic based diseases including neuromuscular diseases, neurodegenerative diseases and chromosome aberrations (**Table 2-3**).

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 2-4**). 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 (**Figure 2-3**) 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 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. 2007; Whitley et al. 2005), 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 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 (**Figure 2-2**), 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 or 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 non-specific 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.

ACKNOWLEDGEMENTS

Dr. Schork and his laboratory are supported in part by the following research grants: The National Institute on Aging Longevity Consortium [grant number U19 AG023122-01]; The NIMH-funded Genetic Association Information Network Study of Bipolar Disorder National [grant number 1 R01 MH078151-01A1]; National Institutes of Health grants: grant numbers N01 MH22005, U01 DA024417-01, P50 MH081755-

01; the Scripps Translational Sciences Institute Clinical Translational Science Award [grant number UL1 RR025774], the Price Foundation and Scripps Genomic Medicine.

Chapter 2, in full, is a reprint of the material as it appears in: Mary E Winn, Matthew A Zapala, Iiris Hovatta, Victoria B Risbrough, Elizabeth Lillie, Nicholas J Schork (2010) The effects of globin on microarray-based gene expression analysis of mouse blood. Mammalian Genome 21, 268-275. The dissertation author was the primary researcher and author of this paper. The original publication is available at http://www.springerlink.com/content/k84136v553l02507/.

TABLES

Table 2-1. Distribution of probe sets detected in: 1.) globin-reduced samples only; 2.) whole blood samples only; or 3.) both whole and globin-reduced samples across all detectable probes (n = 5400). Detectable probes were ranked according to their average, normalized expression intensities in both whole and globin-reduced RNA.

	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

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

	Whole Blood RNA		Globin Reduced RNA	
GeneGO Disease (by Biomarker)	**Network Objects**	**p-value***	**Network Objects**	**p-value***
Hemic and Lymphatic Diseases	251/1927	$5.75x10^{-09}$	597/1927	$1.02x10^{-11}$
Hematologic Diseases	195/1316	$1.21x10^{-11}$	424/1316	$1.04x10^{-10}$
Neoplasms	626/5827	$3.45x10^{-09}$	1571/5827	$4.24x10^{-10}$
Anemia	73/344	$2.24x10^{-11}$	135/344	$2.25x10^{-09}$
Bone Marrow Diseases	114/748	$1.25x10^{-07}$	256/748	$2.60x10^{-09}$
Spherocytosis, Hereditary	9/13	$3.16x10^{-07}$	13/13	$1.50x10^{-08}$
Tay-Sachs Disease	8/11	$8.45x10^{-07}$	11/11	$2.40x10^{-07}$
Gangliosidoses	8/13	$5.53x10^{-06}$	12/13	$5.99x10^{-07}$
Anemia, Hemolytic, Congenital	40/153	$1.81x10^{-09}$	66/153	$6.44x10^{-07}$
Myelodysplastic Syndromes	89/630	$7.19x10^{-05}$	210/630	$7.85x10^{-07}$

*GeneGO hypergeometric model

Table 2-3. Non-hematologic 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. All diseases are significantly enriched with a GeneGO hypergeometric p-value ≤ 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.

GeneGO Disease (by Biomarkers)	Whole Blood Network Objects	Globin-reduced Network Objects	p-value*
Alzheimer 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 Disease	38/283	64/283	9.44×10^{-7}
Schizophrenia	106/752	208/752	$< 2.20 \times 10^{-16}$

*McNemar Test comparing the proportion of network objects identified before and after globin-reduction.

Table 2-4. Data set Characteristic for human-mouse gene set enrichment analysis.

Study	Microarray	Globin Reduction Method	# of Globin-reduced Samples	# of Orthologous Probe Sets	Probe Sets Present in ≥80% of Samples	GeneGO Network Objects	Significant GeneGO Pathways ($p < 0.001$)
Mouse	Affymetrix Mouse 430 2.0	GLOBINclear	18	21287	5428	4134	267
GSE2888	Affymetrix HG U133A	Affymetrix Protocol	14	17042	6048	4996	279
GSE16728	Affymetrix HG U133 Plus 2.0	GLOBINclear	10	17042	9363	7179	296

FIGURES

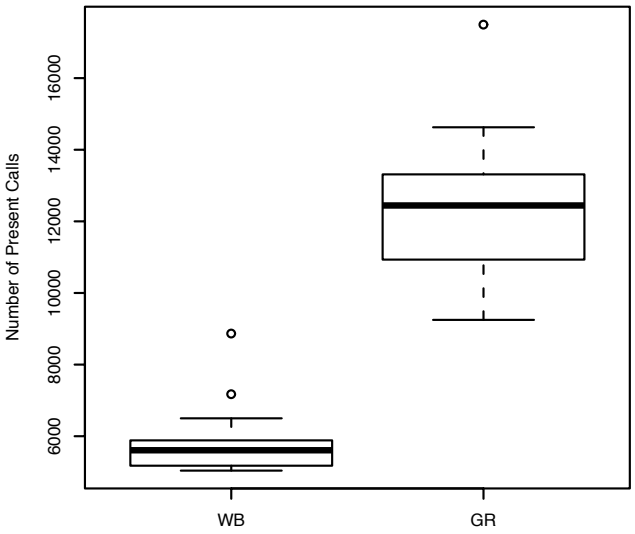

Figure 2-1. Box plots of present calls in whole blood RNA and globin-reduced blood RNA samples. The boxes represent the lower quartile through the upper quartile, while the whiskers extend to 1.5 times the interquartile range. Open circles denote outliers. A bold line denotes the median. WB – whole blood RNA. GR – globin-reduced blood RNA.

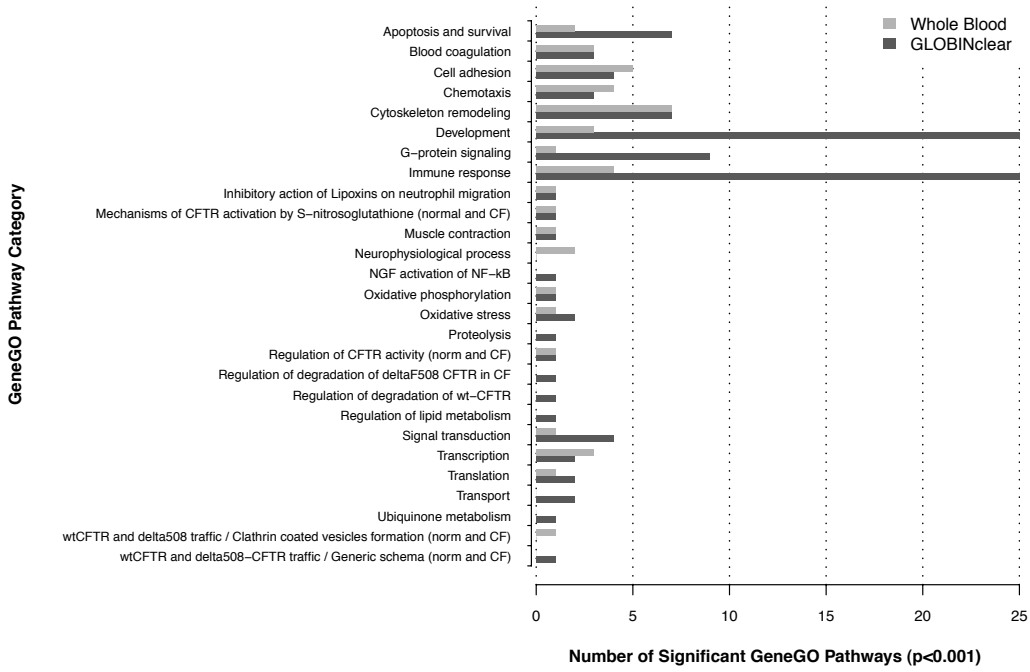

Figure 2-2. Number of significantly enriched GeneGO Pathways Maps at a threshold of $p \leq 0.001$ in mouse whole blood RNA (gray) and globin-reduced RNA (black). The associated false discovery rate is less than 0.01 for a p-value threshold less than or equal to 0.001.

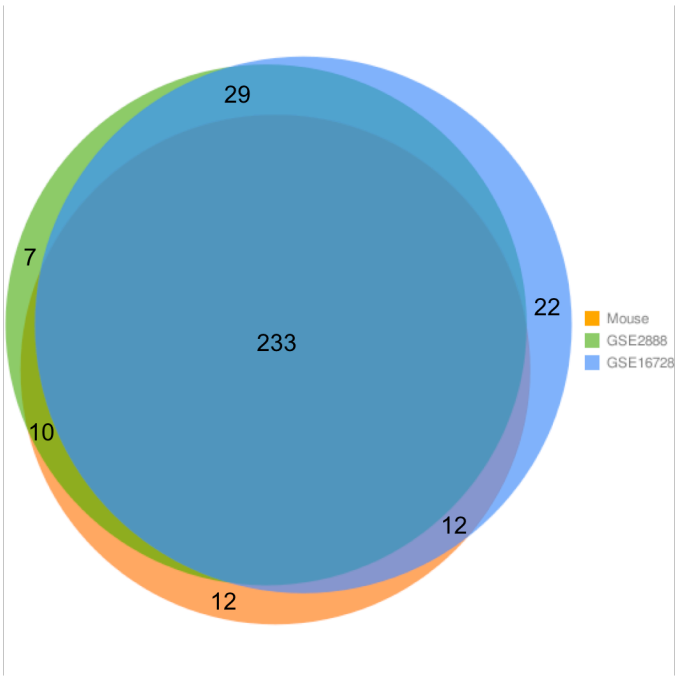

Figure 2-3. Venn diagram comparing significant GeneGO Pathway Maps ($p \leq 0.001$) enriched in globin-reduced mouse RNA and globin-reduced human (GSE2888 and GSE16728) RNA. The associated false discovery rate is approximately 0.01 for a p-value threshold less than or equal to 0.001.

REFERENCES

Affymetrix GeneChip Expression Analysis Data Analysis Fundamentals.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (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, Chen CK, Tsai CS, Chao AS, Lai CH (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, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, Herndon DN, Lowry SF, Maier RV, Schoenfeld DA, Moldawer LL, Davis RW, Tompkins RG, Bankey P, Billiar T, Camp D, Chaudry I, Freeman B, Gamelli R, Gibran N, Harbrecht B, Heagy W, Heimbach D, Horton J, Hunt J, Lederer J, Mannick J, McKinley B, Minei J, Moore E, Moore F, Munford R, Nathens A, O'Keefe G, Purdue G, Rahme L, Remick D, Sailors M, Shapiro M, Silver G, Smith R, Stephanopoulos G, Stormo G, Toner M, Warren S, West M, Wolfe S, Young V (2005) Application of genome-wide expression analysis to human health and disease. Proc Natl Acad Sci U S A 102, 4801-4806

Coppola G, Karydas A, Rademakers R, Wang Q, Baker M, Hutton M, Miller BL, Geschwind DH (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, Zander T, Schultze JL (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, Schultze JL (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, Hamilton-Dutoit S, Wolf H, Orntoft TF (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, Ellsworth RE, Ellsworth DL (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, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80

Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N, Han M, Liew CC, Tsuang MT (2005) Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci U S A 102, 15533-15538

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (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 U S A 100, 9608-9613

Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD, Edenberg HJ, Tsuang MT, Salomon DR, Nurnberger JI, Jr., Niculescu AB (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, Reitman M, Emilsson V, Schadt EE (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 microarray 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

Radich JP, Mao M, Stepaniants S, Biery M, Castle J, Ward T, Schimmack G, Kobayashi S, Carleton M, Lampe J, Linsley PS (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, Montroni I, del Governatore M, Caira A, Taffurelli M, Santini D, Coppola D, Guidotti L, Carinci P, Strippoli P (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, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530-536

Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA (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, Chen JC, Christiani DC (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, Monk BJ, Lockhart DJ, Burger RA, Hampton GM (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 U S A 98, 1176-1181

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

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

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO (2003) Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 100, 1896-1901

Wright C, Bergstrom D, Dai H, Marton M, Morris M, Tokiwa G, Wang Y, Fare T (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. In Affymetrix Technical Note

Zapala MA, Hovatta I, Ellison JA, Wodicka L, Del Rio JA, Tennant R, Tynan W, Broide RS, Helton R, Stoveken BS, Winrow C, Lockhart DJ, Reilly JF, Young WG, Bloom FE, Barlow C (2005) Adult mouse brain gene expression patterns bear an embryologic imprint. Proc Natl Acad Sci U S A 102, 10357-10362

CHAPTER 3

Correlation of Blood and Brain Gene Expression in Inbred Mouse Strains

INTRODUCTION

The development of genomic biomarkers for neural-based diseases is hampered by the inability to collect neural tissue from living, human subjects. In order to overcome this limitation, many microarray-based gene expression studies utilize blood as the tissue source (Alter et al. 2011; Takahashi et al. 2010; Scherzer et al. 2007). Despite the increased use of blood (i.e. whole blood, lymphocytes, PBMCs, leukocytes, lymphoblastoid cell lines), it is unknown to what extent the results from these studies yield viable biomarker candidates for the diagnosis of neural-based disease or whether what blood cell-type or types most accurately reflect neural-based disease associated gene expression profiles. Whole blood has been touted as the ideal tissue source because it requires less technical processing and is believed to more directly reflect true gene expression levels at the time of sample collection (Fan and Hegde 2005). Recent studies also suggest whole blood is more amenable to long term RNA storage than other blood-derived cells such as PBMCs (Debey-Pascher et al. 2011). On the other hand, whole blood is a very heterogeneous tissue composed of a variety of cell types, which may or may not hinder the ability to reliably and/or accurately measure gene expression levels using high throughput microarray techniques. For example, the large number of reticulocytes found in whole blood samples lead to high levels of globin gene expression. The relatively high level of globin gene expression hampers microarray-based gene expression results thus requiring the need for further processing via globin reduction prior to microarray processing (Winn et al. 2010, Field et al. 2007). Furthermore, the cell types found in whole blood may respond to or reflect primary molecular disturbance of disease synergistically or in opposite directions thus mitigating true disturbance in disease-

associated gene expression profiles. Blood gene expression is also highly susceptible to time of day (Sukumaran et al. 2010; Whitney et al. 2003), diet (Leonardson et al. 2010), and other sources of intra- and inter-individual variability. Given the natural and technically induced variation associated with microarray-based gene expression profiling of peripheral whole blood along with its relative popularity in studying neural-based disease, it is necessary that we have a better understanding of its ability to accurately and reliably reflect primary perturbations in neural tissue.

There are few studies aimed at understanding the correlation of disease associated gene expression candidates in blood and disease associated gene expression candidates in brain (Cai et al. 2010; Davies et al. 2009; Jasinska et al. 2009). Davies et al. and Jasinska et al. highlight the potential of blood-based gene expression using mouse and vervet monkey models, respectively, while Cai et al. are less optimistic about the ability of blood to identify neural-based disease associated genes in humans. In an attempt to further our understanding of the overlap between blood and brain, we conducted gene expression studies on blood and 5 refined brain regions from 6 widely used inbred mouse strains. Correlations between variation in gene expression and variation in behavioral phenotypes were pursued and the results from these gene expression studies used to assess the validity of blood as a surrogate for brain tissue and possibly generate candidate genes for human studies.

The ultimate motivation for the choice of mouse strains analyzed in this study is the frequency with which the strains are used and the availability of historical phenotypic information on the strains. Strains were selected that represent phenotypic extremes of the neuropsychiatric phenotypes of interest to our research group (**Table 3-1**) as well as strains that are genealogically closely related but differ

with regard to some key behaviors. These strains include the inbred strains previously studied by us (129S6/SvImJ or 129SvEvTac, A/J, C57BL/6J, C3H/H3J, DBA/2J, and FVB/NJ) and used by others to create the BxD RI strains utilized by WebQTL (C57BL/6J and DBA/2J; Wang et al. 2003), consomic strains by Singer et al. (A/J and C57BL/6J), background strains for most knockout, transgenic, and genetrap models as well as chemically induced mutants (several 129 strains, C57BL/6J, DBA/2J, FVB/NJ, CBA/J), and most of the strains used by the eight-way RI strains of the Collaborative Cross (C57BL/6J, A/J, 129S1/SvImJ, NOD/LtJ, NZO/HILtj, CAST/EiJ, PWD/PhJ, WSB/EiJ) (Churchill et al. 2004).

The anatomical brain regions that regulate various neuropsychiatric traits are fairly well understood, particularly fear and anxiety, even though the molecular mechanisms remain to be determined. Correspondingly, previous studies in our lab show that it is very important to dissect small enough anatomical structures for gene expression analysis of brain-related phenotypes (Zapala et al. 2005, Hovatta et al. 2007, Hovatta et al. 2005). These previously employed brain regions for gene expression profiling include bed nucleus of stria terminalis (BNST), hippocampus, hypothalamus, periaqueductal gray (PAG), and pituitary gland. They were chosen for their roles in several neuropsychiatric phenotypes: prepulse inhibition, fear-potentiated startle, and anxiety.

Prepulse inhibition (PPI) is a measure of sensorimotor gating that is reduced in patients with some neuropsychiatric disorders, such as schizophrenia and obsessive-compulsive disorder (Braff et al. 2001). Neural circuits that regulate PPI have been intensively studied in rats, and it has been shown that regions such as the hippocampus, prefrontal cortex, basolateral amygdala, nucleus accumbens, striatum,

ventral tegmental area, ventral pallidum, globus pallidus, substantia nigra, thalamus, pedunculopontine nucleus, and the colliculi are implicated in the regulation of PPI (reviewed in Swerdlow et al. 2001). In this study we concentrate the gene expression profiling effort on the hippocampus as this brain region is amenable to hand dissection under a dissection microscope and enough total RNA can be obtained for gene expression profiling without the need for amplification.

Brain regions that regulate aspects of anxiety are also fairly well known. Hippocampus was chosen because studies of the behavioral effects of anxiolytic drugs used to treat human anxiety disorders including benzodiazepines, barbiturates, and selective serotonin re-uptake inhibitors suggest that anxiety is related to the septo-hippocampal system (Gray and McNaughton 2000). Using electrical stimulation of the brain it is possible to elicit escape behavior or defensive aggression in animals by stimulating the medial hypothalamus (Panksepp 1982), which controls the autonomic aspects of anxiety as well. BNST has been shown to play a role in anxiety-related processes (Somerville et al. 2010) and known to serve as a relay between the limbic system and the hypothalamic–pituitary–adrenal axis, a key regulator in response to stress (Choi et al. 2007). Fear-potentiated startle, a measure of conditioned fear and a well-studied phenotype in the context of anxiety (Davis et al. 1993), has been shown to be blocked by chemical lesions of the PAG (Fendt et al. 1996).

METHODS

The data sets employed here were previously described (Winn et al. 2010, Hovatta et al. 2007). 10 mice per strain were used for the assessment of the

behavioral phenotypes: anxiety-like behavior (light-dark box and open field test), fear potentiated startle, and prepulse inhibition; 3 mice per strain were used for the assessment of blood gene expression profiles; 2 mice per strain were used for the assessment of neural tissue gene expression profiles. As behavioral testing and brain gene expression were collected earlier (Hovatta et al. 2007, Hovatta et al. 2005), different animals were used for blood gene expression profiling (Winn et al. 2010). This approach is appropriate for several reasons. First, behavioral testing of animals is likely to change their brain gene expression levels. If we first measure the behavior of an animal, the brain and blood gene expression pattern is altered and consequently we cannot reliably measure the baseline gene expression levels. The reciprocal is likely also true due to stress induced by blood collection procedures. Second, a key component of our approach is based on the analysis of multiple inbred strains. Because each animal within an inbred strain is genetically identical, measurements from any individual of the same sex and age handled similarly is meaningful (and could be considered a repeated measure), thus abrogating the need to perform different types of analyses on a single animal and risking altering the baseline gene expression patterns.

Blood 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, USA) (129S1/SvImJ, A/J, C57BL/6J, C3H/HeJ, DBA/2J, 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 collected via cardiac puncture. The blood was transferred to an EDTA tube and then Trizol LS reagent immediately added (3:1 Trizol:blood) creating a solution in a 15 ml tube that was stored at -80°C for no more than two weeks.

Blood Sample Processing and Globin Reduction

The extraction of total RNA from the blood was performed using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Total RNA quantity and quality was assessed by spectrophotometer and the Agilent nano RNA chip. Alpha and beta globin mRNA were reduced from a portion of the total RNA samples using the GLOBINclear™ Mouse/Rat kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions with the recommended start quantity of 10 µg of total RNA.

Brain Sample Collection and RNA Processing

All animal procedures were performed according to protocols approved by the Salk Institute for Biological Studies Institutional Animal Care and Use Committee. Seven-week-old male inbred mice were received from the Jackson Laboratory (Bar Harbor, ME, USA) (A/J, C3H/HeJ, C57BL/6J, DBA/2J, and FVB/NJ) or from Taconic Farms (Germantown, NY, USA) (129S6/SvEvTac) and individually housed for 1 week before dissections were conducted. All brain dissections were done between 11:00 and 17:00 hours on a petri dish filled with ice using a dissection microscope. The dissected brain regions for gene expression analysis included hypothalamus (Hypo), hippocampus (Hippo), pituitary gland (Pit), periaqueductal gray (PAG), and bed

nucleus of the stria terminalis (BNST). Hippocampus samples were directly frozen on dry ice and stored at -80°C. The smaller brain structures were collected in RNA Later buffer (Ambion, Austin, TX, USA) and samples from two to five animals were pooled and stored at -80°C. At least two independent replicate samples for each strain and brain region using independent animals were dissected. If samples were pooled, at least two independent pools were collected. The extraction of total RNA from the tissues was performed using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA), in accordance with the manufacturer's instructions.

Sample Amplification and Microarray Hybridization

Gene expression analysis was performed on all globin-depleted blood samples and brain tissue samples using Mouse 430 2.0 arrays (Affymetrix, Santa Clara, CA, USA) containing 45,101 probe sets. Sample labeling, hybridization, and scanning were performed as previously described (Zapala et al. 2005). For blood samples, three biological replicate samples from independent mice were prepared for each strain for a total of 18 mice. For brain tissue samples, two replicate samples from independent animals were prepared for each strain and each tissue (analysis of BNST for C3H/HeJ was performed in triplicate).

Microarray Data Processing and Analysis

Data processing and analysis was 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). Raw expression values for each set of tissue-specific microarrays were individually GCRMA normalized via the Bioconductor *gcrma* package (Wu et al. 2004). 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 in 2 out 3 blood samples per strain or 2 out 2 brain tissue samples per strain were utilized for analysis (**Table 3-2**).

In order to address hybridization artifacts due to strain-specific differences, GeSNP (http://porifera.ucsd.edu/~cabney/cgi-bin/geSNP.cgi) was used to identify possible sequence differences between strains (Greenhall et al. 2007). Identified probes were removed from the chip description file (cdf) prior to normalization and MAS5 calling.

Statistical Data Analysis

Multivariate distance matrix regression (MDMR) was used to assess the proportion of variance explained by strain, tissue source (brain) or cell counts (blood), and individual mouse (http://polymorphism.scripps.edu/~cabney/cgi-bin/mmr.cgi) (Zapala et al. 2006). MDMR correlation analyses compare phenotype data to gene expression data in order to identify genes whose expression mirrors the phenotypic differences across inbred strains. Blood cell counts were collected from the Mouse Phenome Database (http://phenome.jax.org/) (**Table 3-3**). We used hierarchical clustering, as performed in R, to assess: 1.) the relationship between microarray gene expression profiles across the 6 strains in the 6 tissues, 2.) the relationship between microarray gene expression profiles across genes displaying a strain-specific effect in

blood, and 3.) the expression patterns of genes significantly associated with behavioral phenotypes in brain and one or more of the neural tissues. Gene expression profiles from each tissue were individually assessed by Bivariate Correlated Errors Scatter Analysis (Akritas et al. 1996) to identify genes associated with behavioral phenotypes. Bivariate Correlated Errors Scatter Analysis takes into account error in the dependent and independent variable, in this case error associated with the collection of the microarray gene expression profiles and the behavioral phenotypes. Lastly, genes identified as significant by Bivariate Correlated Errors Scatter Analysis were imported into MetaCore™ Gene Expression and Pathway Analysis (version 5.0, St. Joseph, MI, USA) for gene network analysis.

Behavioral Tests

Each strain was assessed for anxiety and related phenotypes using the light-dark box (LD), open field test (OF), fear potentiated startle (FPS), and prepulse inhibition (PPI). Light-dark box variables studied included: time to emerge, latency to emerge, time in dark, percent in dark, time in light, percent in light, and number of transitions. The control phenotype, distance travelled in dark, was used as a covariate for LD behavioral phenotypes. The control phenotype, distance travelled, was used as a covariate for OF behavioral phenotypes. Covariates for FPS and PPI included context, shock response, and startle. Locomotor activity served as a negative control phenotype. Due to the high correlation between the behavioral phenotypes (i.e. light-dark phenotypes and open field phenotypes) (**Figure 3-1**) and no significant covariates as assessed by linear regression, statistical data analyses were restricted to LD_% time in light, LD_% time in dark, FPS, and PPI.

Open Field/Locomotor Activity

Each mouse was placed in the bottom, left hand corner of their respective enclosure at the start of the test session. The movements of the mice were tracked for 5-60 min, with data being stored in 5-min blocks, respectively (Polytrack digitizer (San Diego Instruments). To analyze the locomotor data, an arbitrary maze was created that consists of a center (20 x 20 cm; or 40 x 40 pixels), 4 corners (10 x 10 cm; or 20 x 20 pixels), and 4 rectangular areas along the walls (20 x 10 cm; or 40 x 20 pixels). Four dependent measures were calculated. To assess the overall amount of locomotor activity, transitions between the 9 regions of the maze were counted. The geometric patterns of locomotor activity was quantified by the spatial scaling exponent, d, as described in detail elsewhere (Paulus and Geyer 1991). Briefly, the spatial scaling exponent, d, quantifies the extent to which a sequence of movements are along a straight line (d=1) or within a circumscribed area (d=2). The time spent in the center (min) was also calculated as the primary variable used to quantify anxiety-like responses. In addition, the mean duration per response was defined as the average time (s) spent in the center during each entry into the center and was also calculated to normalize for locomotor differences (for further details see Dulawa et al 1999; Ralph-Williams et al 2003; Ralph et al 2001).

Startle and Prepulse Inhibition Testing

Animals were always tested according to a pre-determined sequence that counterbalances groups with respect to time of day and stabilimeter chamber. Animals were brought to an adjacent room to the startle testing room 1 h before testing, weighed, and placed in the stabilimeter at the appropriate time. A constant

background noise of 65 dB was provided to avoid uncontrolled variations and to enable comparable tests by other laboratories. For assessing PPI in mice, both prepulse intensity (67 – 81 dB) and interstimulus interval (ISI) between (range from 20-1080 ms) prepulse to pulse onsets were varied (SR-LAB Startle System (SDI). A short block of varied intensity startle stimuli (90, 105, 120 dB) were also included. Prepulse stimuli were 20 ms in duration and startle stimuli were 40 ms in duration. Unless otherwise specified, the ISI was 100 ms (onset-onset). All acoustic stimuli were broadband, thereby avoiding complications due to standing waves associated with sine wave tones of specific frequencies. A typical test session began with a 5-min acclimation period, followed by 5 consecutive blocks of test trials. Blocks 1 and 5 consisted of 120 dB pulse alone trials (5 each). Block 2 consisted of 5 each of 90, 105, and 120 dB startle pulses in a pseudorandom order. Block 3 contained 5 each of 120 dB and 105 dB startle trials, 5 each of 120 dB startle pulse preceded 100 ms by either a 69 or 73 dB prepulse, and 5 each of 105 dB startle pulse preceded 100 ms by either a 69 or 73 dB prepulse. Block 4 contained 6 startle trials (120 dB pulse alone) and 5 each of prepulse (73 dB preceding 120 dB pulse) trials with varied prepulse- pulse onset intervals (ISI; 20, 50, 100, 200, 500 ms). “No stimulus” trials, in which data are recorded without any stimuli were presented between each stimulus trial. No prepulse-alone trials were required, because any potentially detectable responses to the prepulse stimuli were recorded by the system during the ISI period. The intertrial interval (ITI) between stimulus-containing trials ranged from 7-23 s with an average of 15 s. Extensive details can be found elsewhere (Geyer and Dulawa 2003).

Fear Potentiated Startle

Fear potentiated startle was assessed as described previously (Risbrough et al 2009). Separated training and test sessions were used, with a training day followed 24 hrs later by a testing day. All mice were exposed to 5 training and testing days, with the 5th testing day providing the final assessment of FPS. During the training session mice were exposed to 20 CS-US pairings (the CS was a combined 30 sec light and 80 dB 4 KHz tone ending with the 0.4 mA 0.5 s foot shock). During the testing session mice were presented with 24 acoustic startle pulse trials (110 dB startle pulse, 40 ms) with half of the pulses presented at the end of a CS presentation (cue trials) and the other half presented with no CS (no cue trials). %Fear potentiated startle was calculated as [(startle magnitude during Cue trial/Startle Magnitude during No Cue Trial)*100]-100. Details can be found in Risbrough et al. 2009. Data used for gene expression association was from the final test day after the 5th training session.

Home Cage Activity

The activity of the mice in their home cage was measured as a behavioral control phenotype. Increased anxiety-like behavior is associated with lower activity in the brightly lit open-field chamber. However, the activity of the animals in a home cage is not associated with anxiety-like behavior. Different inbred strains differ with regard to home cage activity (Carter et al. 2001), and the strain order for this phenotype is different from the observed anxiety-related behavioral phenotype, and the strain order for PPI (Willott et al. 2003).

RESULTS

Influence of Strain, Cell Count, and Tissue Type on Gene Expression Profiles

To understand the effects of strain, blood cell counts, and tissue type on natural variation in blood and neural tissues, we analyzed blood and brain gene expression levels using MDMR. Each tissue type (blood and neural) was assessed individually due to batch differences in the raw intensity levels unable to be addressed by microarray pre-processing methods. Blood cell counts (**Table 3-4**) or neural tissue (**Table 3-5**) explained the most significant portion of variance (HGB: PVE = 0.317, p-value < 0.0001; PIT: PVE = 0.881, p-value < 0.0001). On average, cell type explained the most variation in blood (0.1798) and neural tissue (0.2474) followed by strain (0.1452, 0.0092) and individual (0.0588, 0.0167). The individuals that explained the most variation in blood (S3, S7, and S17) were most often from the strains that explained the most variation (S3 = 129Sv/ImJ; S7 = C57BL/6J; S17 = FVB). In brain, individuals that explained the most variation were from the two tissues that explained the most variation, pituitary and hippocampus. Overall, there was very little difference between strains within each tissue type as demonstrated by hierarchical clustering analyses (**Figure 3-2**) while the relationship between each strain was not maintained across blood and/or neural tissue.

Next, we performed an analysis of variance (ANOVA) using blood gene expression to identify strain-specific genes and to compare them to the previous brain gene expression results reported by Hovatta et al. 478 probe sets (6.7% of the 7108 probe sets present in blood) displayed a significant ($p < 0.01$, $q < 0.1$) strain-specific effect. To see if strain-specific effects in blood recapitulate known genetic relationships between inbred mouse strains, we constructed a dendrogram of gene

expression relatedness using the 478 strain-specific probe sets (**Figure 3-3**), which was able to partially recapitulate strain relatedness as captured by Hovatta et al. Specifically, a total of 36 strain-specific probe sets overlapped between blood and brain (**Table 3-6**); 338 of the probe sets exhibiting strain-specific effects in blood exhibited region-specific effects in brain. These data suggest that although genetic differences between inbred mouse strains account for a portion of gene expression differences in both blood and brain, blood-specific differences are more likely to capture variation due to 'secondary' perturbations or other sources rather than the primary neural insult.

Gene Expression-Behavioral Phenotype Correlations

We went on to evaluate the overlap between gene expression-behavioral phenotype correlations in blood and brain employing Bivariate Correlated Errors Scatter analysis to identify strain-specific expression patterns associated with anxiety. Bivariate Correlated Error Scatter analysis takes into account the measurement error associated with not only gene expression intensities but also behavioral phenotype testing. As seen in **Figure 3-4** and **Figure 3-5**, there was quite a bit of variance associated with the collection of behavioral phenotypes. Nonetheless, significant ($p<0.05$) anxiety-specific effects were identified in all 6 tissues (**Table 3-7**). None of the probe sets were identified as significant across all 6 tissues. Only 1 probe set (1425858_at) was significant across blood and 3 neural tissues, while 30 other probe sets were significant across blood and 1 or 2 neural tissues (**Table 3-8**). Closer inspection of individual gene expression intensities (**Figure 3-6**) showed differences

between the strains were often slight and not correlated in the same direction in blood and brain.

Finally, pathway-level analyses were pursued to determine the extent to which the genes associated with behavioral phenotypes in blood and brain are associated with known biological pathways and diseases. First, GeneGo Pathway Maps enriched among genes identified in blood were identified and evaluated in each neural tissue (**Table 3-9**, **Table 3-10**, **Table 3-11**, and **Table 3-12**). Many pathways significantly enriched in blood were significantly enriched in neural tissues (% Time in Dark = 6/18; FPS = 13/16; % Time in Light = 12/24; PPI = 8/16). Those pathways identified in both blood and brain included pathways involving apoptosis and survival, cell adhesion, cell cycle, cytoskeleton remodeling, development, g-protein signaling, muscle contraction, regulation of lipid metabolism, signal transduction, transcription, and transport. Second, GeneGo Pathway Maps enriched among genes identified in neural tissues were identified and evaluated in blood, specifically neurophysiological pathways (**Table 3-13**). 19 neurophysiological pathways were significantly associated with one or more behavioral phenotypes and significantly enriched in one or more neural tissues. None of the pathways were significantly enriched in blood and only 5 of those 19 pathways contained 1 gene identified in blood (RHOA; 1437628_s_at). Lastly, GeneGo Diseases (by Biomarker) enriched among genes identified in neural tissues were identified and evaluated in blood (**Table 3-14**). In MetaCore, GeneGo Diseases (by Biomarker) are organized into a hierarchical structure starting with broad disease categories and moving down into specific diseases. The broad category Psychiatry and Psychology and diseases falling under it were assessed here. 25 diseases/disease categories were significantly identified in one or more

behavioral phenotypes and one or more neural tissues. Although not significant, 18, 9, 10, and 11 Psychiatry and Pyschology genes were associated with % Time in Dark, FPS, % Time in Light, and PPI, respectively, in blood. Only two diseases were significantly associated with a behavioral phenotype in blood; *Schizophrenia* and *Schizophrenia and Disorders with Psychotic Features* were associated with fear potentiated startle (p = 0.0068 and 0.0069).

DISCUSSION

The relevance of blood-based gene expression biomarkers depends on the strength of the correlation between gene expression levels in the blood and disease-related phenotypes and/or gene expression levels in the primary tissue involved in the pathogenesis of the disease. Here we assessed whole blood gene expression in regards to anxiety-related phenotypes in mice and evaluated the ability of blood to identify genes and pathways associated with anxiety-related phenotypes in neural tissues. We demonstrate blood gene expression profiles only capture a very small subset of genes associated with inbred mouse strains (1.6%) (Hovatta et al. 2007) and behavioral phenotypes (0.6% on average) in neural tissue, while in regards to neurophysiological pathways there was no overlap. This is likely due to a significant difference in the number of probe sets detected as present between blood and neural tissues (**Table 3-2**). Overall, 5153 genes were detected as present across all 6 tissues, while 11828 were present in all 5 neural tissues. Regardless of these differences, looking at the results from the perspective of blood gene expression profiles, 7.5% of genes associated with inbred mouse strains in blood (36/478) overlapped with neural associated genes, while 33% (% Time in Dark), 81% (FPS),

and 50% (% Time in Light and PPI) of pathways associated with behavioral phenotypes in blood were also significantly identified in one or more neural tissue. These results suggest blood gene expression profiles struggle to effectively capture 'primary' disease perturbations despite the fact many of the genes and pathways identified in blood are relevant to the phenotypes and diseases of interest.

Biologically relevant genes identified in blood and brain include many genes associated with the GABAergic neurotransmitter system such as pyridoxine 5'-phosphate oxidase (*Pnpo*). Significantly associated with PPI in blood and PAG and shown to be associated with schizophrenia in a Japanese population (Song et al. 2007), *Pnpo* is the rate limiting enzyme in vitamin B6 synthesis which in turn plays a key role in serotonin, epinephrine, norepinephrine and gamma-aminobutyric acid (GABA) biosynthesis. Vitamin B6 treatment has shown to upregulate the GABAergic system in mice (Yoo et al. 2011). GABA(A) receptors, the primary target of the psychoactive drug benzodiazepine, are known to be reduced in the hippocampus of patients with panic disorders as well as a mouse models of anxiety (Crestani et al. 1999) while GABA(A) receptor agonists elimination fear potentiated startle in mice (Risbrough et al . 2003). *Gabarap*, associated with % Time in Dark in blood and pituitary, clusters neurotransmitter receptors by mediating interactin with the cytoskeleton. Reduction of *Vps13a* (1440146_at), associated with % Time in Dark in blood and hippocampus, leads to the upregulation of GABA(A) receptors in the mouse hippocampus (Kurano et al. 2006), while *Sept11*, also associated with % Time in Dark but in blood and PAG, is known to play a role in GABAergic synaptic connectivity (Li et al. 2009).

A closer look at the significantly enriched pathways and diseases validates the utility of blood gene expression not directly highlighted by the lack of overlap between blood and brain enriched pathways. For example, although only 7 of 932 network objects in the Schizophrenia disease ontology in the MetaCore database are associated with FPS in blood, these 932 network objects are associated with 52/55 of the GeneGo Pathway Maps associated with behavioral phenotypes in blood.

Despite the potential, or lack there of, to identify biologically relevant, neural-based disease genes and pathways using blood gene expression profiles, this study is not without its limitations. One of the most significant limitations is the fact blood and neural-tissues were not collected at the same time. This is more important from a technical standpoint rather than a biological standpoint (Bryant et al. 2011), as processing each tissue type in batches leads to the inability to correct for differences in microarray intensity and may have led to the significant difference between the number of genes detected as present in blood and the number of genes detected as present in neural tissues. Another limitation is the number of mice and the strains of mice utilized for gene expression profiling (3 per strain) and behavioral testing (10 per strain). Increasing the number of mice screened should reduce variability thus improving the ability to identify small differences in gene expression levels and behavioral phenotypes in such highly related, inbred mouse strains while including more genetically and phenotypically diverse strains enhance the differences between strains.

ACKNOWLEDGEMENTS

Chapter 3, in part, is currently being prepared for submission for publication. Mary E Winn, Matthew A Zapala, Iiris Hovatta, Victoria B Risbrough, Elizabeth Lillie, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

TABLES

Table 3-1. Mouse strains studied and their associated anxiety levels (Mozhui et al. 2010, Hovatta et al. 2005).

Strain	Phenotype
129S1/SvImJ	High Anxiety
A/J	High Anxiety
C3H/HeJ	Intermediate Anxiety
C57BL/6J	Low Anxiety, Low Pre-pulse Inhibition
DBA/2J	High Anxiety, High Pre-pulse Inhibition
FVB/nJ	Low Anxiety

Table 3-2. Probe sets present across each tissue type. Probe sets were considered present in blood if scored as P in 2 out of 3 replicates per strain. Probe sets were considered present in each brain tissue if scored P in 2 out of 2 replicates per strain.

	Blood	BNST	Hippo	Hypo	PAG	Pit
Blood	7108					
BNST	5726	16544				
Hippo	5735	15256	16902			
Hypo	5811	15482	15580	17321		
PAG	5590	14673	14610	15060	15568	
Pituitary	5739	13093	13193	13602	12653	15416

Table 3-3. Blood hematology parameters from the Mouse Phenome Database Graubert 1 dataset collected from 9-week-old male mice.

	129S1/SvImJ	A/J	C57Bl/6J	C3H/HeJ	DBA/2J	FVB/NJ
Total WBC[*]	10.3	5.60	7.57	8.77	7.53	6.69
Neutrophils[*]	3.41	1.07	2.43	2.91	1.58	0.945
Lymphocytes[*]	6.51	4.33	4.89	5.03	5.44	5.54
% Neutrophils	31.4	19.4	33.0	34.2	20.0	13.5
% Lymphocytes	64.5	77.1	63.5	56.6	72.4	83.5
Platelets[*]	492	929	687	702	590	795
Hemoglobin[$]	15.5	13.9	15.7	15.3	14.9	13.9

[*] units/vol $x10^3$
[$] g/dL

Table 3-4. Results from Multivariate Distance Martix Regression (MDMR) analysis using mouse whole blood samples. N: number of samples; NPERMS: number of permutations; SS(TRACE): sum of squares; FSTAT: F-statistic; PVAL: p-value; PVE: portion of variance explained. Predictors of variance included cell counts as collected from the Mouse Phenome Database (purple), strain (blue), and individual mouse (white).

	Strain	N	NPERMS	SS(TRACE)	FSTAT	PVAL	PVE
HGB		18	10000	0.003	7.438	$< 1x10^{-4}$	0.317
129S1/SvImJ		18	10000	0.003	5.987	$< 1x10^{-4}$	0.272
WBC		18	10000	0.002	5.834	$< 1x10^{-4}$	0.267
HCT		18	10000	0.002	5.543	$< 1x10^{-4}$	0.257
S3	129	18	10000	0.002	4.621	0.006	0.224
LYM		18	10000	0.002	4.490	$< 1x10^{-4}$	0.219
NEUT		18	10000	0.002	4.216	$< 1x10^{-4}$	0.209
PLT		18	10000	0.002	4.142	0.0001	0.206
C57BL/6J		18	10000	0.002	4.061	0.0008	0.202
EOS		18	10000	0.002	3.986	0.0001	0.199
BASO		18	10000	0.002	3.735	0.0010	0.189
S7	C57	18	10000	0.002	3.734	0.0620	0.189
MPV		18	10000	0.002	3.323	0.0044	0.172
RBC		18	10000	0.002	3.095	0.0059	0.162
A/J		18	10000	0.001	2.817	0.0190	0.150
FVB/nJ		18	10000	0.001	2.602	0.0349	0.140
RDW		18	10000	0.001	2.420	0.0381	0.131
MONO		18	10000	0.001	2.318	0.0510	0.127
MCHC		18	10000	0.001	2.055	0.0903	0.114
RETIC		18	10000	0.001	1.940	0.1166	0.108
S17	FVB	18	10000	0.001	1.908	0.1325	0.107
MCH		18	10000	0.001	1.802	0.1624	0.101
MCV		18	10000	0.001	1.732	0.1771	0.098
S6	AJ	18	10000	0.001	1.455	0.2092	0.083
S9	C57	18	10000	0.001	1.304	0.2502	0.075
S1	129	18	10000	0.001	1.282	0.2572	0.074
DBA/2J		18	10000	0.001	1.028	0.4802	0.060
C3H/HeJ		18	10000	0.000	0.793	0.6065	0.047
S8	C57	18	10000	0.000	0.743	0.4586	0.044
S5	AJ	18	10000	0.000	0.664	0.4951	0.040
S18	FVB	18	10000	0.000	0.556	0.5922	0.034
S10	C3H	18	10000	0.000	0.516	0.6221	0.031
S2	129	18	10000	0.000	0.485	0.6612	0.029
S14	DBA	18	10000	0.000	0.457	0.6804	0.028
S4	AJ	18	10000	0.000	0.336	0.7955	0.021
S16	FVB	18	10000	0.000	0.299	0.8293	0.018
S13	DBA	18	10000	0.000	0.264	0.8580	0.016
S15	DBA	18	10000	0.000	0.261	0.8541	0.016
S11	C3H	18	10000	0.000	0.259	0.8639	0.016
S12	C3H	18	10000	0.000	0.205	0.8964	0.013

Table 3-5. Results from Multivariate Distance Martix Regression (MDMR) analysis using mouse brain samples. N: number of samples; NPERMS: number of permutations; SS(TRACE): sum of squares; FSTAT: F-statistic; PVAL: p-value; PVE: portion of variance explained. Predictors of variance included brain tissue source (purple), strain (blue), and individual mouse (white).

	Tissue	Strain	NOBS	NPERMS	SS(TRACE)	FSTAT	PVAL	PVE
PIT			61	10000	0.229	434.953	$< 1x10^{-4}$	0.881
HIPPO			61	10000	0.049	13.640	$< 1x10^{-4}$	0.188
PAG			61	10000	0.022	5.425	0.0012	0.084
BNST			61	10000	0.017	4.233	0.0081	0.067
S54	C57	Pit	61	10000	0.017	4.016	0.0188	0.064
S55	C57	Pit	61	10000	0.016	3.945	0.0376	0.063
S60	FVB	Pit	61	10000	0.016	3.866	0.0499	0.061
S61	FVB	Pit	61	10000	0.016	3.828	0.0628	0.061
S52	AJ	Pit	61	10000	0.016	3.793	0.0693	0.060
S58	DBA	Pit	61	10000	0.016	3.776	0.0721	0.060
S56	C3H	Pit	61	10000	0.016	3.762	0.0763	0.060
S53	AJ	Pit	61	10000	0.016	3.749	0.0794	0.060
S59	DBA	Pit	61	10000	0.015	3.739	0.0812	0.060
S50	129	Pit	61	10000	0.015	3.612	0.1228	0.058
S51	129	Pit	61	10000	0.015	3.610	0.1176	0.058
S57	C3H	Pit	61	10000	0.015	3.589	0.1318	0.057
HYPO			61	10000	0.004	1.022	0.4912	0.017
S18	C57	Hippo	61	10000	0.004	0.840	0.2783	0.014
S24	FVB	Hippo	61	10000	0.004	0.814	0.2933	0.014
S25	FVB	Hippo	61	10000	0.003	0.783	0.2981	0.013
S20	C3H	Hippo	61	10000	0.003	0.777	0.3055	0.013
S14	129	Hippo	61	10000	0.003	0.771	0.3045	0.013
S17	AJ	Hippo	61	10000	0.003	0.771	0.3124	0.013
S23	DBA	Hippo	61	10000	0.003	0.760	0.3137	0.013
S19	C57	Hippo	61	10000	0.003	0.758	0.3179	0.013
S16	AJ	Hippo	61	10000	0.003	0.753	0.3223	0.013
S22	DBA	Hippo	61	10000	0.003	0.742	0.3206	0.012
S15	129	Hippo	61	10000	0.003	0.742	0.3136	0.012
S21	C3H	Hippo	61	10000	0.003	0.739	0.3248	0.012
C57BL/6J			61	10000	0.003	0.725	0.5855	0.012
129S6/SvEvTac			61	10000	0.003	0.610	0.612	0.010
A/J			61	10000	0.002	0.527	0.643	0.009
DBA/2J			61	10000	0.002	0.504	0.6538	0.008
FVB/nJ			61	10000	0.002	0.497	0.6458	0.008
C3H/HeJ			61	10000	0.002	0.489	0.6515	0.008
S43	C57	PAG	61	10000	0.002	0.383	0.5726	0.006
S39	129	PAG	61	10000	0.002	0.383	0.5704	0.006
S42	C57	PAG	61	10000	0.002	0.370	0.582	0.006
S38	129	PAG	61	10000	0.002	0.370	0.5908	0.006
S45	C3H	PAG	61	10000	0.002	0.364	0.5881	0.006
S40	AJ	PAG	61	10000	0.002	0.355	0.5956	0.006
S49	FVB	PAG	61	10000	0.002	0.354	0.5927	0.006
S41	AJ	PAG	61	10000	0.002	0.351	0.6099	0.006
S8	C3H	BNST	61	10000	0.002	0.347	0.602	0.006

Table 3-5. Continued.

	Tissue	Strain	NOBS	NPERMS	SS(TRACE)	FSTAT	PVAL	PVE
S44	C3H	PAG	61	10000	0.001	0.327	0.6259	0.006
S47	DBA	PAG	61	10000	0.001	0.303	0.642	0.005
S46	DBA	PAG	61	10000	0.001	0.300	0.6493	0.005
S2	129	BNST	61	10000	0.001	0.295	0.6582	0.005
S48	FVB	PAG	61	10000	0.001	0.287	0.6584	0.005
S6	C57	BNST	61	10000	0.001	0.282	0.6748	0.005
S4	AJ	BNST	61	10000	0.001	0.281	0.6737	0.005
S5	C57	BNST	61	10000	0.001	0.264	0.6897	0.004
S11	DBA	BNST	61	10000	0.001	0.251	0.7003	0.004
S10	DBA	BNST	61	10000	0.001	0.249	0.7091	0.004
S7	C3H	BNST	61	10000	0.001	0.237	0.7169	0.004
S9	C3H	BNST	61	10000	0.001	0.228	0.718	0.004
S3	AJ	BNST	61	10000	0.001	0.227	0.7264	0.004
S1	129	BNST	61	10000	0.001	0.215	0.7384	0.004
S13	FVB	BNST	61	10000	0.001	0.206	0.7424	0.003
S12	FVB	BNST	61	10000	0.001	0.198	0.7558	0.003
S30	C57	Hypo	61	10000	0.000	0.105	0.8357	0.002
S26	129	Hypo	61	10000	0.000	0.101	0.8369	0.002
S27	129	Hypo	61	10000	0.000	0.092	0.8443	0.002
S31	C57	Hypo	61	10000	0.000	0.089	0.8472	0.002
S37	FVB	Hypo	61	10000	0.000	0.083	0.8542	0.001
S33	C3H	Hypo	61	10000	0.000	0.078	0.8516	0.001
S32	C3H	Hypo	61	10000	0.000	0.068	0.8603	0.001
S36	FVB	Hypo	61	10000	0.000	0.067	0.8609	0.001
S35	DBA	Hypo	61	10000	0.000	0.065	0.8659	0.001
S28	AJ	Hypo	61	10000	0.000	0.062	0.8639	0.001
S29	AJ	Hypo	61	10000	0.000	0.049	0.8758	0.001
S34	DBA	Hypo	61	10000	0.000	0.048	0.8744	0.001

Table 3-6. Strain-specific genes identified in blood and neural tissues.

Probe Set	Gene Symbol	Gene Name
1427077_a_at	AP2B1	adaptor-related protein complex 2, beta 1 subunit
1433860_at	C5orf22	chromosome 5 open reading frame 22
1415796_at	DAZAP2	DAZ associated protein 2
1420862_at	DCTN4	dynactin 4 (p62)
1424324_at	ESCO1	establishment of cohesion 1 homolog 1 (S. cerevisiae)
1431020_a_at	FGFR1OP2	FGFR1 oncogene partner 2
1417714_x_at	HBA1	hemoglobin, alpha 1
1417714_x_at	HBA2	hemoglobin, alpha 2
1419964_s_at	HDGF	hepatoma-derived growth factor
1419041_at	ITFG1	integrin alpha FG-GAP repeat containing 1
1455905_at	KIAA0100	KIAA0100
1450740_a_at	MAPRE1	microtubule-associated protein, RP/EB family, member 1
1419909_at	MPHOSPH9	M-phase phosphoprotein 9
1434396_a_at	MYL6	myosin, light chain 6, alkali, smooth muscle and non-muscle
1435914_at	NCOR1	nuclear receptor corepressor 1
1432332_a_at	NUDT19	nudix (nucleoside diphosphate linked moiety X)-type motif 19
1441937_s_at	PINK1	PTEN induced putative kinase 1
1428381_a_at	PPDPF	pancreatic progenitor cell differentiation and proliferation factor homolog (zebrafish)
1442148_at	PSIP1	PC4 and SFRS1 interacting protein 1
1438390_s_at	PTTG1	pituitary tumor-transforming 1
1438069_a_at	RBM5	RNA binding motif protein 5
1434933_at	RC3H1	ring finger and CCCH-type domains 1
1460670_at	RIOK3	RIO kinase 3 (yeast)
1416779_at	SDPR	serum deprivation response
1452439_s_at	SRSF2	serine/arginine-rich splicing factor 2
1419741_at	SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae)
1438963_s_at	TFPT	TCF3 (E2A) fusion partner (in childhood Leukemia)
1452686_s_at	TMEM222	transmembrane protein 222
1417912_at	TMEM93	transmembrane protein 93
1419738_a_at	TPM2	tropomyosin 2 (beta)
1437666_x_at 1420494_x_at	UBC	ubiquitin C
1416156_at	VCL	vinculin
1433748_at	ZDHHC18	zinc finger, DHHC-type containing 18
1449552_at	ZFR	zinc finger RNA binding protein

Table 3-7. Number of significant genes ($p<0.05$) identified by Bivariate Correlated Errors Scatter Analysis. Numbers in () represent the percentage of present probe sets.

Tissue	% Time in Dark	% Time in Light	Fear Potentiated Startle	Pre-Pulse Inhibition
Blood	113 (1.5)	56 (0.8)	35 (0.5)	52 (0.7)
BNST	479 (2.9)	255 (1.5)	281 (1.7)	491 (3.0)
Hippocampus	319 (1.9)	239 (1.4)	264 (1.6)	341 (2.0)
Hypothalamus	408 (2.4)	291 (1.7)	250 (1.4)	263 (1.5)
PAG	333 (2.1)	227 (1.5)	216 (1.4)	290 (1.9)
Pituitary	526 (3.4)	213 (1.4)	505 (3.3)	510 (3.3)

Table 3-8. Genes shared between blood and at least 1 neural tissue (p≤0.05)

Brain Tissue	Probe Set	p-value (Blood)	p-value (Brain)	Gene Symbol	Gene Name
% Time in Dark					
BNST Hippo	1419112_at	0.0404	<0.01 0.05	Nlk	nemo like kinase
BNST Hypo Pit	1425858_at	0.01	<0.01 <0.01 <0.01	Ube2m	ubiquitin-conjugating enzyme E2M
BNST	1453207_at	0.02	0.03		
BNST	1453784_at	0.02	0.02	Ilkap	integrin-linked kinase-associated serine/threonine phosphatase 2C
Hippo	1428836_at	0.01	0.04		
Hippo	1440146_at	<0.01	0.03	Vps13a	vacuolar protein sorting 13A (yeast)
Hypo	1429004_at	0.01	0.03	Phip	pleckstrin homology domain interacting protein
Hypo PAG	1447883_x_at	0.03	0.04 <0.01	Map1lc3a	microtubule-associated protein 1 light chain 3 alpha
PAG	1451259_at	0.05	0.05	Rexo2	REX2, RNA exonuclease 2 homolog (S. cerevisiae)
PAG	1451290_at	<0.01	0.01	Map1lc3a	microtubule-associated protein 1 light chain 3 alpha
PAG	1460626_at	0.02	0.03	Sept11	septin 11
Pit	1416937_at	0.02	0.05	Gabarap	gamma-aminobutyric acid receptor associated protein
Pit	1422807_at	0.04	0.03	Arf5	ADP-ribosylation factor 5
Pit	1425837_a_at	0.01	0.04	Ccrn4l	CCR4 carbon catabolite repression 4-like (S. cerevisiae)
Pit	1428068_at	0.02	0.02	Samm50	sorting and assembly machinery component 50 homolog (S. cerevisiae)
Pit	1442989_at	0.04	0.01		
Pit	1449579_at	0.05	0.05	Shy3yl1	Sh3 domain YSC-like 1
% Time in Light					
BNST	1419112_at	0.03	0.01	Nlk	nemo like kinase
BNST Hypo Pit	1425858_at	0.02	<0.01 <0.01 <0.01	Ube2m	ubiquitin-conjugating enzyme E2M
Hippo	1428181_at	0.05	0.04	Etfb	electron transferring flavoprotein, beta polypeptide
Hippo	1448020_at	0.05	0.02	Rap1a	RAS-related protein-1a
PAG	1447883_x_at	0.03	0.01	Map1lc3a	microtubule-associated protein 1 light chain 3 alpha
PAG	1451290_at	0.02	0.02	Map1lc3a	microtubule-associated protein 1 light chain 3 alpha
Pit	1424598_at	0.02	0.04	Ddx6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
Pit	1453784_at	0.03	<0.01	Ilkap	integrin-linked kinase-associated serine/threonine phosphatase 2C

Table 3-8. Continued.

Brain Tissue	Probe Set	p-value (Blood)	p-value (Brain)	Gene Symbol	Gene Name
Fear Potentiated Startle					
Hypo	1423746_at	0.02	0.04	Txndc5	thioredoxin domain containing 5
PAG	1452077_at	0.03	0.03	Ddx3y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
Pit			0.03		
Pit	1416614_at	0.05	0.03	Eid1	EP300 interacting inhibitor of differentiation 1
Pit	1428534_at	0.05	0.05	Nr2c2ap	nuclear receptor 2C2-associated protein
Pre-pulse Inhibition					
BNST	1425858_at	0.03	<0.01	Ube2m	ubiquitin-conjugating enzyme E2M
Hypo			<0.01		
Pit			<0.01		
Hippo	1440880_at	<0.01	0.05	Mppe1	metallophosphoesterase 1
Hypo	1416034_at	0.05	0.05	Cd24a	nectadrin
PAG	1415793_at	0.02	0.01	Pnpo	pyridoxine 5'-phosphate oxidase
PAG	1415856_at	<0.01	<0.01	Emb	embigin
Pit	1427060_at	0.05	0.01	Mapk3	mitogen-activated protein kinase 3
Pit	1437615_s_at	0.05	0.03	Vps37c	vacuolar protein sorting 37C (yeast)
Pit	1448204_at	0.03	0.05	Sav1	salvador homolog 1 (Drosophila)

Table 3-9. GeneGO Pathway Maps significantly enriched (p<0.05) among genes associated with % Time in Dark in blood.

Maps	Tissue	Rank	pValue	Ratio	
Oxidative phosphorylation	Blood	1	$<1x10^{-3}$	4	105
	BNST	145	NS	3	105
	Hippo	290	NS	1	105
	Hypo	297	NS	1	105
	PAG	232	NS	1	105
	Pit	73	NS	4	105
Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases	Blood	2	0.002	2	23
	BNST	185	NS	1	23
	Pit	190	NS	1	23
Muscle contraction_S1P2 receptor-mediated smooth muscle contraction	Blood	3	0.004	2	30
	BNST	216	NS	1	30
	Hypo	141	NS	1	30
	Pit	228	NS	1	30
Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	Blood	4	0.004	3	111
	BNST	8	0.001	7	111
	Hippo	45	0.044	3	111
	Hypo	10	0.005	5	111
	PAG	233	NS	1	111
	Pit	1	$<1x10^{-3}$	11	111
Development_MAG-dependent inhibition of neurite outgrowth	Blood	5	0.006	2	37
	BNST	11	0.001	4	37
	Hypo	45	0.052	2	37
	PAG	124	NS	1	37
Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity	Blood	6	0.008	2	45
	BNST	293	NS	1	45
	Hypo	223	NS	1	45
	Pit	121	NS	2	45
Cell adhesion_Integrin-mediated cell adhesion and migration	Blood	7	0.009	2	48
	BNST	309	NS	1	48
	Hypo	240	NS	1	48
	Pit	8	0.001	5	48
Pentose phosphate pathway/ Rodent version	Blood	8	0.01	2	51
	BNST	316	NS	1	51
	PAG	185	NS	1	51
	Pit	147	NS	2	51
Pentose phosphate pathway	Blood	9	0.011	2	52
	BNST	321	NS	1	52
	PAG	188	NS	1	52
	Pit	153	NS	2	52

Table 3-9. Continued.

Maps	Tissue	Rank	pValue	Ratio	
Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	Blood	10	0.011	2	53
	BNST	53	0.03	3	53
	Hippo	55	NS	2	53
	PAG	193	NS	1	53
	Pit	22	0.009	4	53
Development_WNT signaling pathway. Part 2	Blood	11	0.011	2	53
	BNST	144	NS	2	53
	Hippo	225	NS	1	53
	Hypo	81	NS	2	53
	Pit	157	NS	2	53
Apoptosis and survival_Endoplasmic reticulum stress response pathway	Blood	12	0.011	2	53
	BNST	324	NS	1	53
	Hippo	222	NS	1	53
	Hypo	80	NS	2	53
	PAG	18	0.015	3	53
	Pit	155	NS	2	53
Blood coagulation_GPVI-dependent platelet activation	Blood	13	0.012	2	55
	BNST	149	NS	2	55
	Hypo	261	NS	1	55
	PAG	199	NS	1	55
Muscle contraction_ACM regulation of smooth muscle contraction	Blood	14	0.012	2	56
	BNST	334	NS	1	56
	Hippo	238	NS	1	56
	Hypo	263	NS	1	56
	Pit	162	NS	2	56
Development_Role of IL-8 in angiogenesis	Blood	15	0.013	2	58
	BNST	9	0.001	5	58
Regulation of lipid metabolism_Alpha-1 adrenergic receptors signaling via arachidonic acid	Blood	16	0.019	2	70
	BNST	361	NS	1	70
	Hypo	278	NS	1	70
	Pit	383	NS	1	70
Blood coagulation_GPCRs in platelet aggregation	Blood	17	0.019	2	71
	BNST	78	NS	3	71
	Hippo	264	NS	1	71
	Hypo	109	NS	2	71
Ubiquinone metabolism	Blood	18	0.021	2	74
	BNST	86	NS	3	74
	Hippo	270	NS	1	74
	Hypo	284	NS	1	74
	PAG	223	NS	1	74
	Pit	94	NS	3	74

Table 3-9. Continued.

Maps	Tissue	Rank	pValue	Ratio	
Immune response _CCR3 signaling in eosinophils	Blood	19	0.023	2	77
	BNST	367	NS	1	77
	Hippo	273	NS	1	77
	PAG	68	NS	2	77
	Pit	391	NS	1	77
HETE and HPETE biosynthesis and metabolism	Blood	20	0.024	2	80
	Pit	392	NS	1	80
Muscle contraction_GPCRs in the regulation of smooth muscle tone	Blood	21	0.026	2	83
	BNST	370	NS	1	83
	Hippo	12	0.021	3	83
	Hypo	41	0.049	3	83
	Pit	110	NS	3	83
Transport_Intracellular cholesterol transport in norm	Blood	22	0.027	2	85
	BNST	113	NS	3	85
	Hippo	276	NS	1	85
	Hypo	289	NS	1	85
Cytoskeleton remodeling_Cytoskeleton remodeling	Blood	23	0.038	2	102
	BNST	4	$<1x10^{-3}$	7	102
	Hippo	99	NS	2	102
	Hypo	68	NS	3	102
	PAG	97	NS	2	102
	Pit	4	$<1x10^{-3}$	8	102

Table 3-10. GeneGO Pathway Maps significantly enriched ($p<0.05$) among genes associated with FPS in blood.

Maps	Tissue	Rank	pValue	Ratio	
Immune response_Role of DAP12 receptors in NK cells	Blood	1	$<1x10^{-3}$	3	54
	BNST	222	NS	1	54
	Hippo	256	NS	2	54
	Pit	118	0.009	4	54
Apoptosis and survival_Endoplasmic reticulum stress response pathway	Blood	2	0.002	2	53
	Hippo	82	0.001	4	53
	Hypo	165	NS	1	53
	PAG	176	NS	1	53
	Pit	116	0.008	4	53
Cell adhesion_Role of CDK5 in cell adhesion	Blood	3	0.012	1	9
	Hypo	5	0.002	2	9
IL-1 beta-dependent CFTR expression	Blood	4	0.02	1	16
	BNST	88	NS	1	16
	Hippo	272	NS	1	16
Cytoskeleton remodeling_CDC42 in cellular processes	Blood	5	0.028	1	22
	Pit	293	NS	1	22
Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases	Blood	6	0.029	1	23
	BNST	101	NS	1	23
	Pit	102	0.005	3	23
Cell adhesion_Endothelial cell contacts by non-junctional mechanisms	Blood	7	0.03	1	24
	BNST	104	NS	1	24
	Hypo	13	0.012	2	24
	Pit	299	NS	1	24
Cytoskeleton remodeling_Neurofilaments	Blood	8	0.032	1	25
	BNST	9	0.001	3	25
	Hippo	301	NS	1	25
	Hypo	15	0.013	2	25
	PAG	10	0.012	2	25
	Pit	304	NS	1	25
Cell adhesion_Cadherin-mediated cell adhesion	Blood	9	0.033	1	26
	Hypo	16	0.014	2	26
Cell adhesion_Endothelial cell contacts by junctional mechanisms	Blood	10	0.033	1	26
	BNST	26	0.016	2	26
	Hippo	306	NS	1	26
	Hypo	1	$<1x10^{-3}$	4	26
	Pit	310	NS	1	26

Table 3-10. Continued.

Maps	Tissue	Rank	pValue	Ratio	
Immune response_Antigen presentation by MHC class I	Blood	11	0.035	1	28
	BNST	28	0.019	2	28
	Hippo	311	NS	1	28
	Hypo	19	0.016	2	28
	Pit	318	NS	1	28
Development_Osteopontin signaling in osteoclasts	Blood	12	0.038	1	30
	BNST	29	0.022	2	30
	Hippo	317	NS	1	30
Cell cycle_Spindle assembly and chromosome separation	Blood	13	0.042	1	33
	BNST	32	0.026	2	33
	Hippo	332	NS	1	33
	Pit	130	0.013	3	33
Cell adhesion_Alpha-4 integrins in cell migration and adhesion	Blood	14	0.043	1	34
	Hippo	201	0.035	2	34
	Hypo	26	0.023	2	34
	Pit	132	0.015	3	34
Development_Role of CDK5 in neuronal development	Blood	15	0.043	1	34
	BNST	137	NS	1	34
	Hippo	339	NS	1	34
	Hypo	94	NS	1	34
	PAG	97	NS	1	34
	Pit	138	0.015	3	34
Cell adhesion_Role of tetraspanins in the integrin-mediated cell adhesion	Blood	16	0.047	1	37
	Hippo	350	NS	1	37
	Hypo	103	NS	1	37
	Pit	358	NS	1	37

Table 3-11. GeneGO Pathway Maps significantly enriched (p<0.05) among genes associated with % Time in Light in blood.

Maps	Tissue	Rank	pValue	Ratio	
Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases	Blood	1	0.001	2	23
	Pit	45	NS	1	23
Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	Blood	2	0.001	3	111
	BNST	35	0.046	3	111
	Hippo	80	0.033	3	111
	Hypo	5	0.010	4	111
	PAG	63	NS	2	111
	Pit	2	$<1x10^{-3}$	6	111
Muscle contraction_S1P2 receptor-mediated smooth muscle contraction	Blood	3	0.002	2	30
	Pit	60	NS	1	30
Development_MAG-dependent inhibition of neurite outgrowth	Blood	4	0.002	2	37
	Hippo	180	NS	1	37
	Hypo	103	NS	1	37
Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity	Blood	5	0.003	2	45
	Pit	25	0.030	2	45
Cell adhesion_Integrin-mediated cell adhesion and migration	Blood	6	0.004	2	48
	Hypo	145	NS	1	48
	Pit	6	0.003	3	48
Pentose phosphate pathway/ Rodent version	Blood	7	0.004	2	51
	BNST	158	NS	1	51
	PAG	154	NS	1	51
Pentose phosphate pathway	Blood	8	0.005	2	52
	BNST	160	NS	1	52
	PAG	158	NS	1	52
Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	Blood	9	0.005	2	53
	BNST	164	NS	1	53
	Hippo	240	NS	1	53
	Pit	27	0.041	2	53
Development_WNT signaling pathway. Part 2	Blood	10	0.005	2	53
	BNST	42	0.056	2	53
	Hippo	243	NS	1	53
	Hypo	26	0.061	2	53
	PAG	29	0.058	2	53
Blood coagulation_GPVI-dependent platelet activation	Blood	11	0.005	2	55
	BNST	44	NS	2	55
	Hippo	249	NS	1	55
	Hypo	168	NS	1	55
	PAG	166	NS	1	55

Table 3-11. Continued.

Maps	Tissue	Rank	pValue	Ratio	
Muscle contraction_ACM regulation of smooth muscle contraction	Blood	12	0.005	2	56
	Pit	30	0.045	2	56
Regulation of lipid metabolism_Insulin regulation of glycogen metabolism	Blood	13	0.005	2	56
	Hippo	96	0.048	2	56
	PAG	167	NS	1	56
	Pit	117	NS	1	56
Blood coagulation_GPCRs in platelet aggregation	Blood	14	0.009	2	71
	BNST	50	NS	2	71
	Hippo	40	0.010	3	71
	Hypo	194	NS	1	71
Immune response _CCR3 signaling in eosinophils	Blood	15	0.010	2	77
	Hippo	286	NS	1	77
	PAG	191	NS	1	77
	Pit	139	NS	1	77
Muscle contraction_GPCRs in the regulation of smooth muscle tone	Blood	16	0.011	2	83
	Hippo	18	0.002	4	83
	Hypo	203	NS	1	83
	Pit	16	0.014	3	83
Cytoskeleton remodeling_Cytoskeleton remodeling	Blood	17	0.017	2	102
	BNST	203	NS	1	102
	Hippo	117	NS	2	102
	Hypo	60	NS	2	102
	PAG	52	NS	2	102
	Pit	1	$<1x10^{-3}$	6	102
Cytoskeleton remodeling_Alpha-1A adrenergic receptor-dependent inhibition of PI3K	Blood	18	0.037	1	19
	Pit	38	NS	1	19
Development_FGF2-dependent induction of EMT	Blood	19	0.039	1	20
	Hippo	112	NS	1	20
	PAG	45	NS	1	20
Cytoskeleton remodeling_Role of Activin A in cytoskeleton remodeling	Blood	20	0.039	1	20
	BNST	61	NS	1	20
Cytoskeleton remodeling_ESR1 action on cytoskeleton remodeling and cell migration	Blood	21	0.043	1	22
	Hippo	115	NS	1	22
	Hypo	49	NS	1	22
Development_S1P4 receptor signaling pathway	Blood	22	0.043	1	22
G-protein signaling_Cross-talk between Ras-family GTPases	Blood	23	0.045	1	23
	BNST	66	NS	1	23
	Hippo	7	$<1x10^{-3}$	3	23
Cytoskeleton remodeling_Role of PDGFs in cell migration	Blood	24	0.047	1	24
	Hippo	123	NS	1	24

Table 3-12. GeneGO Pathway Maps significantly enriched ($p<0.05$) among genes associated with PPI in blood.

Maps	Tissue	Rank	pValue	Ratio	
Development_MAG-dependent inhibition of neurite outgrowth	Blood	1	0.001	2	37
	BNST	111	NS	2	37
	Hippo	202	NS	1	37
Regulation of lipid metabolism_Insulin regulation of glycogen metabolism	Blood	2	0.002	2	56
	BNST	166	NS	2	56
	Hippo	311	NS	1	56
	PAG	245	NS	1	56
	Pit	277	NS	2	56
Cell adhesion_Chemokines and adhesion	Blood	3	0.007	2	100
	BNST	5	$<1x10^{-3}$	8	100
	Hippo	167	NS	2	100
	Hypo	76	NS	2	100
	PAG	60	0.039	3	100
	Pit	40	0.004	6	100
Transport_Rab-9 regulation pathway	Blood	4	0.013	1	10
	BNST	24	0.010	2	10
	PAG	90	NS	1	10
Cytoskeleton remodeling_CDC42 in cellular processes	Blood	5	0.028	1	22
	BNST	194	NS	1	22
	Pit	132	0.043	2	22
Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases	Blood	6	0.029	1	23
	BNST	67	0.049	2	23
	Hippo	131	NS	1	23
	Hypo	79	NS	1	23
	Pit	312	NS	1	23
Transcription_Transcription regulation of aminoacid metabolism	Blood	7	0.032	1	25
	BNST	74	0.057	2	25
	Hippo	30	0.026	2	25
	Hypo	91	NS	1	25
	Pit	49	0.006	3	25
Development_S1P2 and S1P3 receptors in cell proliferation and differentiation	Blood	8	0.033	1	26
	BNST	213	NS	1	26
	Pit	152	NS	2	26
Development_Thrombospondin-1 signaling	Blood	9	0.035	1	28
	BNST	79	NS	2	28
	Hippo	154	NS	1	28
Vitamin B6 metabolism	Blood	10	0.038	1	30
	Hypo	104	NS	1	30
	PAG	136	NS	1	30

Table 3-12. Continued.

Maps	Tissue	Rank	pValue	Ratio	
Development_Slit-Robo signaling	Blood	11	0.038	1	30
	BNST	233	NS	1	30
	Hippo	163	NS	1	30
Cytoskeleton remodeling_Fibronectin-binding integrins in cell motility	Blood	12	0.039	1	31
	BNST	85	NS	2	31
	Hippo	168	NS	1	31
	Hypo	107	NS	1	31
	Pit	67	0.011	3	31
Cell adhesion_Alpha-4 integrins in cell migration and adhesion	Blood	13	0.043	1	34
	BNST	251	NS	1	34
	Hippo	181	NS	1	34
	Hypo	119	NS	1	34
	PAG	152	NS	1	34
	Pit	370	NS	1	34
Cell adhesion_Role of tetraspanins in the integrin-mediated cell adhesion	Blood	14	0.047	1	37
	BNST	110	NS	2	37
	Hippo	201	NS	1	37
	Hypo	17	0.020	2	37
	Pit	389	NS	1	37
Signal transduction_cAMP signaling	Blood	15	0.048	1	38
	BNST	284	NS	1	38
	Hippo	55	0.057	2	38
	Hypo	18	0.021	2	38
	Pit	395	NS	1	38
Blood coagulation_Blood coagulation	Blood	16	0.049	1	39
	BNST	291	NS	1	39

Table 3-13. Neurophysiological Process GeneGo Pathway Maps significantly enriched ($p<0.05$) in neural tissues. Neurophysiological GeneGO Pathway Maps identified in blood are shown (gray), whether or not they were significant.

Maps	Phenotype	Tissue	Rank	pValue	Ratio	
Neurophysiological process_ACM regulation of nerve impulse	Dark	Blood	83	0.131	1	46
	FPS	Hippo	27	$<1x10^{-3}$	5	46
		PAG	27	0.037	2	46
	Light	Blood	67	0.088	1	46
		Pit	26	0.032	2	46
	PPI	Hippo	13	0.011	3	46
		Hypo	36	0.029	2	46
Neurophysiological process_ACM regulation of nerve impulse	PPI	Pit	8	$<1x10^{-3}$	6	46
Neurophysiological process_ACM1 and ACM2 in neuronal membrane polarization	Dark	Blood	68	0.115	1	40
	FPS	Hippo	134	0.005	3	40
	Light	Blood	56	0.077	1	40
Neurophysiological process_Circadian rhythm	Light	BNST	34	0.045	2	47
		PAG	25	0.047	2	47
	PPI	PAG	73	0.048	2	47
Neurophysiological process_Corticoliberin signaling via CRHR1	Dark	Hippo	48	0.049	2	50
	FPS	Hippo	75	0.001	4	50
	Light	Hippo	5	$<1x10^{-3}$	4	50
	PPI	Hypo	42	0.034	2	50
		Pit	51	0.007	4	50
Neurophysiological process_Dopamine D2 receptor transactivation of PDGFR in CNS	FPS	Hippo	177	0.021	2	26
		Pit	113	0.007	3	26
	Light	BNST	13	0.015	2	26
	PPI	BNST	1	$<1x10^{-3}$	6	26
		Hippo	32	0.028	2	26
		Hypo	9	0.010	2	26
Neurophysiological process_EphB receptors in dendritic spine morphogenesis and synaptogenesis	Light	Hippo	52	0.020	2	35
	PPI	Hippo	52	0.049	2	35
		Hypo	15	0.018	2	35
Neurophysiological process_GABA-A receptor life cycle	Dark	Blood	37	0.079	1	27
		Hypo	32	0.029	2	27
	FPS	Hippo	90	0.002	3	27
	Light	Blood	27	0.052	1	27
	PPI	BNST	10	0.001	4	27

Table 3-13. Continued.

Maps	Phenotype	Tissue	Rank	pValue	Ratio	
Neurophysiological process_Glutamate regulation of Dopamine D1A receptor signaling	Dark	Hypo	2	0.001	4	45
	FPS	BNST	55	0.046	2	45
		Pit	52	0.001	5	45
	Light	BNST	7	0.004	3	45
		PAG	20	0.043	2	45
	PPI	BNST	9	0.001	5	45
		PAG	67	0.044	2	45
Neurophysiological process_HTR1A receptor signaling in neuronal cells	FPS	BNST	49	0.040	2	42
	Light	Hippo	68	0.029	2	42
Neurophysiological process_Kappa-type opioid receptor in transmission of nerve impulses	FPS	Hippo	189	0.027	2	30
	Light	BNST	16	0.020	2	30
Neurophysiological process_Long-term depression in cerebellum	FPS	PAG	30	0.042	2	49
	PPI	Pit	124	0.037	3	49
Neurophysiological process_Melatonin signaling	FPS	BNST	50	0.042	2	43
	Light	Hippo	70	0.030	2	43
Neurophysiological process_Netrin-1 in regulation of axon guidance	FPS	Pit	156	0.024	3	41
Neurophysiological process_NMDA-dependent postsynaptic long-term potentiation in CA1 hippocampal neurons	Dark	BNST	34	0.019	4	80
		Hypo	39	0.045	3	80
		PAG	30	0.044	3	80
	FPS	BNST	3	<1x10-3	5	80
		Hippo	135	0.005	4	80
	FPS	PAG	4	0.002	4	80
	Light	Hippo	16	0.002	4	80
	PPI	PAG	5	<1x10-3	5	80
		Pit	115	0.032	4	80
Neurophysiological process_nNOS signaling in neuronal synapses	PPI	PAG	35	0.020	2	29
Neurophysiological process_PGE2-induced pain processing	Dark	Blood	74	0.123	1	43
		Hippo	32	0.037	2	43
	Light	Hippo	20	0.002	3	43
Neurophysiological process_Receptor-mediated axon growth repulsion	Dark	Blood	80	0.128	1	45
		Hippo	33	0.041	2	45
	FPS	BNST	54	0.046	2	45
		PAG	25	0.036	2	45
	Light	Blood	65	0.086	1	45
		Hippo	21	0.003	3	45
	PPI	Blood	20	0.056	1	45

Table 3-13. Continued.

Maps	Phenotype	Tissue	Rank	pValue	Ratio
Neurophysiological process_Thyroliberin in cell hyperpolarization and excitability	Dark	Blood	54	0.101	1 35
	FPS	Hippo	119	0.003	3 35
	Light	Blood	42	0.067	1 35
	PPI	Hippo	49	0.049	2 35
		Hypo	14	0.018	2 35
		Pit	31	0.002	4 35

Table 3-14. GeneGO Diseases (by Biomarkers) significantly enriched ($p<0.05$) in neural tissues. GeneGO Diseases (by Biomarkers) identified in blood are shown (gray), whether or not they were significant.

Disease	Phenotype	Tissue	Rank	p-value	Ratio	
Affective Disorders, Psychotic	Dark	Hippo	2	<1x10-4	21	632
		Hypo	3	<1x10-4	29	632
		PAG	90	0.0095	18	632
	FPS	Blood	455	NS	1	632
		BNST	92	0.0075	17	632
		Hippo	16	<1x10-4	24	632
		Hypo	100	0.0162	13	632
	Light	Hippo	8	<1x10-4	18	632
		Hypo	14	0.0002	20	632
		PAG	104	0.0064	14	632
	PPI	Blood	279	NS	1	632
		BNST	212	0.0166	24	632
		Hippo	8	0.0014	20	632
		Hypo	73	0.0095	15	632
		PAG	71	0.0420	14	632
		Pit	47	0.0004	30	632
Agoraphobia	Dark	Hippo	62	0.0061	2	10
		Hypo	89	0.0130	2	10
	FPS	BNST	94	0.0082	2	10
		Hippo	129	0.0070	2	10
		Hypo	47	0.0047	2	10
		Pit	15	0.0015	3	10
	PPI	Pit	202	0.0235	2	10
		Hippo	54	0.0095	2	10
		Hypo	55	0.0061	2	10
	Light	Hippo	62	0.0039	2	10
		Hypo	88	0.0070	2	10
Anxiety Disorders	Dark	Blood	289	NS	1	174
		Hippo	30	0.0012	8	174
		Hypo	161	0.0366	7	174
	FPS	BNST	206	0.0357	6	174
	Light	Blood	197	NS	1	174
		Hippo	78	0.0066	6	174
		Hypo	92	0.0076	7	174
	PPI	Blood	141	NS	1	174
		Hippo	83	0.0166	7	174
		Hypo	95	0.0190	6	174
		PAG	27	0.0086	7	174
Bipolar Disorder	Dark	Hippo	1	<1x10-4	21	630
		Hypo	7	<1x10-4	27	630
		PAG	89	0.0092	18	630
	FPS	Blood	454	NS	1	630
		BNST	90	0.0072	17	630
		Hippo	14	<1x10-4	24	630
		Hypo	98	0.0158	13	630

Table 3-14. Continued.

Disease	Phenotype	Tissue	Rank	p-value	Ratio	
Bipolar Disorder (continued)	Light	Hippo	7	<1x10-4	18	630
		Hypo	33	0.0014	18	630
		PAG	103	0.0063	14	630
	PPI	Blood	278	NS	1	630
		BNST	207	0.0160	24	630
		Hippo	7	0.0014	20	630
		Hypo	70	0.0092	15	630
		PAG	69	0.0410	14	630
		Pit	46	0.0004	30	630
Borderline Personality Disorder	Dark	Pit	110	0.0247	2	10
	FPS	BNST	97	0.0082	2	10
Depression	FPS	Hippo	274	0.0428	3	60
	Light	Hypo	212	0.0428	3	60
	PPI	Hypo	136	0.0359	3	60
Depression, Postpartum	FPS	BNST	31	0.0006	2	3
Depressive Disorder	Dark	Blood	433	NS	2	751
		Hippo	73	0.0092	17	751
		Hypo	21	0.0009	26	751
	FPS	BNST	46	0.0019	21	751
		Hippo	11	<1x10-4	27	751
		Hypo	87	0.0129	15	751
	Light	Hippo	73	0.0055	15	751
		Hypo	15	0.0003	22	751
	PPI	Blood	304	NS	1	751
		Hippo	90	0.0192	19	751
		Hypo	96	0.0193	16	751
		Pit	79	0.0015	32	751
Depressive Disorder, Major	Dark	Blood	432	NS	2	744
		Hippo	100	0.0178	16	744
		Hypo	27	0.0017	25	744
	FPS	BNST	43	0.0017	21	744
		Hippo	10	<1x10-4	27	744
		Hypo	84	0.0119	15	744
	Light	Hippo	107	0.0119	14	744
		Hypo	23	0.0006	21	744
	PPI	Blood	302	NS	1	744
		Hippo	84	0.0176	19	744
		Hypo	93	0.0178	16	744
		Pit	74	0.0013	32	744
Mental Disorders	Dark	Blood	116	NS	18	2945
		BNST	12	0.0010	82	2945
		Hippo	35	0.0016	52	2945
		Hypo	14	0.0002	76	2945
		PAG	127	0.0245	58	2945
		Pit	164	0.0450	87	2945

Table 3-14. Continued.

Disease	Phenotype	Tissue	Rank	p-value	Ratio	
Mental Disorders (continued)	FPS	Blood	402	NS	9	2945
		BNST	40	0.0014	59	2945
		Hippo	42	0.0005	57	2945
		Hypo	151	0.0292	41	2945
		PAG	39	0.0009	50	2945
		Pit	83	0.0265	86	2945
	Light	Blood	211	NS	10	2945
		Hippo	190	0.0408	37	2945
		Hypo	20	0.0005	57	2945
		PAG	160	0.0162	42	2945
		Pit	34	0.0199	38	2945
	PPI	Blood	42	NS	11	2945
		BNST	28	0.0001	98	2945
		Hippo	53	0.0094	59	2945
		Hypo	97	0.0199	47	2945
		PAG	15	0.0018	56	2945
		Pit	25	0.0001	101	2945
Mood Disorders	Dark	Blood	469	NS	2	982
		Hippo	12	0.0003	25	982
		Hypo	4	<1x10-4	38	982
		PAG	152	0.0309	23	982
	FPS	Blood	461	NS	1	982
		BNST	30	0.0005	27	982
		Hippo	9	<1x10-4	32	982
		Hypo	93	0.0151	18	982
	Light	Hippo	13	0.0001	22	982
		Hypo	6	<1x10-4	29	982
		PAG	205	0.0276	17	982
	PPI	Blood	177	NS	3	982
		BNST	223	0.0177	34	982
		Hippo	97	0.0237	23	982
		Hypo	116	0.0276	19	982
		Pit	82	0.0017	39	982
Neurotic Disorders	PPI	Pit	181	0.0174	4	41
Obsessive-Compulsive Disorder	Dark	Hippo	33	0.0015	5	69
	FPS	BNST	138	0.0159	4	69
Panic Disorder	Dark	Blood	186	NS	1	94
		Hippo	127	0.0271	4	94
		PAG	98	0.0155	5	94
	FPS	BNST	50	0.0020	6	94
	Light	Blood	134	NS	1	94
		Hippo	37	0.0021	5	94
		Hypo	186	0.0338	4	94
	PPI	Blood	83	NS	1	94
		Hippo	20	0.0030	6	94
		Hypo	53	0.0056	5	94
		PAG	6	0.0003	7	94
		Pit	219	0.0274	6	94
Personality Disorders	FPS	BNST	57	0.0044	3	24

Table 3-14. Continued.

Disease	Phenotype	Tissue	Rank	p-value	Ratio	
Psychiatry and Psychology	Dark	Blood	130	NS	18	2980
		BNST	19	0.0014	82	2980
		Hippo	39	0.0020	52	2980
		Hypo	18	0.0003	76	2980
		PAG	140	0.0305	58	2980
	FPS	Blood	404	NS	9	2980
		BNST	47	0.0019	59	2980
		Hippo	46	0.0007	57	2980
		Hypo	124	0.0223	42	2980
		PAG	45	0.0012	50	2980
		Pit	101	0.0346	86	2980
	Light	Blood	216	NS	10	2980
		Hippo	215	0.0478	37	2980
		Hypo	26	0.0007	57	2980
		PAG	171	0.0196	42	2980
		Pit	38	0.0238	38	2980
	PPI	Blood	44	NS	11	2980
		BNST	32	0.0002	98	2980
		Hippo	60	0.0121	59	2980
		Hypo	110	0.0243	47	2980
		PAG	13	0.0014	57	2980
		Pit	23	0.0001	102	2980
Psychoses, Substance-Induced	FPS	BNST	145	0.0182	2	15
Schizophrenia	Dark	Blood	263	NS	5	932
		Hippo	20	0.0008	23	932
		Hypo	1	<1x10-4	38	932
		PAG	10	<1x10-4	32	932
	FPS	Blood	288	0.0068	7	932
		Hippo	38	0.0004	25	932
		PAG	22	0.0001	24	932
	Light	Blood	252	NS	3	932
		Hippo	119	0.0165	16	932
		Hypo	7	<1x10-4	28	932
		PAG	29	<1x10-4	24	932
	PPI	Blood	81	NS	4	932
		BNST	104	0.0028	36	932
		Hippo	47	0.0072	24	932
		Hypo	46	0.0041	21	932
		PAG	1	<1x10-4	29	932
		Pit	286	0.0483	31	932
Schizophrenia and Disorders with Psychotic Features	Dark	Blood	264	NS	5	936
		Hippo	21	0.0008	23	936
		Hypo	2	<1x10-4	38	936
		PAG	11	<1x10-4	32	936
	FPS	Blood	289	0.0069	7	936
		Hippo	39	0.0004	25	936
		PAG	23	0.0001	24	936

Table 3-14. Continued.

Disease	Phenotype	Tissue	Rank	p-value	Ratio	
Schizophrenia and Disorders with Psychotic Features (continued)	Light	Blood	254	NS	3	936
		Hippo	123	0.0171	16	936
		Hypo	8	<1x10-4	28	936
		PAG	30	<1x10-4	24	936
	PPI	Blood	84	NS	4	936
		BNST	108	0.0030	36	936
		Hippo	49	0.0075	24	936
		Hypo	47	0.0043	21	936
		PAG	2	<1x10-4	29	936
Schizophrenia, Paranoid	Dark	PAG	173	0.0452	2	22
		Pit	90	0.0169	3	22
Self-Injurious Behavior	Dark	BNST	158	0.0432	5	94
		Hippo	128	0.0271	4	94
		Hypo	55	0.0067	6	94
		PAG	99	0.0155	5	94
	FPS	BNST	24	0.0003	7	94
		Hippo	134	0.0075	5	94
		PAG	107	0.0217	4	94
		Pit	41	0.0077	7	94
	Light	BNST	91	0.0150	4	94
		Hippo	1	<1x10-4	8	94
	PPI	Pit	131	0.0081	7	94
Stress Disorders, Post-Traumatic	Dark	Pit	137	0.0296	2	11
	FPS	Pit	87	0.0277	2	11
Stress Disorders, Traumatic	Dark	Pit	148	0.0350	2	12
	FPS	Pit	96	0.0328	2	12
Stress, Psychological	Dark	Pit	115	0.0251	1	1
Suicide, Attempted	PPI	Pit	12	<1x10-4	4	8

FIGURES

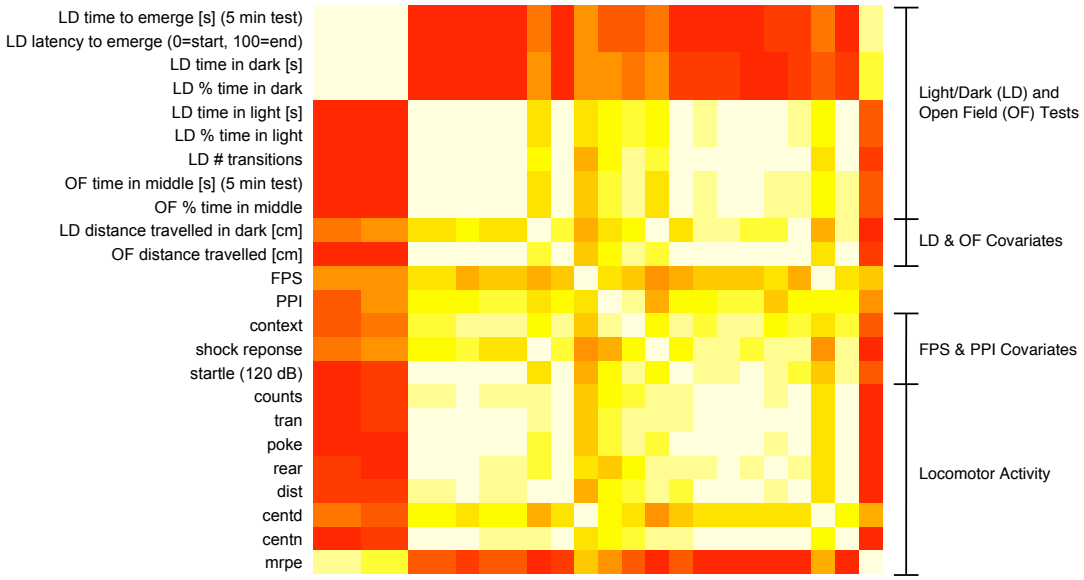

Figure 3-1. Correlations between the behavioral phenotypes. Light yellow represent near perfect positive correlation while bright red represents near perfect negative correlation.

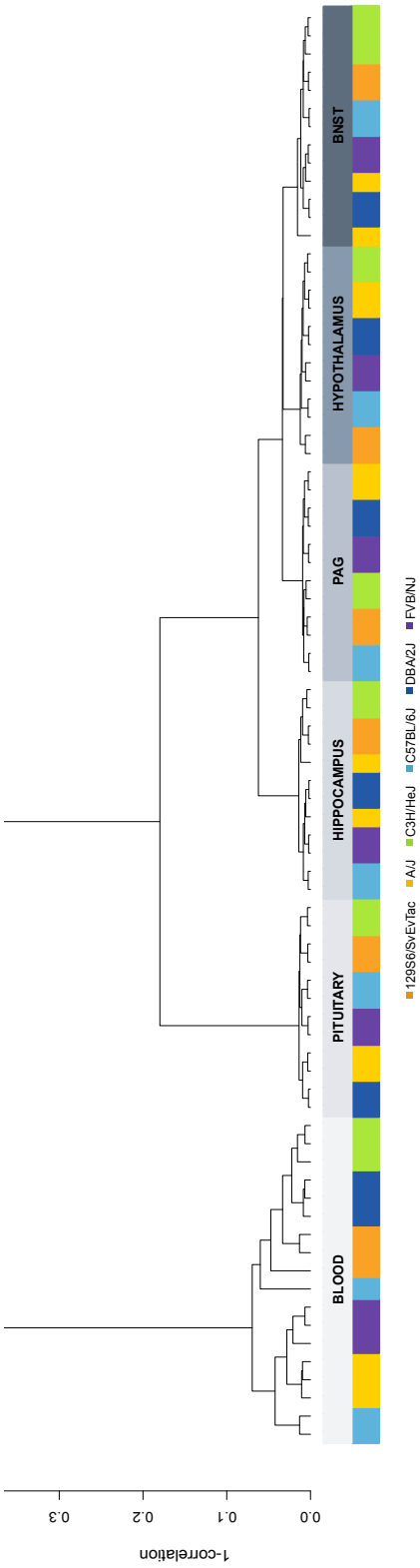

Figure 3-2. Unsupervised hierarchical clustering of gene expression profiles in blood and neural tissues. Gold: 129S1/SvImJ(blood) or 129S6/SvEvTac (neural); Yellow: A/J; Green: C3H/HeJ; Light blue: C57BL/6J; Dark blue: DBA/2J; Purple: FVB/nJ.

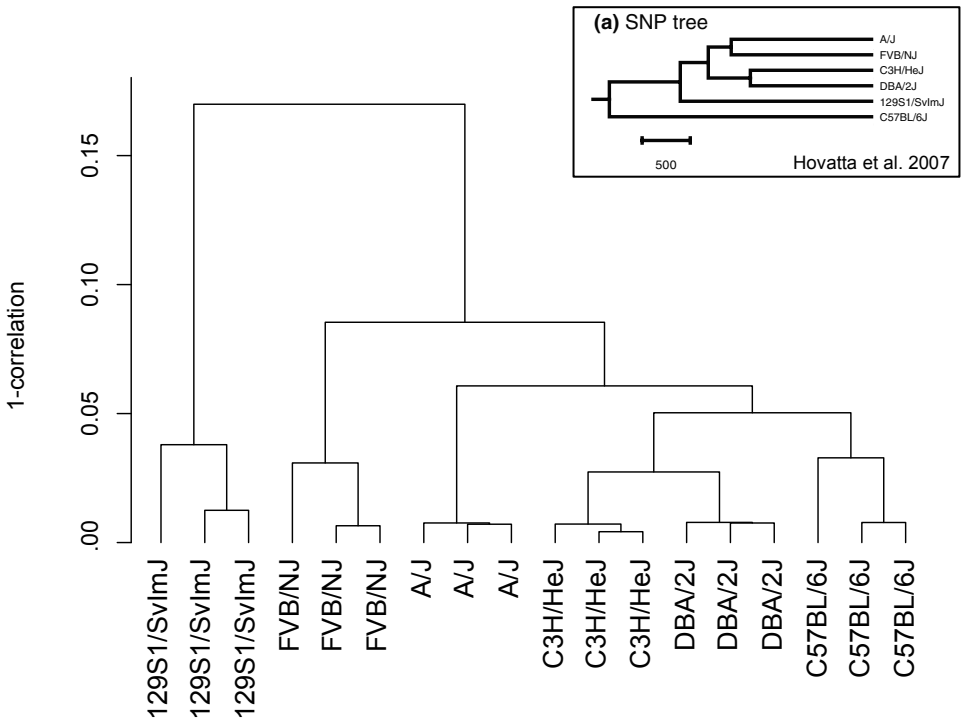

Figure 3-3. Unsupervised hierarchical clustering of strain-specific gene expression profiles in blood (n=478 probe sets exhibiting significant strain-effects). Inset dendrogram from Hovatta et al. 2007 highlighting the strain relationships between inbred mouse strains.

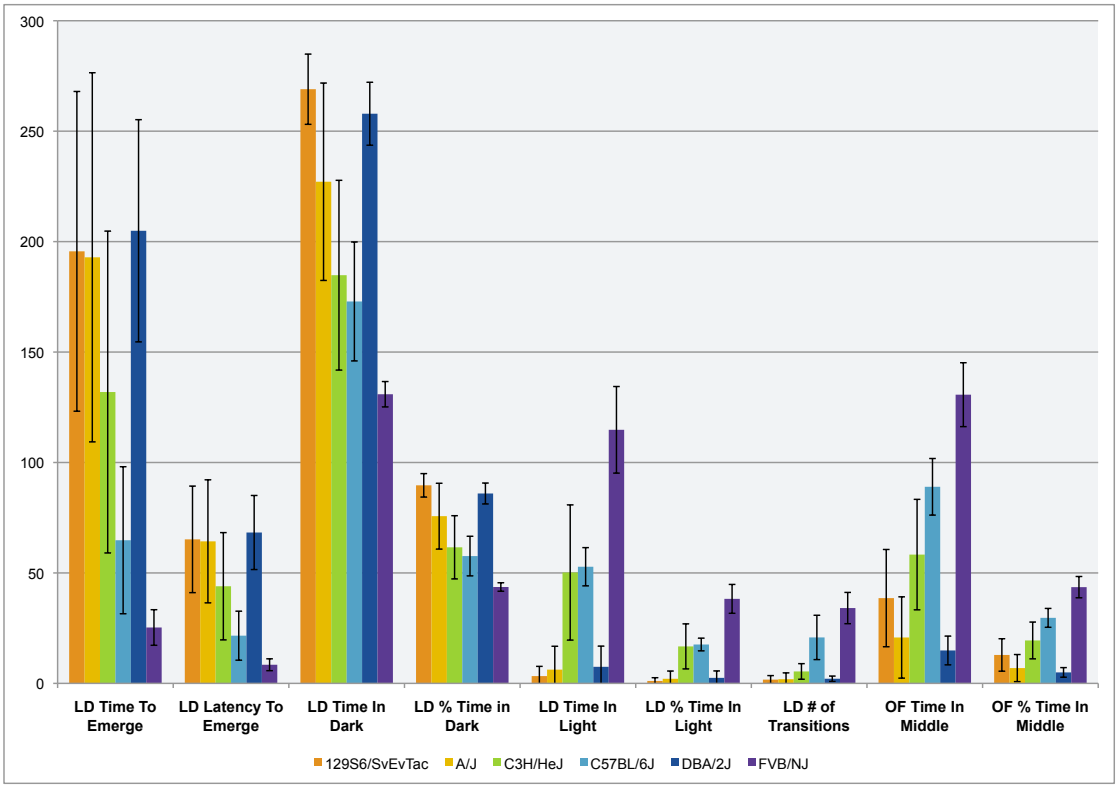

Figure 3-4. Light-dark box (LD) and open field (OF) behavioral phenotypes of inbred mouse strains. Error bars = 95% Confidence Interval. Gold: 129S1/SvImJ(blood) or 129S6/SvEvTac (neural); Yellow: A/J; Green: C3H/HeJ; Light blue: C57BL/6J; Dark blue: DBA/2J; Purple: FVB/nJ.

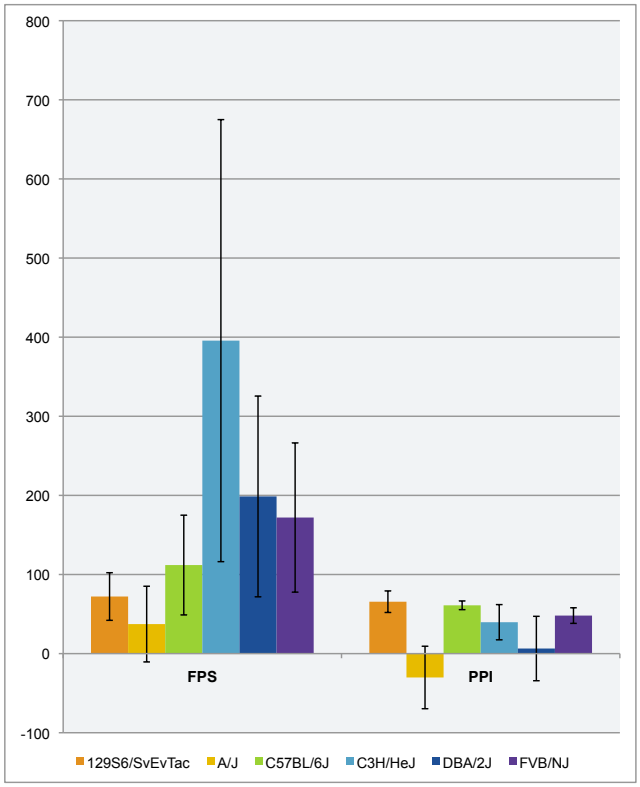

Figure 3-5. Fear potentiated startle (FPS) and pre-pulse inhibition (PPI) behavioral phenotypes of inbred mouse strains. Error bars = 95% Confidence Interval. Gold: 129S1/SvImJ(blood) or 129S6/SvEvTac (neural); Yellow: A/J; Green: C3H/HeJ; Light blue: C57BL/6J; Dark blue: DBA/2J; Purple: FVB/nJ.

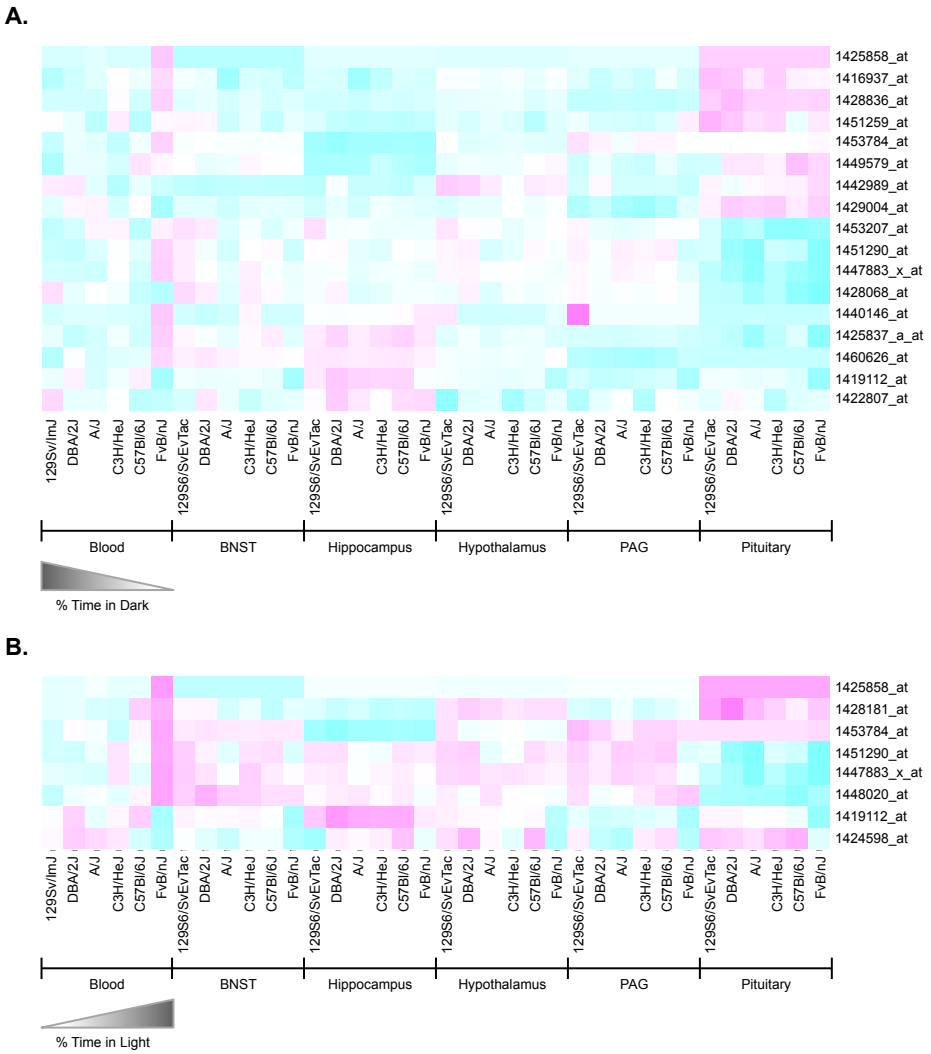

Figure 3-6. Heat maps of probe sets significantly associated with four anxiety-related phenotypes in blood and at least one neural tissue: A.) % time in dark, B.) % time in light, C.) FPS, and D.) PPI. The heat maps are clustered according to the y-axis. The y-axis shows the probe set identifiers. The x-axis is organized by strain and tissue. Strains are organized by phenotype, as shown by scale. Blue represents low signal intensity and pink represents high signal intensity; a more intense color means the relatively higher or lower the signal intensity.

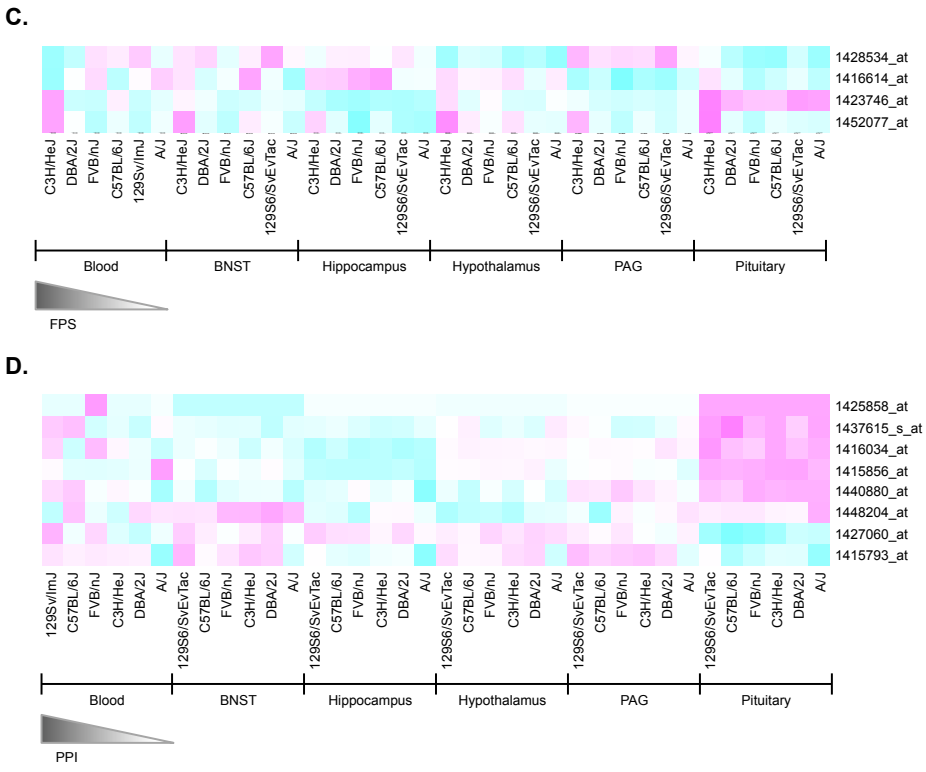

Figure 3-6. Heat maps of probe sets significantly associated with four anxiety-related phenotypes in blood and at least one neural tissue: A.) % time in dark, B.) % time in light, C.) FPS, and D.) PPI. The heat maps are clustered according to the y-axis. The y-axis shows the probe set identifiers. The x-axis is organized by strain and tissue. Strains are organized by phenotype, as shown by scale. Blue represents low signal intensity and pink represents high signal intensity; a more intense color means the relatively higher or lower the signal intensity. Continued.

REFERENCES

Akritas MG, Bershady MA (1996) Linear Regression for Astronomical Data with Measurement Errors and Intrinsic Scatter. The Astrophysical Journal 470, 706-706

Alter MD, Kharkar R, Ramsey KE, Craig DW, Melmed RD, Grebe TA, Bay RC, Ober-Reynolds S, Kirwan J, Jones JJ, Turner JB, Hen R, Stephan DA (2011) Autism and increased paternal age related changes in global levels of gene expression regulation. PLoS ONE 6, e16715-e16715

Braff DL, Geyer MA, Swerdlow NR (2001) Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. Psychopharmacology 156, 234-258

Bryant PA, Smyth GK, Robins-Browne R, Curtis N (2011) Technical variability is greater than biological variability in a microarray experiment but both are outweighed by changes induced by stimulation. PLoS ONE 6, e19556-e19556

Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R, Horvath S (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11, 589-589

Carter TA, Del Rio JA, Greenhall JA, Latronica ML, Lockhart DJ, Barlow C (2001) Chipping away at complex behavior: transcriptome/phenotype correlations in the mouse brain. Physiology & Behavior 73, 849-857

Choi DC, Furay AR, Evanson NK, Ostrander MM, Ulrich-Lai YM, Herman JP (2007) Bed Nucleus of the Stria Terminalis Subregions Differentially Regulate Hypothalamic-Pituitary-Adrenal Axis Activity: Implications for the Integration of Limbic Inputs. The Journal of Neuroscience 27, 2025-2034

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, Tarantino LM, Threadgill D, Toth LA, Valdar W, de Villena FP, Warden C, Whatley S, Williams RW, Wiltshire T, Yi N, Zhang D, Zhang M, Zou F (2004) The Collaborative Cross, a

community resource for the genetic analysis of complex traits. Nat Genet 36, 1133-1137

Crestani F, Lorez M, Baer K, Essrich C, Benke D, Laurent JP, Belzung C, Fritschy J-M, Luscher B, Mohler H (1999) Decreased GABAA-receptor clustering results in enhanced anxiety and a bias for threat cues. Nat Neurosci 2, 833-839

Davies MN, Lawn S, Whatley S, Fernandes C, Williams RW, Schalkwyk LC (2009) To What Extent is Blood a Reasonable Surrogate for Brain in Gene Expression Studies: Estimation from Mouse Hippocampus and Spleen. Front Neurosci 3

Davis M, Falls WA, Campeau S, Kim M (1993) Fear-potentiated startle: a neural and pharmacological analysis. Behavioural Brain Research 58, 175-198

Debey-Pascher S, Hofmann A, Kreusch F, Schuler G, Schuler-Thurner B, Schultze JL, Staratschek-Jox A (2011) RNA-stabilized whole blood samples but not peripheral blood mononuclear cells can be stored for prolonged time periods prior to transcriptome analysis. The Journal of Molecular Diagnostics: JMD 13, 452-460

Dulawa SC, Grandy DK, Low MJ, Paulus MP, Geyer MA (1999) Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 19, 9550-9556

Fan H, Hegde PS (2005) The transcriptome in blood: challenges and solutions for robust expression profiling. Current Molecular Medicine 5, 3-10

Fendt M, Koch M, Schnitzler H-U (1996) Lesions of the central gray block conditioned fear as measured with the potentiated startle paradigm. Behavioural Brain Research 74, 127-134

Field LA, Jordan RM, Hadix JA, Dunn MA, Shriver CD, Ellsworth RE, Ellsworth DL (2007) Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. Clinical Biochemistry 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, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5, R80-R80

Geyer MA, Dulawa SC (2003) Assessment of murine startle reactivity, prepulse inhibition, and habituation. Current Protocols in Neuroscience Chapter 8, Unit 8.17-Unit 18.17

Greenhall JA, Zapala MA, Caceres M, Libiger O, Barlow C, Schork NJ, Lockhart DJ (2007) Detecting genetic variation in microarray expression data. Genome Research 17, 1228-1235

Hovatta I, Tennant RS, Helton R, Marr RA, Singer O, Redwine JM, Ellison JA, Schadt EE, Verma IM, Lockhart DJ, Barlow C (2005) Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice. Nature 438, 662-666

Hovatta I, Zapala MA, Broide RS, Schadt EE, Libiger O, Schork NJ, Lockhart DJ, Barlow C (2007) DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. Genome Biology 8, R25-R25

Jasinska AJ, Service S, Choi O-w, DeYoung J, Grujic O, Kong S-y, Jorgensen MJ, Bailey J, Breidenthal S, Fairbanks LA, Woods RP, Jentsch JD, Freimer NB (2009) Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits. Human Molecular Genetics 18, 4415-4427

Kurano Y, Nakamura M, Ichiba M, Matsuda M, Mizuno E, Kato M, Izumo S, Sano A (2006) Chorein deficiency leads to upregulation of gephyrin and GABA(A) receptor. Biochemical and Biophysical Research Communications 351, 438-442

Leonardson AS, Zhu J, Chen Y, Wang K, Lamb JR, Reitman M, Emilsson V, Schadt EE (2010) The effect of food intake on gene expression in human peripheral blood. Human Molecular Genetics 19, 159-169

Li X, Serwanski DR, Miralles CP, Nagata K-i, De Blas AL (2009) Septin 11 is present in GABAergic synapses and plays a functional role in the cytoarchitecture of neurons and GABAergic synaptic connectivity. The Journal of Biological Chemistry 284, 17253-17265

Mozhui K, Karlsson R-M, Kash TL, Ihne J, Norcross M, Patel S, Farrell MR, Hill EE, Graybeal C, Martin KP, Camp M, Fitzgerald PJ, Ciobanu DC, Sprengel R, Mishina M, Wellman CL, Winder DG, Williams RW, Holmes A (2010) Strain Differences in Stress Responsivity Are Associated with Divergent Amygdala Gene Expression and Glutamate-Mediated Neuronal Excitability. The Journal of Neuroscience 30, 5357-5367

Ralph RJ, Paulus MP, Fumagalli F, Caron MG, Geyer MA (2001) Prepulse inhibition deficits and perseverative motor patterns in dopamine transporter knock-out mice: differential effects of D1 and D2 receptor antagonists. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 21, 305-313

Ralph-Williams RJ, Paulus MP, Zhuang X, Hen R, Geyer MA (2003) Valproate attenuates hyperactive and perseverative behaviors in mutant mice with a dysregulated dopamine system. Biological Psychiatry 53, 352-359

Risbrough VB, Brodkin JD, Geyer MA (2003) GABA-A and 5-HT1A receptor agonists block expression of fear-potentiated startle in mice. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology 28, 654-663

Risbrough VB, Geyer MA (2005) Anxiogenic treatments do not increase fear-potentiated startle in mice. Biological Psychiatry 57, 33-43

Risbrough VB, Geyer MA, Hauger RL, Coste S, Stenzel-Poore M, Wurst W, Holsboer F (2009) CRF1 and CRF2 receptors are required for potentiated startle to contextual but not discrete cues. Neuropsychopharmacology. 34(6):1494-503

Scherzer CR, Eklund AC, Morse LJ, Liao Z, Locascio JJ, Fefer D, Schwarzschild MA, Schlossmacher MG, Hauser MA, Vance JM, Sudarsky LR, Standaert DG, Growdon JH, Jensen RV, Gullans SR (2007) Molecular markers of early Parkinson's disease based on gene expression in blood. Proceedings of the National Academy of Sciences 104, 955-960

Somerville LH, Whalen PJ, Kelley WM (2010) Human bed nucleus of the stria terminalis indexes hypervigilant threat monitoring. Biological Psychiatry 68, 416-424

Song H, Ueno S-i, Numata S, Iga J-i, Shibuya-Tayoshi S, Nakataki M, Tayoshi SY, Yamauchi K, Sumitani S, Tomotake T, Tada T, Tanahashi T, Itakura M, Ohmori T (2007) Association between PNPO and schizophrenia in the Japanese population. Schizophrenia Research 97, 264-270

Sukumaran S, Xue B, Jusko WJ, Dubois DC, Almon RR (2010) Circadian variations in gene expression in rat abdominal adipose tissue and relationship to physiology. Physiological Genomics 42A, 141-152

Swerdlow NR, Geyer MA, Braff DL (2001) Neural circuit regulation of prepulse inhibition of startle in the rat: current knowledge and future challenges. Psychopharmacology 156, 194-215

Takahashi M, Hayashi H, Watanabe Y, Sawamura K, Fukui N, Watanabe J, Kitajima T, Yamanouchi Y, Iwata N, Mizukami K, Hori T, Shimoda K, Ujike H, Ozaki N, Iijima K, Takemura K, Aoshima H, Someya T (2010) Diagnostic classification of schizophrenia by neural network analysis of blood-based gene expression signatures. Schizophrenia Research 119, 210-218

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO (2003) Individuality and variation in gene expression patterns in human blood. Proceedings of the National Academy of Sciences of the United States of America 100, 1896-1901

Willott JF, Tanner L, O'Steen J, Johnson KR, Bogue MA, Gagnon L (2003) Acoustic startle and prepulse inhibition in 40 inbred strains of mice. Behavioral Neuroscience

117, 716-727

Winn ME, Zapala MA, Hovatta I, Risbrough VB, Lillie E, Schork NJ (2010) The effects of globin on microarray-based gene expression analysis of mouse blood. Mammalian Genome 21, 268-275

Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F (2004) A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association 99, 909-917

Yoo DY, Kim W, Kim DW, Yoo K-Y, Chung JY, Youn HY, Yoon YS, Choi SY, Won M-H, Hwang IK (2011) Pyridoxine enhances cell proliferation and neuroblast differentiation by upregulating the GABAergic system in the mouse dentate gyrus. Neurochemical Research 36, 713-721

Zapala MA, Hovatta I, Ellison JA, Wodicka L, Del Rio JA, Tennant R, Tynan W, Broide RS, Helton R, Stoveken BS, Winrow C, Lockhart DJ, Reilly JF, Young WG, Bloom FE, Lockhart DJ, Barlow C (2005) Adult mouse brain gene expression patterns bear an embryologic imprint. Proceedings of the National Academy of Sciences of the United States of America 102, 10357-10362

Zapala MA, Schork NJ (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proceedings of the National Academy of Sciences of the United States of America 103, 19430-19435

CHAPTER 4

Gene Expression Profiling of Human Whole Blood Samples with the Illumina WG-DASL Assay

ABSTRACT

Background

Microarray-based gene expression analysis of peripheral whole blood is a common strategy in the development of clinically relevant biomarker panels for a variety of human diseases. However, the results of such an analysis are often plagued by decreased sensitivity and reliability due to the effects of relatively high levels of globin mRNA in whole blood. Globin reduction assays have been shown to overcome such effects, but they require large amounts of total RNA and may induce distinct gene expression profiles. The Illumina whole genome DASL assay can detect gene expression levels using partially degraded RNA samples and has the potential to detect rare transcripts present in highly heterogeneous whole blood samples without the need for globin reduction. We assessed the utility of the whole genome DASL assay in an analysis of peripheral whole blood gene expression profiles.

Results

We find that gene expression detection is significantly increased with the use of whole genome DASL compared to the standard IVT-based direct hybridization. Additionally, globin-probe negative whole genome DASL did not exhibit significant improvements over globin-probe positive whole genome DASL. Globin reduction further increases the detection sensitivity and reliability of both whole genome DASL and IVT-based direct hybridization with little effect on raw intensity correlations. Raw intensity correlations between total RNA and globin reduced RNA were 0.955 for IVT-based direct hybridization and 0.979 for whole genome DASL.

Conclusions

Overall, the detection sensitivity of the whole genome DASL assay is higher than the IVT-based direct hybridization assay, with or without globin reduction, and should be considered in conjunction with globin reduction methods for future blood-based gene expression studies.

BACKGROUND

Peripheral whole blood is an attractive source of mRNA for the identification, examination, and development of disease biomarkers via microarray-based gene expression (Rockett et al. 2004). In fact, many studies have explored the utility of gene expression patterns in whole blood for the purposes of classifying or predicting clinical conditions (Hoang et al. 2010; Lin et al. 2009; Takahashi et al. 2010). However, the sensitivity and specificity of microarray assays using peripheral whole blood are reduced due to the relatively high proportion of globin mRNA present in total RNA, which obscures the detection of transcripts expressed at low levels in whole blood (Fan et al. 2004; Wright et al. 2008). While globin reduction assays have been shown to overcome these effects when used in conjunction with Affymetrix microarrays (Vartanian et al. 2009) and the standard Illumina direct hybridization assay (Debey et al. 2006; Tian et al. 2009), globin reduction assays require large amounts of total RNA (Vartanian et al. 2009), fail to completely eliminate globin transcripts (Vartanian et al. 2009), and may induce distinct gene expression profiles (Liu et al. 2006). Consequently, methods of developing blood-based gene expression biomarker panels that do not involve globin reduction are needed. Developing a microarray-based gene expression assay that does not rely on globin reduction or other methods of sample fractionation, such as the isolation of PBMCs or other cell types from the blood, should reduce sample variability introduced by sample handling and preparation. This will result in a more accurate reflection of the transcriptome at the time of blood draw, and will reduce time and cost.

There are ways to eliminate the need for globin reduction including 1.) the removal of globin probes from the microarray; and 2.) the elimination of globin

transcript amplification. Originally developed for the profiling of partially degraded and fixed RNA samples, the highly sensitive and reproducible Illumina cDNA-mediated annealing, selection, extension and ligation (DASL) assay (Fan et al. 2004; April et al. 2009) uses random priming and a modifiable oligo pool for cDNA synthesis. Random priming in conjunction with PCR amplification may allow for the increased detection of low abundance transcripts. In addition, removing globin-specific oligos from the DASL Assay Oligo Pool (DAP) should decrease noise associated with the high abundance of globin mRNA transcripts and potentially eliminate the necessity of globin reduction. Currently, the DAP is available with and without globin-specific oligos. In order to assess the need for globin reduction with the Illumina DASL assay, we compared microarray gene expression profiles of peripheral blood total RNA and globin-reduced RNA amplified via in vitro transcription (IVT)-based direct hybridization, DASL with globin-specific oligos, and DASL without globin-specific oligos.

METHODS SUMMARY

Peripheral whole blood samples were collected from eight human donors in PAXGene blood RNA tubes. RNA was isolated after freezing and storage and then prepared for gene expression analysis using the Illumina Human-Ref8 v3.0 Beadchip. Alpha and beta globin were reduced from a portion of the total RNA using the GLOBINclear assay (Ambion, Austin, TX, USA). Two methods of microarray target preparation were examined: Illumina IVT-based direct hybridization (IVT) and Illumina Whole-Genome DASL (WG-DASL) (**Figure 4-1**). The differences between IVT and WG-DASL are outlined in **Table 4-1**. Two DASL Assay Oligo pools (DAP) were utilized for DASL target preparation: the DASL Assay Oligo Pool with globin probes

(DAP +) and the DASL Asssay Oligo Pool without globin probes (DAP-). Comparisons involving the number of genes whose expression levels were detected and the actual levels of expression of the genes were made across the different platforms. A more complete description of the methods is provided in the Methods section.

RESULTS

Comparison between IVT and WG-DASL with and without globin reduction

Following target amplification as outlined in **Figure 4-1**, samples were hybridized with the Illumina Human-Ref8 v 3.0 following the manufacturer's instructions. Each target preparation method was assessed for performance by the number of probes detected as present (Detection p-value < 0.05) (**Figure 4-2**). Probes are generally detected as present if the probe intensity is significantly increased in comparison to the array background intensity. As noted, high levels of background due to the presence of globin transcripts in whole blood are known to decrease the number of significantly detected probes. The WG-DASL target preparation method significantly improved detection sensitivity compared to IVT (p-value = 2.13×10^{-9} from an analysis of variance (ANOVA)). Globin reduction decreased probe detection variability with both IVT and WG-DASL target preparation methods. The removal of globin probes from the DASL assay oligo pool (DAP-) resulted in a moderate increase in the number of probes detected but had no significant affect on detection variability (p-value = 0.680, ANOVA) as compared to the DAP+ target preparation method. Overall, 8677 probes were detected across all samples by the

five target preparation methods (**Figure 4-3**), but only 867 probes were detected by IVT alone. 2604 probes were detected by WG-DASL alone.

NanoDrop Spectrophotometer 260/280 ratios were moderately decreased following globin reduction with an average ratio equal to 2.06 prior and 1.97 post globin reduction (**Table 4-2**). However, raw intensity correlations indicate that whole and globin-reduced blood yield similar expression profiles with both IVT and DASL DAP+ assays. Overall raw intensity values increased in globin reduced samples (**Figure 4-4: A, C-D**) despite the failure of GLOBINclear to completely eliminate the two most abundant globin transcripts, hemoglobin alpha (HBA2) and hemoglobin beta (HBB). The removal of globin probes from the DASL Assay Oligo Pool (DAP-) (**Figure 4-4: B**) had little effect on gene expression profiles compared to DAP+ ($R2 = 0.993$) despite the near complete elimination of HBA2 and HBB.

Expression patterns maintained across target preparation methods

IVT target amplification is approximately linear while WG-DASL is approximately logarithmic, making it difficult to compare expression intensities directly. Thus, it was important in our analyses that the sample-to-sample relations are maintained among each target preparation method. Despite the differences in target amplification, sample relations were preserved across the five target preparation methods as shown by unsupervised hierarchical clustering (**Figure 4-5**). For example, with both IVT and WG-DASL, expression profiles for Sample 3 and Sample 7 exhibited the greatest differences from the other six samples, while for the IVT or WG-DASL whole blood RNA clustered separately from globin reduced RNA.

DISCUSSION

The analysis of whole blood, microarray-based gene expression profiles is often hindered by low sensitivity and high variability due to high levels of globin mRNA transcripts. These issues have been addressed by the development of globin-reduction methods, which specifically target and remove globin transcripts prior to array hybridization. However, studies have shown that globin reduction, like other methods of sample fractionation, may alter expression profiles (Liu et al. 2006), require large amounts of sample input, increase sample variability (Vartanian et al. 2009), and lead to increased costs. Thus, the ability to assay whole blood without sample fractionation or globin reduction may result in improved gene expression profile quality and decrease cost.

Here we describe the utility of a highly sensitive, whole-genome assay in the assessment of whole blood gene expression. Our results suggest that gene expression detection sensitivity is significantly increased with the whole-genome cDNA-mediated annealing, selection, extension and ligation (WG-DASL) assay as compared to IVT-based direct hybridization (IVT). The increased detection sensitivity of WG-DASL may be due to, 1.) random priming allowing for cDNA synthesis along the length of mRNA transcripts, or 2.) the ability to produce larger amounts of cDNA with PCR amplification. Regardless, attempts to further improve detection sensitivity and decrease expression variability through the selective removal of globin probes from the DASL assay oligo pool (DAP-) did not exhibit any large improvements over globin-probe positive DASL (DAP+). Our study also confirms the positive effect of globin reduction on microarray quality when used in conjunction with the Illumina BeadChip and standard IVT-based hybridization (Tian et al. 2009), while showing that

the positive effect of globin reduction extends to WG-DASL as well. However, as shown by unsupervised hierarchical clustering analysis, globin reduction appears to mildly influence gene expression profiles produced by both IVT and WG-DASL assays. Whether this is due to the induction of a globin reduction-specific profile (Liu et al. 2006), reduced RNA quality due to globin reduction, or the result of decreased noise is unknown, and should be taken into consideration while planning blood-based gene expression experiments.

CONCLUSIONS

Overall, our results suggest that the detection sensitivity of the WG-DASL assay is higher than the IVT-based direct hybridization assay, with or without globin reduction, and should be considered in conjunction with globin reduction methods for future blood-based gene expression studies. However, further investigation into the ability of the WG-DASL assay to distinguish between disease populations using whole blood is needed, as our study was not designed to address such issues.

METHODS

Blood collection and RNA isolation

For each sample, 2.5 ml whole blood was collected in a PAXgene Blood RNA collection tube (Qiagen, Valencia, CA, USA) and stored frozen at -80ºC prior to RNA isolation. RNA isolation was performed using the PAXGene Blood RNA Isolation System (Qiagen, Valencia, CA, USA). RNA quantity and quality were assessed by NanoDrop® Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) before and after globin reduction as well as before and after RNA amplification. For the 8

samples isolated, the total RNA yield ranged from 5.8 – 13.8 ug (average 7.9 ug +/- 1.0 ug), while A260/A280 ratios revealed all samples appeared to be of sufficient quality for microarray analysis (1.93 – 2.10) (**Table 4-2**), despite a moderate decrease in quality following globin reduction.

Globin Reduction

Alpha and beta globin mRNA were reduced from a portion of the total RNA samples using the GLOBINclear™ Human kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions with the recommended start quantity of 2 µg of total RNA. Each sample was processed twice then globin-reduced RNA pooled prior to RNA amplification and hybridization.

RNA amplification and hybridization

Whole blood total RNA and globin-reduced samples were assayed at both Scripps Genomic Medicine (La Jolla, CA, USA) and Illumina (San Diego, CA, USA) for IVT and DASL-based labelling, hybridization, and scanning, respectively (**Table 4-1**). Briefly, the WG-DASL method utilizes biotinylated random nonamer and oligo (dT) primers to convert 10-200 ng input RNA to cDNA. The biotinylated cDNA is then immobilized to a streptavidin-coated solid support and annealed to a pool of gene-specific oligonucleotides (DAP) for extension and ligation followed by PCR amplification with a biotinylated and a fluorophore-labeled universal primer. Finally, the single-stranded PCR products are eluted and hybridized to an Illumina BeadChip. For this study, 250 ng and 100 ng input RNA were utilized for IVT and DASL, respectively.

Gene expression analysis was performed on all whole blood RNA and globin-reduced samples using Human-Ref8 v3.0 Beadchips (Illumina, San Diego, CA, USA) containing 24,526 probes. All arrays were scanned with the Illumina BeadArray Reader and read into Illumina GenomeStudio® software (version 1.1.1). Individual samples were assayed once for all IVT analyses and twice for all DASL analyses. Given the limited amount of mRNA, replicates were only performed for the DASL assay due to its relative novelty as compared to the IVT assay. All replicates were highly correlated (average R2 = .9925). All raw data is available on the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo, [GSE 28064]).

Microarray data analysis

Raw intensities values were exported from GenomeStudio® software (version 1.1.1) for data processing and analysis in R (http://www.R-project.org) and Bioconductor (http://www.bioconductor.org) (Gentleman et al. 2004). Data quality and sample relations were assessed using the Bioconductor lumi package (Du et al. 2008). Probes with a Detection p-value less than 0.05 were considered present. Analysis of Variance (ANOVA) was used to assess the consistency of present/absent calls across the different sample preparation methods. Correlation coefficients were calculated from the raw intensity levels to assess the similarity of expression profiles.

ABBREVIATIONS

cDNA: complementary deoxyribonucleic acid; DAP+: DASL Assay Oligo Pool with globin probes; DAP-: DASL Assay Oligo Pool without globin probes; DASL: cDNA-mediated annealing, selection, extension and ligation; GR: RNA following

globin reduction by GLOBINclear; HBA2: hemoglobin, alpha 2; HBB: hemoglobin, beta; HBD: hemoglobin, delta; HBE1: hemoglobin, epsilon; HBG1: hemoglobin, gamma A; HBG2: hemoglobin, gamma G ;HBM: hemoglobin, mu; HBQ: hemoglobin, theta 1; HBZ: hemoglobin, zeta; IVT: in vitro-transcription; mRNA: messenger RNA; WB: total RNA from peripheral whole blood; WG: whole genome.

ACKNOWLEDGMENTS

This work was supported in part by the following research grants: U19 AG023122-05; R01 MH078151-03; N01 MH22005; U01 DA024417-01; P50 MH081755-01; R01 AG030474-02; N01 MH022005; R01 HL089655-02; R01 MH080134-03; U54 CA143906-01; UL1 RR025774-03 as well as the Price Foundation and Scripps Genomic Medicine.

Chapter 4, in full, is a reprint of the material as it appears in: Mary E. Winn, Marian Shaw, Craig April, Brandy Klotzle, Jian-Bing Fan, Sarah S. Murray, Nicholas J. Schork. (2011) Gene Expression Profiling of Human Whole Blood Samples with the Illumina WG-DASL Assay. BMC Genomics 12, 412-412. The dissertation author was the primary researcher and author of this paper. MEW participated in the design of the study, performed all data analysis, and drafted the manuscript. MS carried out the IVT-based microarray assays. CA and BK carried out the DASL-based microarray assays. JF and SSM participated in the design of the study. NJS conceived of and participated in the design and coordination of the study and helped draft the manuscript. All authors read and approved the final manuscript.

TABLES

Table 4-1. Summary of IVT and WG-DASL Methods.

	IVT	WG-DASL
Target Preparation Protocol Name	*In Vitro* Transcription	cDNA-mediated annealing, selection, extension and ligation
Total RNA Input Amount	50-100 ng	10-200 ng
Priming Method	Reverse Transcription off polyA tail	Poly(T) and random priming with biotinylated nonamers
Amplification	*In Vitro* Transcription (Linear)	PCR (Exponential)
Hybridization	Illumina BeadChip	

Table 4-2. RNA quality as assessed by 260/280 ratio. RNA quality was assessed before and after globin reduction as well as before and after amplification. tRNA; total RNA; GC RNA; GLOBINclear treated RNA or globin reduced RNA.

Sample ID	Before Amplification	Amplified	1st Globin Reduction	2nd Globin Reduction	Pooled Globin Reduction	Diluted to 20 ng/ul	Average
C00023 (tRNA)	2.03	2.01				2.07	2.04
C00023 (GC RNA)	1.98	1.99	2.02	1.89	2.00	1.94	1.97
C00027 (tRNA)	2.06	2.01				2.06	2.04
C00027 (GC RNA)	1.98	2.01	2.02	1.95	1.94	1.89	1.97
C00169 (tRNA)	2.04	2.12				1.89	2.02
C00169 (GC RNA)	1.96	1.95	1.88	1.90	2.02	1.84	1.93
C00179 (tRNA)	2.04	1.99				2.12	2.05
C00179 (GC RNA)	1.95	1.98	1.90	2.01	2.01	2.02	1.98
C00275 (tRNA)	2.03	1.95				2.14	2.04
C00275 (GC RNA)	1.99	1.99	1.87	1.91	1.93	2.01	1.95
C00304 (tRNA)	2.05	1.99				2.25	2.10
C00304 (GC RNA)	1.98	2.00	1.89	2.11	1.91	1.96	1.98
C00311 (tRNA)	2.04	2.10				2.12	2.09
C00311 (GC RNA)	2.02	2.00	2.04	1.93	1.97	2.01	2.00
C00342 (tRNA)	2.01	2.07				2.21	2.10
C00342 (GC RNA)	2.03	2.03	1.97	1.99	1.93	1.97	1.99

FIGURES

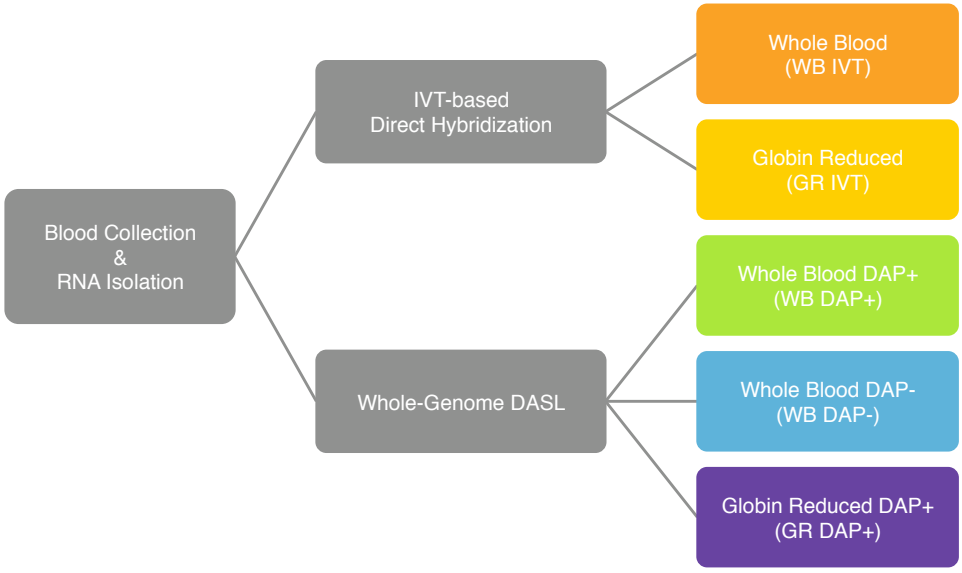

Figure 4-1. Flow diagram of study design. A PAXGene blood tube was collected from 8 individuals then frozen and stored for later processing. RNA was isolated and microarray targets prepared by one of five different methods: IVT-based direct hybridization with total RNA (WB IVT), IVT-based direct hybridization with globin-reduced RNA (GR IVT), whole-genome DAP+ DASL with total RNA (WB DAP+), whole-genome DAP- DASL with total RNA (WB DAP-), and whole-genome DAP+ DASL with globin-reduced RNA (GR DAP+).

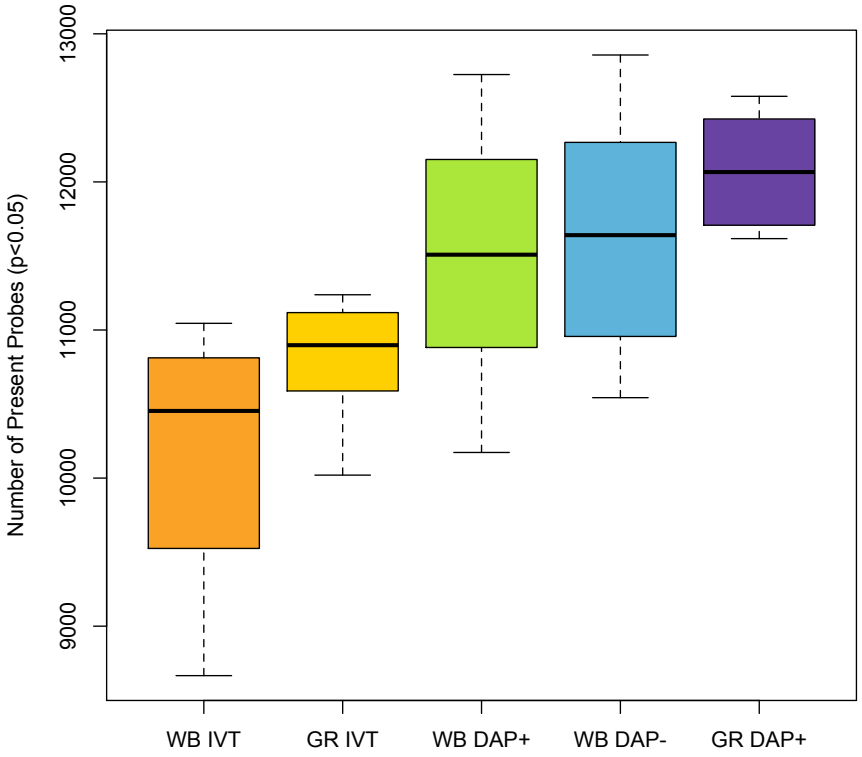

Figure 4-2. Box plots of present calls**.** The number of detected probes (detection p-value < 0.05) per target preparation method are shown. The boxes represent the lower quartile through the upper quartile, while the whiskers extend to 1.5 times the interquartile range. A bold line denotes the median. WB IVT and GR IVT (n=8). WB DASL+, WB DASL-, and GR DAP+ (n=16).

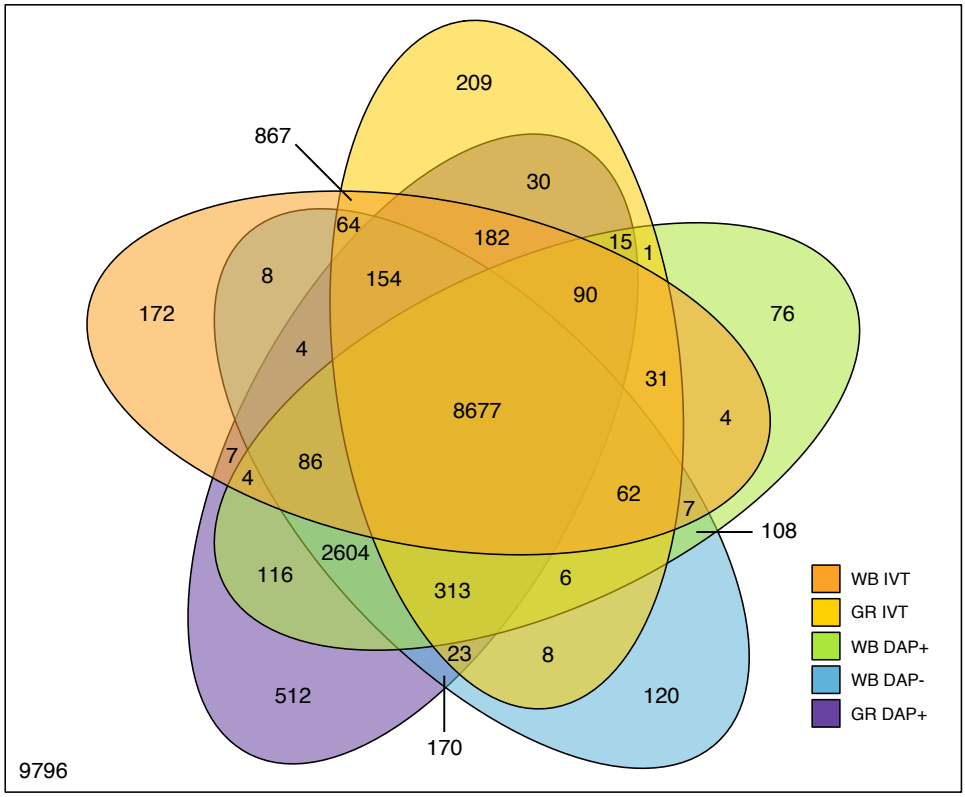

Figure 4-3. Overlap of detected probes**.** Probes detected as present across all eight samples per target preparation method are compared. WB IVT: IVT-based direct hybridization with total RNA, GR IVT: IVT-based direct hybridization with globin-reduced RNA, WB DAP+: whole-genome DAP+ DASL with total RNA, WB DAP-: whole-genome DAP- DASL with total RNA, and GR DAP+: whole-genome DAP+ DASL with globin-reduced RNA.

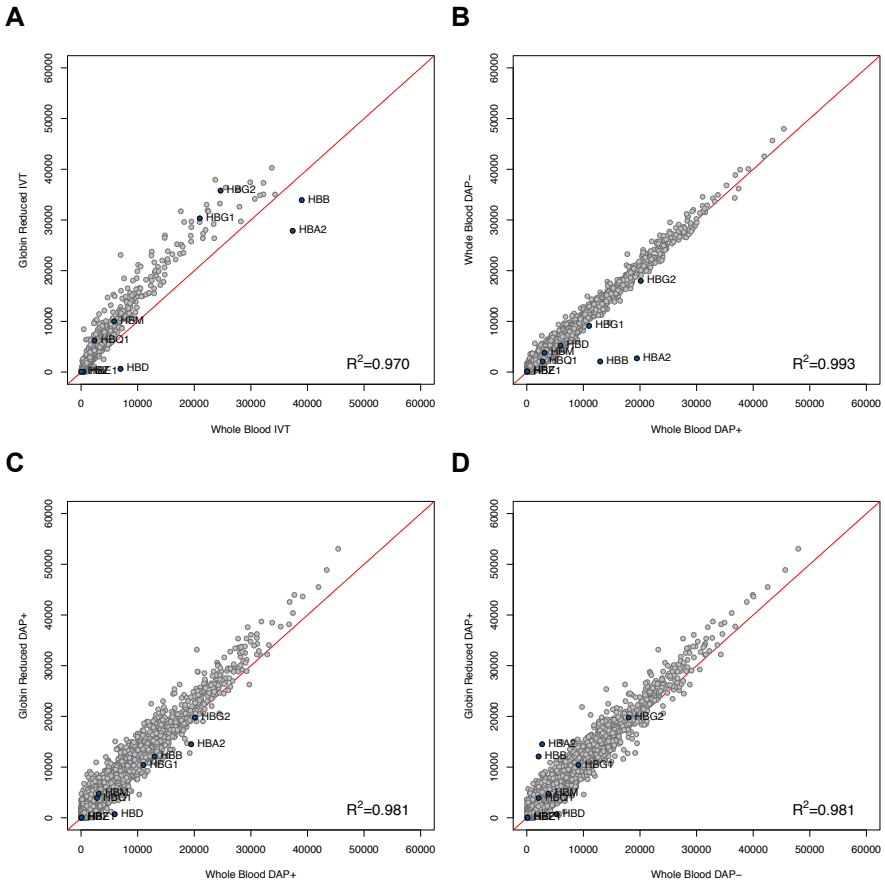

Figure 4-4. Raw intensity scatter plots**.** Raw intensities for all probes (n=24526) were compared for (A) whole blood RNA and globin reduced RNA with IVT, (B) whole blood RNA with DAP+ and whole blood RNA with DAP-, (C) whole blood RNA and globin reduced RNA with DAP+, and (D) whole blood RNA with DAP- and globin reduced RNA with DAP-. Correlations for sample 1 are depicted. Average correlations for paired WB IVT versus GR IVT, WB DAP+ versus WB DAP-, WB DAP+ versus GR DAP+, and WB DAP- versus GR DAP- samples are 0.955, 0.992, 0.976, and 0.979, respectively. All 8 hemoglobin genes assayed on Illumina BeadChip Human-Ref v3.0 are labelled: HBA2, HBB, HBD, HBE1, HBG1, HBG2, HBM, HBQ, and HBZ. GLOBINclear specifically targets only HBA2 and HBB for reduction.

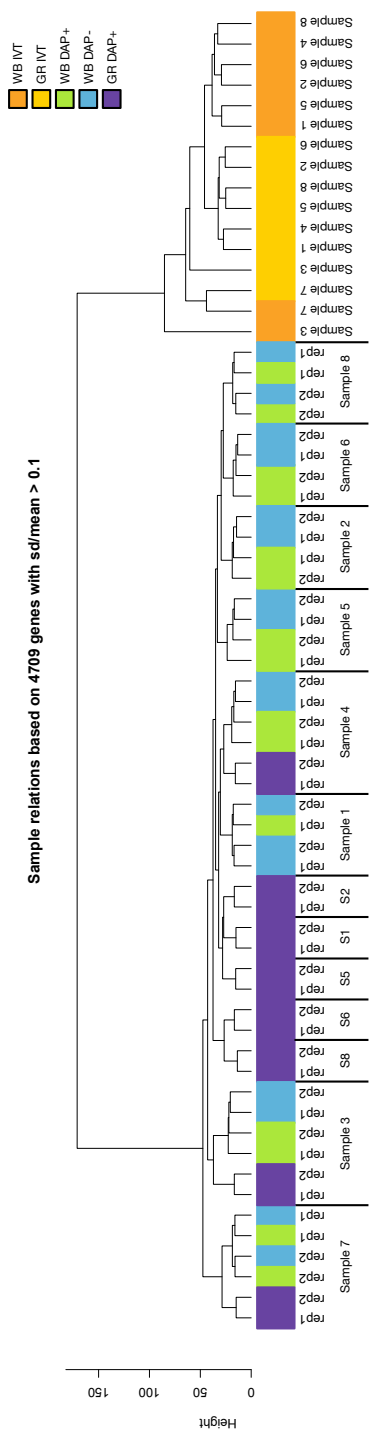

Figure 4-5. Sample relations as assessed by unsupervised hierarchical clustering. Dendrogram reflecting the clustering of the individual samples and the different sample preparation methods. The dendrogram was constructed using hierarchical clustering methods as implemented in the Bioconductor lumi package.

REFERENCES

April C, Klotzle B, Royce T, Wickham-Garcia E, Boyaniwsky T, Izzo J, Cox D, Jones W, Rubio R, Holton K, Matulonis U, Quackenbush J, Fan J (2009) Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. PLoS ONE 4:e8162

Debey S, Zander T, Brors B, Popov A, Eils R, Schultze JL (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

Du P, Kibbe WA, Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics 24:1547-1548

Fan J, Yeakley JM, Bibikova M, Chudin E, Wickham E, Chen J, Doucet D, Rigault P, Zhang B, Shen R, McBride C, Li H, Fu X, Oliphant A, Barker DL, Chee MS (2004) A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res 14:878-885

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Hoang LT, Lynn DJ, Henn M, Birren BW, Lennon NJ, Le PT, Duong KTH, Nguyen TTH, Mai LN, Farrar JJ, Hibberd ML, Simmons CP (2010) The early whole-blood transcriptional signature of dengue virus and features associated with progression to dengue shock syndrome in Vietnamese children and young adults. J Virol 84:12982-12994

Lin D, Hollander Z, Ng RT, Imai C, Ignaszewski A, Balshaw R, Freue GC, Wilson-McManus JE, Qasimi P, Meredith A, Mui A, Triche T, McMaster R, Keown PA, McManus BM (2009) Whole blood genomic biomarkers of acute cardiac allograft rejection. J Heart Lung Transplant 28:927-935

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

Rockett JC, Burczynski ME, Fornace AJ, Herrmann PC, Krawetz SA, Dix DJ (2004) Surrogate tissue analysis: monitoring toxicant exposure and health status of inaccessible tissues through the analysis of accessible tissues and cells. Toxicol Appl Pharmacol 194:189-199

Takahashi M, Hayashi H, Watanabe Y, Sawamura K, Fukui N, Watanabe J, Kitajima T, Yamanouchi Y, Iwata N, Mizukami K, Hori T, Shimoda K, Ujike H, Ozaki N, Iijima

K, Takemura K, Aoshima H, Someya T (2010) Diagnostic classification of schizophrenia by neural network analysis of blood-based gene expression signatures. Schizophr Res 119:210-218

Tian Z, Palmer N, Schmid P, Yao H, Galdzicki M, Berger B, Wu E, Kohane IS (2009) A practical platform for blood biomarker study by using global gene expression profiling of peripheral whole blood. PLoS ONE 4:e5157

Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA (2009) Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. BMC Genomics 10:2

Wright C, Bergstrom D, Dai H, Marton M, Morris M, Tokiwa G, Wang Y, Fare T (2008) Characterization of globin RNA interference in gene expression profiling of whole-blood samples. Clin Chem 54:396-405

CHAPTER 5

Comparison of Lymphocyte and Brain Gene Expression Patterns in Young Autistic Cases and Controls

ABSTRACT

There has been great interest in the identification of gene expression differences in whole blood, or cell types extracted from blood, between diseased and non-diseased individuals. Such differences might reflect underlying molecular pathologies associated with a disease or act, either, as clinically accessible biomarkers of disease susceptibility or surrogate endpoints for pathogenic processes. It is therefore important to not only assess the strength of associations between gene expression patterns and a disease, but also their potential biological relevance to the disease. We assessed gene expression differences in lymphocytes obtained from young autistic cases and controls to explore the potential blood-based cell type analyses have in producing biologically relevant biomarkers and surrogate endpoints for autism. We contrast our results with gene expression differences recently reported in an expression study involving neural tissues in autism as well as the literature as a whole. We find evidence for gene expression differences in pathways of relevance to autism pathology, but these differences are not entirely consistent with those found in the brain. In particular we find development, transcription, translation, and apoptosis and survival pathways to be among the most enriched in lymphocytes and brain. Our evaluation of the clinical utility of a lymphocyte-based classifier of autism suggests that there is potential, but replication studies are in order.

INTRODUCTION

There is growing interest in the identification of biomarkers of disease susceptibility, clinical outcomes, and drug response that are relatively easy to interrogate (Addona et al. 2011; Perlis 2011). Many diseases, such as autism and related neurodevelopmental or neuropsychiatric diseases, have primary lesions in tissues that are difficult to access (e.g., neural tissues), hence motivating researchers to consider the utility of biomarkers in more accessible tissues such as blood (Noelker et al. 2011; Le-Niculescu et al. 2009; Rockett et al. 2004). Assaying blood for potential biomarkers with modern transcriptomic, proteomic, and epigenomic technologies has, in fact, led to a number of recent notable successes (Feinberg et al. 2010; Rosenberg et al. 2010; Teschendorff et al. 2008). Despite these successes, there are a number of issues surrounding the use of blood-based assays for differentiating individuals with and without, e.g., autism or another clinical condition.

Blood is composed of a number of cell types known to manifest unique gene and protein expression, as well as epigenomic, profiles (Miao et al. 2008; Jacobsen et al. 2006; Palmer et al. 2006). These unique profiles are not only influenced by genetic factors (Yang et al. 2010; Göring et al. 2007) but also environmental factors, such as diet (Leonardson et al. 2010). Thus, not only is it the case that variation in the fraction of blood that is composed of different cell types across individuals can impact, e.g., blood gene expression profiles (Palmer et al. 2006; Whitney et al. 2003), but the degree to which different blood cell types express common genes amongst themselves and with respect to different tissues raises questions about the biological coherence or relevance of expression patterns in the blood with respect to expression patterns in other tissues.

One way to avoid issues associated with the heterogeneity of cell types in the blood is to focus on one particular blood cell type or cell family. Peripheral blood mononuclear cells (PBMCs) and lymphocytes have been studied widely in this regard (Baine et al. 2011; Gupta et al 2011; Bowden et al. 2006; Vawter et al. 2004). Given their role in immune system activity, the interrogation of perturbations or genomic alterations in lymphocytes is seen as logical given that lymphocytes more or less act as 'sentinels' of immune system dysregulation and other disturbances possibly associated with a disease (Fan et al. 2005). In addition, since many genes are known to be expressed in both lymphocytes and other tissues – and possibly ubiquitously expressed across all cell types and tissues – it may be the case that pathogenic molecular genetic disturbances that manifest in lymphocytes are also present in cell types more directly relevant to the fundamental lesions associated with a disease (Rollins et al. 2010). Such common disturbances may not necessarily occur at the individual gene level, but possibly at higher levels of biological organization, such as at the pathway or molecular physiologic process level (Subramanian et al. 2005).

Ultimately, however, even if a biologically-sound lymphocyte or whole blood-based biomarker profile is found to differentiate individuals with and without a particular condition such as autism or autism-related clinical phenotypes for diagnostic, prognostic, or therapeutic purposes, a good question is whether or not such differentiation can be achieved through other, possibly less invasive and less costly, procedures (Rosenberg et al. 2010). This question, though not directly related to the biological relevance of blood cell profiles, or processes in blood cells, to other cell types, does bear on the clinical utility and motivation for blood-based biomarker assays.

We assessed lymphocyte gene expression patterns in young individuals with and without autism. We compared the results of our study with those of a previous report investigating brain gene expression patterns in autism (Chow et al. 2011, submitted) as well as genes implicated in previous genome-wide association, copy number variation, and gene expression studies. We considered not only individual gene expression differences, but also pathway level differences. Finally, we compared the classification accuracy of lymphocyte-based gene expression patterns with the accuracy of classifiers based on brain gene expression.

RESULTS

Differential gene expression and gene set enrichment analysis between autism cases and controls in lymphocytes

We analyzed lymphocyte gene expression profiles from 290 individuals (total number of arrays = 347) collected by the Autism Center of Excellence (La Jolla, CA, USA). This is a diverse data set containing both single or longitudinal time points as well as single (child only), duo (mother- or father-child), or trio (mother, father, and child) samples for young children (12-24 months of age) at risk for autism spectrum disorder (ASD), developmental delay (DD), or language delay (LD). Given the heterogeneous nature of the data set, this study focuses on 76 male, first time-point samples (cases=45, controls=31; **Table 5-1**).

We identified 2321 genes as differentially expressed in autism cases relative to controls (p-value < 0.05; FDR ~ 0.163)(See Supplementary Material Table 5-S1). Among the top 20 up- and/or down-regulated genes were genes potentially involved in autism pathogenesis, including genes involved in signal transduction (PGHD,

MNK2, CSNK1G2, CHD8, GPR44, TRAT1, FCGR1B, IFIT3, OAS1), transport (SORL1, ABCA7, VAMP2, KCNG1), and anti-apoptotic/pro-cell cycle genes (CABIN1, CHD8, UBA3, NAMPT, ARG1, IFI27). Of the 2312 differentially expressed genes, 67% and 23% were known to be expressed in the brain and fetal brain, respectively (eGenetics/SANBI EST database; http://biomart.org).

Differentially expressed genes were then subjected to enrichment analysis via MetaCore^{TM} (GeneGO Inc, St. Joseph, MI, USA). The most significant GeneGo Pathway Maps (n = 68, p<0.001, FDR<0.005) (**Table 5-2A**) were heavily populated by development pathways (26/68), while GeneGO Process Networks (n = 17, p <0.001, FDR<0.01) (**Table 5-2B**) were most often related to cell cycle (4/17).

Comparing lymphocyte gene expression to brain gene expression via differential expression and gene set enrichment analysis

To test whether these findings are comparable to those from brain gene expression studies, we compared our results to those identified by Chow et al. (Chow et al., submitted) in which they identified 2017 genes to be differentially expressed in autism cases relative to controls. Comparing GeneGO Pathway Maps enrichment between the two tissues (**Table 5-3**), the same development pathway, A2A receptor signaling, was the most significant among differentially expressed brain and lymphocyte genes. In general, 9 of the top 15 pathways enriched among differentially expressed genes in lymphocytes were significantly enriched in the brain (**Table 5-3A**), while 12 of the top 15 pathways enriched among differentially expressed genes in brain were significantly enriched in lymphocytes (**Table 5-3B**).

We next looked at the individual genes differentially expressed in the same direction in both lymphocytes and brain (**Figure 5-1**); 20 genes were up-regulated (**Table 5-4**) while 56 genes were down-regulated (**Table 5-5**). Genome-wide association, copy-number variation, and/or gene expression analyses had previously identified many of these genes (33%) to be dysregulated or altered in autism cases. Among the genes up-regulated in both the brain and lymphocytes, apoptosis and survival pathways were among the most significant (**Table 5-6A**), while the most significant GeneGO Process Networks were heavily populated by immune response and inflammation processes (**Table 5-6B**). Down-regulated genes were heavily enriched for development pathways (**Table 5-7A**) and a variety of process networks, including several cell cycle networks (**Table 5-7B**).

Diagnostic classification of autism using lymphocyte and brain gene expression

Finally, we applied class prediction tools, as implemented in BRB-Array Tools, to assess the ability of differentially expressed genes to separate autism cases from controls using lymphocyte and brain gene expression (**Figure 5-2**). Using Receiver Operator Curve (ROC) analysis to test the significance of each model, a 67-gene model developed from lymphocyte gene expression yielded a cross-validated AUC equal to 0.668 (p=0.007) (**Figure 5-2A**, blue)(See Supplementary Material Table 5-S2). A 116-gene model developed from brain gene expression yielded a cross-validated AUC equal to 0.704 (p=0.024) (**Figure 5-2B**, blue)(See Supplementary Material Table 5-S3). These two models were not significantly different (p=0.757).

Testing the lymphocyte and brain-based models on the opposite tissue, the 67-gene lymphocyte model tested on brain gene expression yielded an AUC equal to 0.574 (p=0.240) (**Figure 5-2B**, green). The 116-gene brain model tested on lymphocyte gene expression yielded an AUC equal to 0.618 (p=0.042) (**Figure 5-2A**, green). These two models were not significantly different (p=0.732). Neither model, when tested on the opposite tissue, was significantly different compared to the original, cross-validated model (p=0.58152, **Figure 5-2A**; p=0.27665, **Figure 5-2B**).

The genes included in each model were enriched for apoptosis and survival, immune response, and development GeneGO Pathway Maps (**Table 5-8**).

DISCUSSION

Understanding the strength of associations between gene expression patterns and a disease, but also their potential biological relevance to the disease, is important to the development of blood-based gene expression diagnostics. In the present study, we assessed differences in lymphocyte gene expression between young autistic cases and controls and compared them to the results of a previous brain gene expression study (Chow et al. 2011, submitted) and the current literature. Young autistic lymphocyte gene expression was found to display dysregulation among development, apoptosis and survival, cell cycle, and immune response pathways and networks, among others. Many of these pathways and networks were shown to be significantly dysregulated in the young autistic brain suggesting lymphocyte gene expression is capable of capturing biologically relevant genetic dysregulation. For instance, the top dysregulated pathway (A2A receptor signaling) in both lymphocytes and brain is a potent biological mediator that affects several cell types including

neuronal cells. In brain, A2AR in highly expressed in the dorso-ventral striatum and at lower levels in the cortex, cortico-striatal terminals and hippocampus (Svenningsson et al., 1997; Rebola et al., 2005) and is involved in neuronal excitability, neurotransmitters release, neuronal synaptic plasticity, cognition and neuro-protection/-inflammation (Wei et al., 2011). In vitro, stimulation of A2AR prevents apoptosis via PKA-cat activation (Huang et al., 2001) that in turn, upon stimulation, enhances PKC-zeta activity and cell survival (Qiu et al., 2000).

Similarly, GM-CSF signaling, ranked 2nd in lymphocytes and 17th in brain, is responsible for the proliferation, differentiation, survival and maturation of immune and neuronal cells via the transcription factors STAT3 and STAT5, that in turn activate proliferative proteins like cyclinD1/3, Pim-1 and anti-apoptotic proteins like Mcl-1 and BCL-2/-x as well as mitogen-activated proteins like ERK1/2 via SHC transforming protein 1 (Kolonics et al., 2001; Choi et al., 2007). A major pathway involved in apoptosis and survival (BAD phosphorylation) was also highly ranked in both lymphocytes and brain. BAD induces apoptosis by inhibiting anti-apoptotic BCL-2-family members BCL-x, Bcl-2 (Bergmann et al., 2002). Overall, development and apoptosis and survival pathways were significantly enriched in the analysis of the commonly dysregulated brain/lymphocytes genes as well as the gene expression classifiers, demonstrating the utility of lymphocyte gene expression in identifying biologically relevant disease profiles at a pathway or network level.

Lymphocytes were less successful at identifying individual genes shown to be significantly dysregulated in the autistic brain. Approximately 4% of the differentially expressed genes in the brain were dysregulated in the same direction in lymphocytes. Nonetheless, most of the genes are involved in developmental and

apoptotic functions, while one third of these common genes were involved or dysregulated in previous autism studies. DCUN1D1 is known to regulate cell growth, viability, and development (Kim et al. 2008), is a risk factor for frontotemporal lobar degeneration (Villa et al. 2009), and maps to an autism susceptibility locus (Mas et al. 2000). Other genes are involved in the dopamine and/or serotonin synthesis pathways (GCH1, YWHAZ) (Wang et al. 2009). YWHAZ, which has been nominally associated with autism (Anderson et al. 2009), has also been implicated in the regulation of neurite outgrowth (Ramser et al. 2010) and more generally in cell cycle regulation and cell growth and death (Mhawech et al. 2005). A previous study (Philippe et al. 1999) identified a candidate genomic region containing TCP1, a gene involved in cytoskeletal maintenance and neurotransmitter trafficking. VEZT is known to play a role in the establishment of adherens junctions thus regulating dendritic formation of hippocampal neurons (Sandra et al. 2010), a region of the brain enlarged in autism (Groen et al. 2010). SLC30A5 is believed to transport zinc and was shown to be down-regulated in both lymphocytes and brain (Chow et al. 2011, submitted, Gregg et al. 2008) as well as deleted among sporadic autism cases (O'Roak et al. 2011). Although it is not clear how several of these genes may play into neural development and the manifestation of autism, the converging evidence presented here suggests further study on the subject is warranted. Regardless, these findings demonstrate that peripheral blood, in particular lymphocytes, may be used to detect a proportion of the genetic dysregulation occurring in the brain of autistic subjects.

While we demonstrate the ability of lymphocyte gene expression to capture genes and pathways known to be dysregulated in neural tissue, it is necessary to show these profiles are capable of classifying autism cases from controls in both a

sensitive and specific manor. The evidence we provide suggests lymphocyte gene expression profiles are able to meet this requirement at a similar level as brain gene expression profiles, thus being a relevant RNA source for the development of gene expression-based biomarkers of autism. Nonetheless, future clinical studies specifically aimed at biomarker development will be needed to refine and test the brain and lymphocyte-based models built here as our study was not specifically designed to address the specific needs associated with development of a clinically relevant classifier and suffers from overall sample size, particularly among the brain. It also remains to be seen if gene expression will significantly add to the classification of autism using widely accepted clinical factors and screening tools such as The One-Year Well-Baby Check Up Approach (Pierce et al. 2011). Rather, gene expression may be more useful in tracking the success or failure of clinical interventions as well as differences in clinical progression or recession once a child is diagnosed with autism. Further studies equating gene expression with QTL and CNV analysis as well as clinical and imaging phenotypes will only enhance our ability to predict and classify autism in infants.

MATERIALS AND METHODS

Subject Identification

All procedures were performed according to protocols approved by the University of California, San Diego Institutional Review Board. All minor subjects assented to the study procedures, and one or both parents or legal guardians of each subject provided written informed consent for their child to participate.

Participants were obtained by: 1.) community referral (e.g., website or outside agency) or 2.) a general population-based screening method called the One-Year Well-Baby Check-Up Approach (Pierce et al. 2011) performed by the participant's pediatrician. Using the latter approach, toddlers as young as 12-months who were at-risk for an Autism Spectrum Disorder (ASD), Language Delay (LD), or Developmental Delay (DD) were recruited and tracked every six months until at least their third birthday, thus allowing for the prospective study of autism beginning at 12 months. Typically developing (TD) and type-1 error (TIE) control subjects were obtained from community referrals.

ASD subjects were diagnosed based on failure of the Autism Diagnostic Observation Schedule (ADOS)(Lord et al. 2001) as well as the clinical judgment of a PhD-level psychologist. While several ASD toddlers were only one year old at the time of blood sampling, all but one have been tracked and diagnosed with an ASD using the toddler module of the ADOS (Luyster et al. 2009) until at least age two, when the diagnosis of autism can be made reliably. Final diagnoses of an ASD for participants older than 30 months were confirmed with the Autism Diagnostic Interview–Revised (Luyster et al. 2009).

Sample Collection and Processing

From each subject, 4ml of venous blood was collected into EDTA-coated collection tubes and immediately transferred to an RNase-free laboratory, where all subsequent procedures took place. Total mRNA was extracted, stabilized, isolated, and stored from each blood sample in a manner as previously described (Glatt et al. 2009, Glatt et al. 2005, Tsuang et al. 2005). Briefly, each blood sample was passed

over a LeukoLOCK™ (Ambion, Austin, TX, USA) filter, which was flushed with PBS and then fully saturated with RNAlater® (Ambion, Austin, TX, USA). Each LeukoLOCK™ filter, containing bound, isolated, stabilized, and purified white blood cells, was sealed and stored in a sterile box at -20°C. Once all samples were collected, LeukoLOCK filters were processed by flushing the filter with TRI reagent® (Ambion, Austin, TX, USA) to lyse the cells and isolate mRNA. Eluted mRNA samples were stored at -20°C until transferred to Scripps Genomic Medicine (La Jolla, CA, USA) for quality assurance and microarray hybridization.

Sample Quantification and Quality Control

The concentration of mRNA in each sample was quantified by the absorption of ultraviolet light at 260 nm. The quantity of mRNA in each sample exceeded the minimally sufficient amount required for microarray hybridization. The purity of each mRNA sample was estimated by the 260:280 nm absorbance ratio, with an acceptable range designated a priori as 1.7-2.1. The quality of each mRNA sample was quantified by the RNA Integrity Number (RIN) and, according to convention, values of 6.0 or greater were deemed acceptable (Schroeder et al. 2006). A total of 339 samples selected for analysis in Wave I had acceptable levels of mRNA quantity, purity, and quality.

Labeling, Hybridization, and Scanning

Lymphocyte total RNA was assayed at Scripps Genomic Medicine (La Jolla, CA, USA) for labelling, hybridization, and scanning using Illumina HumanWG-6 v3.0 expression BeadChips (Illumina, San Diego, CA, USA) per the manufacturer's

instruction. All arrays were scanned with the Illumina BeadArray Reader® and read into Illumina GenomeStudio® software (version 1.1.1). Raw data was exported from Illumina GenomeStudio® for data pre-processing and normalization.

Data Processing

Data processing was performed using the lumi package (Du et al. 2008) for R (http://www.R-project.org) and Bioconductor (http://www.bioconductor.org) (Gentleman et al. 2004). Of the 347 arrays processed, 27 were identified as low-quality based on poor signal intensity (raw intensity box plots and average signal >2 standard deviations below the mean) and poor hierarchical clustering (Oldham et al. 2008) and were removed prior to log2 transformation and quantile normalization. The remaining 320 high-quality arrays were filtered for first time point, male, proband samples with a diagnosis of ASD, PDD-NOS, TD, or TIE (n=76).

Data Analysis and Gene Set Enrichment Analysis

For differential expression analysis, normalized expression values were imported into BRB-Array Tools (http://linus.nci.nih.gov/BRBArrayTools.html). Gene filtering was performed as previously described (Chow et al. 2011, submitted) followed by differential expression analysis via Class comparison between groups of arrays using a random variance model and 10,000 univariate permutation tests. Differentially expressed genes ($p<0.05$) were then assessed for pathway enrichment using the MetaCore software suite (www.genego.com/metacore.php) (GeneGO, Inc., St. Joseph, MI, USA).

Genes imported into MetaCore were filtered for known expression specific to the brain or the fetal brain followed by enrichment analysis of GeneGO Pathway Maps and GeneGO Process Networks using the default background gene list. GeneGO Pathway Maps are defined as sets of linear consecutive signals, or metabolic transformations, that have been confirmed as a whole by inferred relationships or experimental data. GeneGO Process Networks are network models of main cellular processes that are created manually by GeneGO using information from GO processes and GeneGO Pathway Maps.

Multigene models for classifying autistic and control samples were constructed using BRB-Array Tools Class Prediction methods for both lymphocytes and brain gene expression profiles (Radmacher et al. 2002). Class Prediction creates a multivariate predictor for determining which of the two classes a given sample belongs. Leave-one-out cross-validation was used to determine misclassification rate. The Bayesian Compound Covariate Predictor results were exported from BRB-Array Tools and the reliability of this classification assessed via receiver-operator characteristic (ROC) curve analysis using the pROC package for R (Robin et al. 2011). ROC curve analysis was also used to compare different classification models as implemented in the pROC package. Two models were built: one using lymphocyte gene expression, the other using brain gene expression. Then both of these models were used to classify autism cases and controls using the other tissue source.

ACKNOWLEDGEMENTS

Chapter 5, in part, is currently being prepared for submission for publication. Mary E Winn, Tiziano Pramparo, Maggie L Chow, Anthony Wynshaw-Boris, Karen

Pierce, Eric Courchesne, Nicholas J Schork. The dissertation author was the primary investigator and author of this paper.

TABLES

Table 5-1. Sample characteristics for lymphocyte gene expression.

	Cases (n=45)			**Controls** (n=31)		
	All Cases	**ASD**	**PDD-NOS**	**All Controls**	**TD**	**T1E**
Sample size (n)	45	31	14	31	11	20
Age (mean months ± s.d.)	27.10±9.63	24.28±9.98	33.34±4.83	19.12±7.66	24.41±10.06	16.22±3.79
Age (range in months)	12.60-43.50	12.60-43.50	26.87-41.03	12.5-44.93	12.50-44.93	12.93-24.63

Table 5-2. Enriched GeneGo Pathway Maps and GeneGo Process Networks.

A. GeneGo Pathway Maps (FDR<0.01)

GeneGo Pathway Map	p-value	Ratio	
Development_A2A receptor signaling	4.672E-07	13	43
Development_GM-CSF signaling	4.841E-07	14	50
Development_Flt3 signaling	6.289E-07	13	44
Development_PIP3 signaling in cardiac myocytes	1.454E-06	13	47
Translation_Insulin regulation of translation	2.539E-06	12	42
Apoptosis and survival_BAD phosphorylation	2.539E-06	12	42
Immune response_Function of MEF2 in T lymphocytes	3.128E-06	13	50
Development_Role of IL-8 in angiogenesis	3.422E-06	14	58
Transcription_CREB pathway	4.358E-06	12	44
Translation _Regulation of EIF4F activity	6.329E-06	13	53
G-protein signaling_Regulation of p38 and JNK signaling mediated by G-proteins	7.680E-06	11	39
Development_Role of HDAC and calcium/calmodulin-dependent kinase (CaMK) in control of skeletal myogenesis	7.906E-06	13	54
Chemotaxis_Leukocyte chemotaxis	1.818E-05	15	75
G-protein signaling_G-Protein alpha-12 signaling pathway	3.020E-05	10	37
Chemotaxis_CCR4-induced leukocyte adhesion	3.037E-05	9	30
Development_Thrombopoietin-regulated cell processes	3.398E-05	11	45
Regulation of lipid metabolism_Insulin signaling:generic cascades	5.252E-05	11	47
DNA damage_ATM/ATR regulation of G1/S checkpoint	5.370E-05	9	32
Cell cycle_Role of APC in cell cycle regulation	5.370E-05	9	32
Cytoskeleton remodeling_Cytoskeleton remodeling	6.134E-05	17	102
DNA damage_ATM / ATR regulation of G2 / M checkpoint	6.926E-05	8	26
Development_S1P1 receptor signaling via beta-arrestin	7.013E-05	9	33
Cell cycle_Spindle assembly and chromosome separation	7.013E-05	9	33
Chemotaxis_CXCR4 signaling pathway	9.060E-05	9	34
Regulation of degradation of deltaF508 CFTR in CF	9.354E-05	8	27
Development_EDNRB signaling	9.630E-05	11	50
Apoptosis and survival_HTR1A signaling	9.630E-05	11	50
Development_A2B receptor: action via G-protein alpha s	9.630E-05	11	50
Immune response _Immunological synapse formation	1.051E-04	12	59
Development_EPO-induced Jak-STAT pathway	1.158E-04	9	35
Development_Growth hormone signaling via STATs and PLC/IP3	1.158E-04	9	35
Development_IGF-1 receptor signaling	1.165E-04	11	51
Development_ACM2 and ACM4 activation of ERK	1.218E-04	10	43
Development_SSTR2 in regulation of cell proliferation	1.467E-04	9	36
Immune response_IL-5 signalling	1.496E-04	10	44
Cell cycle_Sister chromatid cohesion	1.575E-04	7	22
Cell adhesion_Chemokines and adhesion	1.651E-04	16	100
Development_FGFR signaling pathway	1.681E-04	11	53
Development_A1 receptor signaling	1.681E-04	11	53
Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	1.681E-04	11	53
Development_Delta- and kappa-type opioid receptors signaling via beta-arrestin	2.153E-04	7	23
Signal transduction_PTEN pathway	2.217E-04	10	46
Cytoskeleton remodeling_Fibronectin-binding integrins in cell motility	2.712E-04	8	31
Apoptosis and survival_Role of IAP-proteins in apoptosis	2.712E-04	8	31
Immune response_IL-3 activation and signaling pathway	2.712E-04	8	31
Development_GDNF signaling	2.891E-04	7	24
Cytoskeleton remodeling_FAK signaling	3.303E-04	11	57
Apoptosis and survival_Granzyme B signaling	3.437E-04	8	32
Development_PDGF signaling via STATs and NF-kB	3.437E-04	8	32
Immune response_TCR and CD28 co-stimulation in activation of NF-kB	3.471E-04	9	40

Table 5-2. Continued.

A. Continued.

GeneGo Pathway Map	p-value	Ratio	
Immune response_IL-22 signaling pathway	4.312E-04	8	33
G-protein signaling_Ras family GTPases in kinase cascades (scheme)	4.970E-04	7	26
Immune response_IL-10 signaling pathway	4.970E-04	7	26
Development_Growth hormone signaling via PI3K/AKT and MAPK cascades	5.108E-04	9	42
Cell adhesion_Alpha-4 integrins in cell migration and adhesion	5.360E-04	8	34
Development_NOTCH1-mediated pathway for NF-KB activity modulation	5.360E-04	8	34
Chemotaxis_Inhibitory action of lipoxins on IL-8- and Leukotriene B4-induced neutrophil migration	5.366E-04	10	51
Development_Membrane-bound ESR1: interaction with growth factors signaling	6.138E-04	9	43
G-protein signaling_G-Protein alpha-i signaling cascades	6.382E-04	7	27
Development_Angiopoietin - Tie2 signaling	6.605E-04	8	35
Development_Ligand-independent activation of ESR1 and ESR2	7.332E-04	9	44
Immune response_Regulation of T cell function by CTLA-4	8.073E-04	8	36
Immune response_IL-12-induced IFN-gamma production	8.073E-04	8	36
Signal transduction_Calcium signaling	8.708E-04	9	45
Atherosclerosis_Role of ZNF202 in regulation of expression of genes involved in Atherosclerosis	8.972E-04	6	21
Development_Beta-adrenergic receptors transactivation of EGFR	9.792E-04	8	37
Apoptosis and survival_nAChR in apoptosis inhibition and cell cycle progression	1.015E-03	7	29
NGF activation of NF-kB	1.015E-03	7	29
Chemotaxis_Lipoxin inhibitory action on fMLP-induced neutrophil chemotaxis	1.029E-03	9	46
Immune response_MIF - the neuroendocrine-macrophage connector	1.029E-03	9	46
Apoptosis and survival_DNA-damage-induced apoptosis	1.133E-03	5	15
Development_S1P4 receptor signaling pathway	1.175E-03	6	22
Transcription_Role of heterochromatin protein 1 (HP1) family in transcriptional silencing	1.175E-03	6	22
Cell cycle_Role of 14-3-3 proteins in cell cycle regulation	1.175E-03	6	22
Transcription_Sin3 and NuRD in transcription regulation	1.179E-03	8	38
Transport_Alpha-2 adrenergic receptor regulation of ion channels	1.209E-03	9	47

B. GeneGo Process Networks (FDR<0.01)

GeneGo Process Networks	p-value	Ratio	
Cell cycle_G2-M	3.704E-08	46	205
Signal Transduction_Cholecystokinin signaling	2.000E-06	27	106
Proteolysis_Ubiquitin-proteasomal proteolysis	7.311E-06	35	166
Transcription_Chromatin modification	1.047E-05	29	128
Cell cycle_G1-S	1.259E-05	34	163
Cell cycle_S phase	3.194E-05	31	149
Inflammation_IL-10 anti-inflammatory response	6.473E-05	21	87
Development_Hemopoiesis, Erythropoietin pathway	8.098E-05	28	135
Immune response_TCR signaling	1.238E-04	33	174
Cytoskeleton_Regulation of cytoskeleton rearrangement	1.492E-04	34	183
Cytoskeleton_Cytoplasmic microtubules	2.333E-04	24	115
Development_Regulation of angiogenesis	3.947E-04	38	223
Cell cycle_Mitosis	4.751E-04	32	179
Apoptosis_Apoptotic nucleus	6.167E-04	29	159
Cell adhesion_Leucocyte chemotaxis	6.389E-04	35	205
Apoptosis_Anti-Apoptosis mediated by external signals via MAPK and JAK/STAT	1.006E-03	31	179
Translation_Regulation of initiation	1.064E-03	24	127

Table 5-3. Overlap of the top 15 significantly enriched GeneGO Pathway Maps among differentially expressed genes in lymphocytes and brain.

A. Top 15 GeneGo Pathway Maps among differentially expressed genes in lymphocytes

GeneGO Pathway Map	Lymphocytes				Brain			
	Rank	Ratio		p-value	Rank	Ratio		p-value
Development_A2A receptor signaling	1	13	43	4.67E-07	1	11	43	1.92E-07
Development_GM-CSF signaling	2	14	50	4.84E-07	17	7	50	1.76E-03
Development_Flt3 signaling	3	13	44	6.29E-07	153	4	44	6.93E-02
Development_PIP3 signaling in cardiac myocytes	4	13	47	1.45E-06	6	8	47	2.13E-04
Translation_Insulin regulation of translation	5	12	42	2.54E-06	73	5	42	1.57E-02
Apoptosis and survival_BAD phosphorylation	6	12	42	2.54E-06	25	6	42	3.37E-03
Immune response_Function of MEF2 in T lymphocytes	7	13	50	3.13E-06	183	4	50	1.00E-01
Development_Role of IL-8 in angiogenesis	8	14	58	3.42E-06	9	8	58	9.27E-04
Transcription_CREB pathway	9	12	44	4.36E-06	4	8	44	1.32E-04
Translation _Regulation of EIF4F activity	10	13	53	6.33E-06	56	6	53	1.07E-02
G-protein signaling_Regulation of p38 and JNK signaling mediated by G-proteins	11	11	39	7.68E-06	237	3	39	1.59E-01
Development_Role of HDAC and calcium/calmodulin-dependent kinase (CaMK) in control of skeletal myogenesis	12	13	54	7.91E-06	120	5	54	4.15E-02
Chemotaxis_Leukocyte chemotaxis	13	15	75	1.82E-05	77	7	75	1.66E-02
G-protein signaling_G-Protein alpha-12 signaling pathway	14	10	37	3.02E-05	219	3	37	1.42E-01
Chemotaxis_CCR4-induced leukocyte adhesion	15	9	30	3.04E-05	314	2	30	2.88E-01

B. Top 15 GeneGO Pathway Maps among differentially expressed genes in brain

GeneGO Pathway Map	Brain				Lymphocytes			
	Rank	Ratio		p-value	Rank	Ratio		p-value
Development_A2A receptor signaling	1	11	43	1.92E-07	1	13	43	4.67E-07
Development_A2B receptor: action via G-protein alpha s	2	10	50	7.83E-06	28	11	50	9.63E-05
Transcription_P53 signaling pathway	3	8	39	5.34E-05	431	3	39	3.90E-01
Transcription_CREB pathway	4	8	44	1.32E-04	9	12	44	4.36E-06
Development_Thrombopoietin-regulated cell processes	5	8	45	1.55E-04	16	11	45	3.40E-05
Development_PIP3 signaling in cardiac myocytes	6	8	47	2.13E-04	4	13	47	1.45E-06
Cell adhesion_ECM remodeling	7	8	52	4.37E-04	504	2	52	8.08E-01
Cell cycle_Role of 14-3-3 proteins in cell cycle regulation	8	5	22	8.65E-04	74	6	22	1.18E-03
Development_Role of IL-8 in angiogenesis	9	8	58	9.27E-04	8	14	58	3.42E-06
Cytoskeleton remodeling_Cytoskeleton remodeling	10	11	102	9.27E-04	20	17	102	6.13E-05
Reproduction_GnRH signaling	11	9	72	9.33E-04	109	11	72	2.49E-03
Immune response_MIF - the neuroendocrine-macrophage connector	12	7	46	1.07E-03	70	9	46	1.03E-03
Signal transduction_PTEN pathway	13	7	46	1.07E-03	42	10	46	2.22E-04
Development_HGF signaling pathway	14	7	47	1.22E-03	218	7	47	1.68E-02
Development_Melanocyte development and pigmentation	15	7	49	1.56E-03	358	5	49	1.48E-01

Table 5-4. Genes up-regulated in both lymphocytes and brain.

Gene Symbol	Chr.	Lymphocytes		Brain		Reference
		Fold Change	p-value	Fold Changes	p-value	
C17orf65	17q21.31	1.11	0.0183	1.29	0.0229	
CYP2S1	19q13.1	1.12	0.0267	1.28	0.0352	
CYP4F12	19p13.1	1.23	0.0059	1.61	0.0493	
DCLRE1C	10p13	1.12	0.0497	1.33	0.0482	Pinto et al. (2010)
IL5RA	3p26-p24	1.19	0.0395	1.62	0.0264	Pinto et al. (2010)
KIAA0664	17p13.3	1.18	0.0180	1.35	0.0493	
LAT2	7q11.23	1.18	0.0050	1.34	0.0063	Jacquemont et al. (2006)
LILRB1	19q13.4	1.15	0.0126	1.86	0.0047	
MYBPH	1q32.1	1.22	0.0233	1.46	0.0297	
NFKB2	10q24	1.13	0.0403	1.56 1.43	0.0032 0.0420	
P2RY8	Yp11.3	1.11	0.0475	1.40	0.0224	Marshall et al. (2008)
PIM1	6p21.2	1.13	0.0201	1.40	0.0109	
RAPGEF1	9q34.3	1.13	0.0120	1.68	0.0097	
RUNX1	21q22.3	1.15	0.0267	1.83	0.0004	
SEPN1	1p36.13	1.19	0.0437	1.37	0.0229	
SIGLEC7	19q13.3	1.11	0.0136	1.60	0.0080	
SLC44A4	6p21.3	1.14	0.0255	1.46	0.0379	
SSH1	6p24	1.13	0.0302	1.31	0.0252	
TFAP2A	6p24	1.18	0.0258	2.13	0.0131	
UNC13D	17q25.1	1.21	0.0066	1.32	0.0377	

Table 5-5. Genes down-regulated in both lymphocytes and brain.

Gene Symbol	Chr.	Lymphocytes		Brain		Reference
		Fold Change	p-value	Fold Change	p-value	
ABCE1	4q31	-1.14 -1.14	0.015 0.017	-1.51	0.029	
ACAT2	6q25.3	-1.11	0.022	-1.49	0.005	
ACN9	7q21.3	-1.20	0.023	-1.27	0.018	
AIFM1	Xq25-q26	-1.14	0.010	-1.36	0.008	
ALG13	Xq23	-1.23	0.008	-1.42	0.030	
APIP	11p13	-1.15 -1.27	0.004 0.009	-1.29	0.036	
ARL4A	7p21.3	-1.11	0.025	-1.56	0.025	AGPC et al. (2007)
ATG4C	1p31.3	-1.16	0.013	-1.31	0.016	
C18orf10	18q12.2	-1.20	0.003	-1.35	0.014	Marshall et al. (2008)
C1orf124	1q42.12-q43	-1.18	0.007	-1.30	0.017	
CSNK1A1	5q32	-1.18	0.028	-1.30	0.014	Sarachana et al. (2010)
CYSLTR1	Xq13.2-q21.1	-1.18	0.014	-1.42	0.016	
DCTN6	8p12-p11	-1.16	0.005	-1.48	0.005	
DCUN1D1	3q26.3	-1.20	0.007	-1.51	0.024	Villa et al. (2009)
EEF1B2	2q33-q34	-1.18 -1.14	0.032 0.038	-1.30	0.024	
FAM18B	17p11.2	-1.27 -1.32	0.019 0.021	-1.35	0.031	
FBXO3	11p13	-1.16	0.022	-1.39	0.003	
GCC2	2q12.3	-1.14	0.045	-1.31 -1.26	0.038 0.039	
GNPDA2	4p12	-1.20	0.007	-1.50	0.041	Gregg et al. (2008)
GTPBP8	3q13.2	-1.11 -1.12	0.014 0.045	-1.28	0.015	
GTSF1	12q13.13	-1.16	0.009	-1.40	0.026	
HAT1	2q31.2-q33.1	-1.14 -1.20	0.019 0.022	-1.39	0.031	
HDAC9	7p21.1	-1.10	0.036	-1.26	0.040	Berkel et al. (2010)
HIF1A	14q21-q24	-1.23	0.039	-1.33	0.027	
HINT3	6q22.32	-1.12	0.010	-1.51	0.009	AGPC et al. (2007)
ISCA1P1	5q12.1	-1.12	0.036	-1.56	0.000	
KLHL5	4p14	-1.10	0.039	-1.34 -1.37	0.021 0.026	Gregg et al. (2008)
MIER1	1p31.3	-1.14	0.018	-1.41	0.022	
MTRR	5p15.3-p15.2	-1.15	0.019	-1.26	0.013	Mohammad et al. (2009) AGPC et al. (2007)
NDUFS4	5q11.1	-1.15	0.013	-1.32	0.021	Pinto et al. (2010)
NET1	10p15	-1.10	0.042	-1.38	0.013	
OSGEPL1	2q32.2	-1.12	0.019	-1.27	0.032	
PDCD10	3q26.1	-1.20 -1.25 -1.23	0.002 0.003 0.006	-1.41	0.014	
PIK3CA	3q26.3	-1.11	0.042	-1.39	0.043	
POT1	7q31.33	-1.16	0.019	-1.36	0.038	Marshall et al. (2008)
PRKACB	1p36.1	-1.33	0.010	-1.47	0.005	
PRPF4B	6p25.2	-1.22	0.046	-1.55	0.009	Gregg et al. (2008)

Table 5-5. Continued.

Gene Symbol	Chr.	Lymphocytes		Brain		Reference
		Fold Change	p-value	Fold Change	p-value	
PSMD10	Xq22.3	-1.22 -1.16 -1.28	0.004 0.015 0.024	-1.33	0.005	Piton et al. (2011)
RAD51AP1	12p13.2-p13.1	-1.18	0.000	-1.32	0.037	
RANBP6	9p24.1	-1.15	0.028	-1.33	0.009	
RBM41	Xq22.3	-1.11	0.007	-1.36	0.012	
SC4MOL	4q32-q34	-1.23	0.011	-1.45	0.007	AGPC et al. (2007)
SEC22C	3p22.1	-1.14	0.020	-1.29	0.045	
SFRS13A	1p36.11	-1.18	0.017	-1.56	0.030	
SLC25A40	7q21.12	-1.18 -1.14	0.026 0.027	-1.32	0.041	
SLC30A5	5q12.1	-1.16	0.016	-1.42	0.033	O'Roak et al. (2011) Gregg et al. (2008)
SYF2	1p36.11	-1.22	0.030	-1.23	0.048	
TAF9	5q11.2-q13.1	-1.16	0.025	-1.29	0.019	
TCP1	6q25.3-q26	-1.33	0.011	-1.31	0.046	Philippe et al. (2009)
TMPO	12q22	-1.23	0.004	-1.30 -1.47	0.023 0.016	Gregg et al. (2008)
TPM4	19p13.1	-1.10	0.023	-1.21	0.041	Baron et al. (2006)
TRMT11	6q11.1-q22.33	-1.20	0.007	-1.40	0.026	AGPC et al. (2007)
TWF1	12q12	-1.14	0.006	-1.48	0.037	Wang et al. (2010)
USP16	21q22.11	-1.16	0.008	-1.63	0.012	
VEZT	12q22	-1.11	0.017	-1.29	0.030	
YWHAZ	8q23.1	-1.30	0.010	-1.45	0.003	Anderson et al. (2009)

Table 5-6. GeneGO Pathway Maps and Process Networks enriched among genes up-regulated in both lymphocytes and brain.

A. GeneGO Pathway Maps (FDR < 0.05)

GeneGO Pathway Map	p-value	Ratio	
Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/Bcl-2 pathway	4.606E-05	3	41
Immune response_IL-27 signaling pathway	7.906E-04	2	24
Apoptosis and survival_APRIL and BAFF signaling	1.985E-03	2	38
Transcription_NF-kB signaling pathway	2.090E-03	2	39
Apoptosis and survival_Lymphotoxin-beta receptor signaling	2.308E-03	2	41

B. GeneGO Process Networks (FDR < 0.05)

GeneGO Process Network	pValue	Ratio	
Inflammation_IL-10 anti-inflammatory response	2.595E-05	4	87
Immune response_BCR pathway	1.538E-04	4	137
Immune response_Phagocytosis	9.858E-04	4	223
Inflammation_IL-2 signaling	1.224E-03	3	104
Inflammation_Protein C signaling	1.365E-03	3	108
Inflammation_Inflammasome	1.516E-03	3	112
Inflammation_Amphoterin signaling	1.762E-03	3	118
Inflammation_IL-6 signaling	1.805E-03	3	119
Cell cycle_G1-S Interleukin regulation	2.225E-03	3	128
Development_Hemopoiesis, Erythropoietin pathway	2.590E-03	3	135
Inflammation_IgE signaling	2.758E-03	3	138
Inflammation_MIF signaling	2.873E-03	3	140
Immune response_IL-5 signalling	3.748E-03	2	44
Immune response_TCR signaling	5.304E-03	3	174
Inflammation_Innate inflammatory response	6.013E-03	3	182
Cell cycle_G1-S Growth factor regulation	7.282E-03	3	195
Immune response_Antigen presentation	7.491E-03	3	197
Reproduction_Feeding and Neurohormone signaling	8.933E-03	3	210
Inflammation_Histamine signaling	9.528E-03	3	215
Inflammation_Neutrophil activation	1.053E-02	3	223
Immune response_Phagosome in antigen presentation	1.420E-02	3	249

Table 5-7. GeneGO Pathway Maps and Process Networks enriched among genes down-regulated in both lymphocytes and brain.

A. GeneGO Pathway Maps (FDR < 0.065) *1st three pathways significant with a FDR < 0.025

GeneGO Pathway Map	p-value	Ratio	
Transcription_Role of heterochromatin protein 1 (HP1) family in transcriptional silencing	1.965E-05	3	22
Apoptosis and survival_BAD phosphorylation	1.419E-04	3	42
Development_Role of HDAC and calcium/calmodulin-dependent kinase (CaMK) in control of skeletal myogenesis	3.007E-04	3	54
Apoptosis and survival_Beta-2 adrenergic receptor anti-apoptotic action	1.437E-03	2	23
Cytoskeleton remodeling_Role of PDGFs in cell migration	1.565E-03	2	24
Cell adhesion_Alpha-4 integrins in cell migration and adhesion	3.133E-03	2	34
Inhibitory action of Lipoxin A4 on PDGF, EGF and LTD4 signaling	3.133E-03	2	34
Development_Regulation of telomere length and cellular immortalization	3.318E-03	2	35
Development_Lipoxin inhibitory action on PDGF, EGF and LTD4 signaling	3.318E-03	2	35
G-protein signaling_G-Protein alpha-12 signaling pathway	3.703E-03	2	37
Transcription_Receptor-mediated HIF regulation	4.108E-03	2	39
Translation _Regulation of EIF2 activity	4.108E-03	2	39
Development_A2A receptor signaling	4.976E-03	2	43
Transcription_CREB pathway	5.206E-03	2	44
Development_Ligand-independent activation of ESR1 and ESR2	5.206E-03	2	44
Immune response_Inhibitory action of Lipoxins on pro-inflammatory TNF-alpha signaling	5.440E-03	2	45
Development_Hedgehog signaling	5.678E-03	2	46
Regulation of lipid metabolism_Insulin signaling:generic cascades	5.922E-03	2	47
Development_Leptin signaling via PI3K-dependent pathway	5.922E-03	2	47
Development_PIP3 signaling in cardiac myocytes	5.922E-03	2	47
Regulation of metabolism_Triiodothyronine and Thyroxine signaling	6.170E-03	2	48
Development_Melanocyte development and pigmentation	6.423E-03	2	49
Development_GM-CSF signaling	6.681E-03	2	50
Immune response_Function of MEF2 in T lymphocytes	6.681E-03	2	50
Development_A2B receptor: action via G-protein alpha s	6.681E-03	2	50
Development_IGF-1 receptor signaling	6.943E-03	2	51
ENaC regulation in airways (normal and CF)	7.210E-03	2	52
PGE2 pathways in cancer	8.039E-03	2	55
Regulation of lipid metabolism_Insulin regulation of glycogen metabolism	8.324E-03	2	56

B. GeneGO Process Networks (FDR < 0.20) *No significant networks with a FDR < 0.05

GeneGO Process Network	p-value	Ratio	
Inflammation_IL-6 signaling	2.363E-03	4	119
Development_Skeletal muscle development	4.698E-03	4	144
Inflammation_TREM1 signaling	4.815E-03	4	145

Table 5-7. Continued.

GeneGO Process Network	p-value	Ratio	
Apoptosis_Apoptotic mitochondria	5.771E-03	3	77
Cytoskeleton_Intermediate filaments	6.644E-03	3	81
Cell adhesion_Platelet aggregation	9.119E-03	4	174
Signal transduction_Leptin signaling	1.385E-02	3	106
Cell cycle_Meiosis	1.385E-02	3	106
Cell cycle_G2-M	1.591E-02	4	205
Reproduction_Feeding and Neurohormone signaling	1.724E-02	4	210
Muscle contraction_Nitric oxide signaling in the cardiovascular system	1.762E-02	3	116
Transcription_Chromatin modification	2.283E-02	3	128
Development_Melanocyte development and pigmentation	2.364E-02	2	50

Table 5-8. GeneGO Pathway Maps enriched among genes included in the lymphocyte-based class prediction model (n=67) and brain-based class prediction model (n=116).

GeneGO Pathway Map	Lymphocytes			Brain		
	p-value	Ratio		p-value	Ratio	
Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/Bcl-2 pathway	6.05E-02	1	41	5.85E-05	4	41
Immune response_CXCR4 signaling via second messenger	5.04E-02	1	34	7.24E-04	3	34
Immune response_Inhibitory action of Lipoxins on pro-inflammatory TNF-alpha signaling	2.03E-03	2	45	1.65E-03	3	45
Development_PEDF signaling	2.40E-03	2	49	2.28E-01	1	49
Apoptosis and survival_HTR1A signaling	2.50E-03	2	50	2.32E-01	1	50
Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/IAP pathway	4.03E-02	2	27	8.72E-03	1	27
Immune response_IL-4 - antiapoptotic action	4.46E-02	2	30	1.07E-02	1	30
Transcription_NF-kB signaling pathway	5.77E-02	2	39	1.77E-02	1	39
Immune response_TCR and CD28 co-stimulation in activation of NF-kB	5.91E-02	2	40	1.86E-02	1	40
Development_A2A receptor signaling	6.34E-02	2	43	2.13E-02	1	43

FIGURES

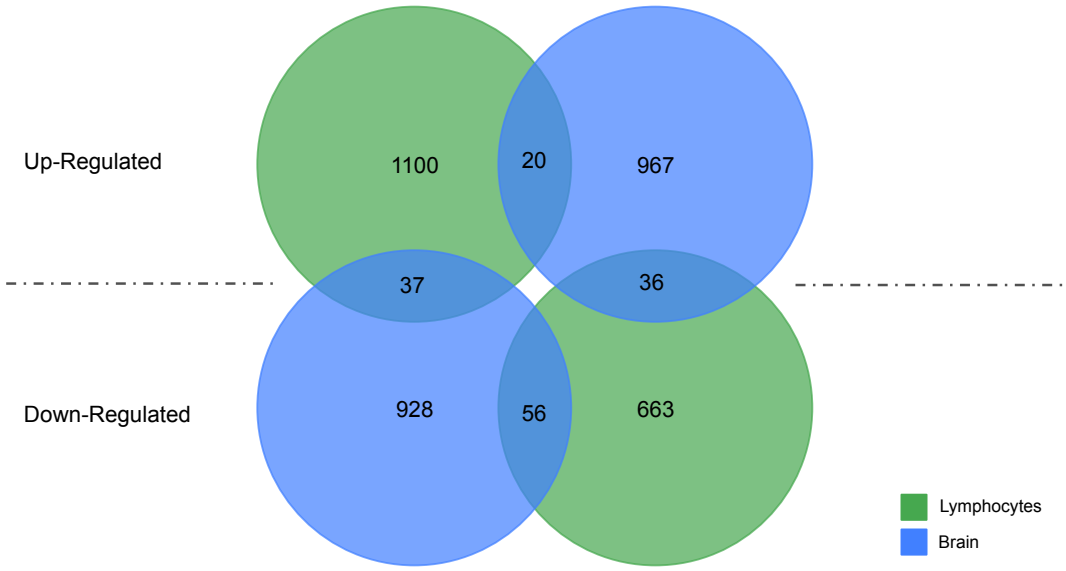

Figure 5-1. Overlap between differentially expressed genes in lymphocytes and brain. Green = lymphocytes (n=2321); Purple = brain (n=2017).

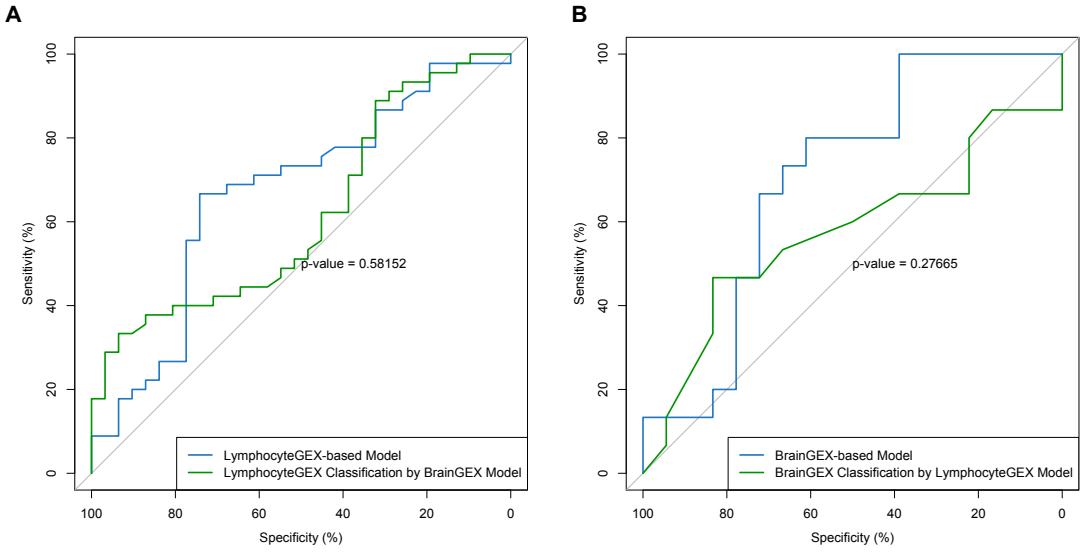

Figure 5-2. Receiver Operating Characteristic (ROC) curves for classification of autism cases and controls using A.) lymphocyte and B.) brain gene expression. A.) Comparison of a classification model built and tested via cross validation with lymphocyte gene expression (blue) (AUC: 0.668, 95% CI: 0.540-0.796) and a classification model built with brain gene expression and tested with lymphocyte gene expression (green) (AUC: 0.618, 95% CI: 0.746-0.489 B.) Comparison of a classification model built and tested via cross validation with brain gene expression (blue) (AUC: 0.704, 95% CI: 0.520-0.888) and a classification model built with lymphocyte gene expression and tested with brain gene expression (green) (AUC: 0.574, 95% CI: 0.785-0.363).

REFERENCES

Addona TA, Shi X, Keshishian H, Mani DR, Burgess M, Gillette MA, Clauser KR, Shen D, Lewis GD, Farrell LA, Fifer MA, Sabatine MS, Gerszten RE, Carr SA (2011) A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. Nature Biotechnology 29, 635-643

Anderson BM, Schnetz-Boutaud NC, Bartlett J, Wotawa AM, Wright HH, Abramson RK, Cuccaro ML, Gilbert JR, Pericak-Vance MA, Haines JL (2009) Examination of association of genes in the serotonin system to autism. Neurogenetics 10, 209-216

Baine MJ, Chakraborty S, Smith LM, Mallya K, Sasson AR, Brand RE, Batra SK (2011) Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. PLoS ONE 6, e17014-e17014

Baron CA, Liu SY, Hicks C, Gregg JP (2006) Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. Journal of Autism and Developmental Disorders 36, 973-982

Bergmann A (2002) Survival signaling goes BAD. Developmental Cell 3, 607-608

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, Bonin M, Riess A, Engels H, Sprengel R, Scherer SW, Rappold GA (2010) Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nature Genetics 42, 489-491

Bowden NA, Weidenhofer J, Scott RJ, Schall U, Todd J, Michie PT, Tooney PA (2006) Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. Schizophrenia Research 82, 175-183

Choi JK, Choi BH, Ha Y, Park H, Yoon SH, Park HC, Park SR (2007) Signal transduction pathways of GM-CSF in neural cell lines. Neuroscience Letters 420, 217-222

Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, Fan JB, April C, Fu XD, Wynshaw-Boris A, Schork NJ, Courchesne E. Age Dependent Brain Gene Expression and Copy Number Anomalies in Autism Suggest Distinct Age-Dependent Pathological Processes. *Submitted*

Du P, Kibbe WA, Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics (Oxford, England) 24, 1547-1548

Fan H, Hegde PS (2005) The transcriptome in blood: challenges and solutions for robust expression profiling. Current Molecular Medicine 5, 3-10

Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. Science Translational Medicine 2, 49ra67-49ra67

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5, R80-R80

Glatt SJ, Chandler SD, Bousman CA, Chana G, Lucero GR, Tatro E, May T, Lohr JB, Kremen WS, Everall IP, Tsuang MT (2009) Alternatively Spliced Genes as Biomarkers for Schizophrenia, Bipolar Disorder and Psychosis: A Blood-Based Spliceome-Profiling Exploratory Study. Current Pharmacogenomics and Personalized Medicine 7, 164-188

Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N, Han M, Liew CC, Tsuang MT (2005) Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci U S A 102, 15533-15538

Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet 39, 1208-1216

Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, Croen LA, Ozonoff S, Hansen R, Pessah IN, Sharp FR (2008) Gene expression changes in children with autism. Genomics 91, 22-29

Groen W, Teluij M, Buitelaar J, Tendolkar I (2010) Amygdala and hippocampus enlargement during adolescence in autism. Journal of the American Academy of Child and Adolescent Psychiatry 49, 552-560

Gupta A, Nagilla P, Le H-S, Bunney C, Zych C, Thalamuthu A, Bar-Joseph Z, Mathavan S, Ayyavoo V (2011) Comparative Expression Profile of miRNA and mRNA in Primary Peripheral Blood Mononuclear Cells Infected with Human Immunodeficiency Virus (HIV-1). PLoS ONE 6, e22730-e22730

Huang NK, Lin YW, Huang CL, Messing RO, Chern Y (2001) Activation of protein kinase A and atypical protein kinase C by A(2A) adenosine receptors antagonizes apoptosis due to serum deprivation in PC12 cells. The Journal of Biological Chemistry 276, 13838-13846

Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, Kaufmann SHE, Ziegler A (2006) Deconfounding microarray analysis - independent measurements of cell type proportions used in a regression model to resolve tissue heterogeneity bias. Methods of Information in Medicine 45, 557-563

Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S, Amiel J, Le Merrer M, Heron D, de Blois MC, Prieur M, Vekemans M, Carter NP, Munnich A, Colleaux L, Philippe A (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J Med Genet 43, 843-849

Kim AY, Bommelje CC, Lee BE, Yonekawa Y, Choi L, Morris LG, Huang G, Kaufman A, Ryan RJ, Hao B, Ramanathan Y, Singh B (2008) SCCRO (DCUN1D1) is an essential component of the E3 complex for neddylation. J Biol Chem 283, 33211-33220

Kolonics A, Apati A, Janossy J, Brozik A, Gati R, Schaefer A, Magocsi M (2001) Activation of Raf/ERK1/2 MAP kinase pathway is involved in GM-CSF-induced proliferation and survival but not in erythropoietin-induced differentiation of TF-1 cells. Cell Signal 13, 743-754

Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD, Edenberg HJ, Tsuang MT, Salomon DR, Nurnberger JI, Jr., Niculescu AB (2009) Identifying blood biomarkers for mood disorders using convergent functional genomics. Molecular Psychiatry 14, 156-174

Leonardson AS, Zhu J, Chen Y, Wang K, Lamb JR, Reitman M, Emilsson V, Schadt EE (2010) The effect of food intake on gene expression in human peripheral blood. Human Molecular Genetics 19, 159-169

Lord C, Rutter M, DiLavore P, Risi S (2001) Autism Diagnostic Observation Schedule (ADOS) manual. Los Angeles, CA: Western Psychological Services

Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, Bishop S, Esler A, Hus V, Oti R, Richler J, Risi S, Lord C (2009) The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. Journal of Autism and Developmental Disorders 39, 1305-1320

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapduram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CEJ, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW (2008) Structural variation of chromosomes in autism spectrum disorder. American Journal of Human Genetics 82, 477-488

Mas C, Bourgeois F, Bulfone A, Levacher B, Mugnier C, Simonneau M (2000) Cloning and expression analysis of a novel gene, RP42, mapping to an autism susceptibility locus on 6q16. Genomics 65, 70-74

Mhawech P (2005) 14-3-3 proteins--an update. Cell Research 15, 228-236

Miao F, Wu X, Zhang L, Riggs AD, Natarajan R (2008) Histone methylation patterns are cell-type specific in human monocytes and lymphocytes and well maintained at core genes. Journal of Immunology (Baltimore, Md.: 1950) 180, 2264-2269

Mohammad NS, Jain JMN, Chintakindi KP, Singh RP, Naik U, Akella RRD (2009) Aberrations in folate metabolic pathway and altered susceptibility to autism. Psychiatric Genetics 19, 171-176

Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M, Steindler C, Pellegrini S, Schanen NC, Warren ST, Geschwind DH (2007) Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Human Molecular Genetics 16, 1682-1698

Noelker C, Hampel H, Dodel R (2011) Blood-based protein biomarkers for diagnosis and classification of neurodegenerative diseases: current progress and clinical potential. Molecular Diagnosis & Therapy 15, 83-102

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature Genetics 43, 585-589

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. Nature Neuroscience 11, 1271-1282

Palmer C, Diehn M, Alizadeh AA, Brown PO (2006) Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 7, 115-115

Perlis RH (2011) Translating biomarkers to clinical practice. Molecular Psychiatry

Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M (1999) Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. Hum Mol Genet 8, 805-812

Pierce K, Carter C, Weinfeld M, Desmond J, Hazin R, Bjork R, Gallagher N (2011) Detecting, studying, and treating autism early: the one-year well-baby check-up approach. The Journal of Pediatrics 159, 458-465.e456-458-465.e456

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI, Jr., Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466, 368-372

Piton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, Henrion E, Laurent S, Noreau A, Thibodeau P, Karemera L, Spiegelman D, Kuku F, Duguay J, Destroismaisons L, Jolivet P, Cote M, Lachapelle K, Diallo O, Raymond A, Marineau C, Champagne N, Xiong L, Gaspar C, Riviere JB, Tarabeux J, Cossette P, Krebs MO, Rapoport JL, Addington A, Delisi LE, Mottron L, Joober R, Fombonne E, Drapeau P, Rouleau GA (2011) Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol Psychiatry 16, 867-880

Qiu RG, Abo A, Steven Martin G (2000) A human homolog of the C. elegans polarity determinant Par-6 links Rac and Cdc42 to PKCzeta signaling and cell transformation. Current Biology: CB 10, 697-707

Radmacher MD, McShane LM, Simon R (2002) A paradigm for class prediction using gene expression profiles. Journal of Computational Biology: A Journal of Computational Molecular Cell Biology 9, 505-511

Ramser EM, Wolters G, Dityateva G, Dityatev A, Schachner M, Tilling T (2010) The 14-3-3zeta protein binds to the cell adhesion molecule L1, promotes L1 phosphorylation by CKII and influences L1-dependent neurite outgrowth. PLoS ONE 5, e13462

Rebola N, Canas PM, Oliveira CR, Cunha RA (2005) Different synaptic and subsynaptic localization of adenosine A2A receptors in the hippocampus and striatum of the rat. Neuroscience 132, 893-903

Rockett JC, Burczynski ME, Fornace AJ, Herrmann PC, Krawetz SA, Dix DJ (2004) Surrogate tissue analysis: monitoring toxicant exposure and health status of inaccessible tissues through the analysis of accessible tissues and cells. Toxicology and Applied Pharmacology 194, 189-199

Rollins B, Martin MV, Morgan L, Vawter MP (2010) Analysis of whole genome biomarker expression in blood and brain. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics 153B, 919-936

Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT, Sehnert AJ, Yau M, Kraus WE, Newby LK, Schwartz RS, Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A, Winn ME, Schork NJ, Topol EJ (2010) Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. Annals of Internal Medicine 153, 425-434

Sanda M, Ohara N, Kamata A, Hara Y, Tamaki H, Sukegawa J, Yanagisawa T, Fukunaga K, Kondo H, Sakagami H (2010) Vezatin, a potential target for ADP-ribosylation factor 6, regulates the dendritic formation of hippocampal neurons. Neuroscience Research 67, 126-136

Sarachana T, Zhou R, Chen G, Manji HK, Hu VW (2010) Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. Genome Med 2, 23

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology 7, 3-3

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545-15550

Sullivan PF, Fan C, Perou CM (2006) Evaluating the comparability of gene expression in blood and brain. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics 141B, 261-268

Svenningsson P, Hall H, Sedvall G, Fredholm BB (1997) Distribution of adenosine receptors in the postmortem human brain: an extended autoradiographic study. Synapse (New York, N.Y.) 27, 322-335

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, Feuk L, Qian C, Bryson SE, Jones MB, Marshall CR, Scherer SW, Vieland VJ, Bartlett C, Mangin LV, Goedken R, Segre A, Pericak-Vance MA, Cuccaro ML, Gilbert JR, Wright HH, Abramson RK, Betancur C, Bourgeron T, Gillberg C, Leboyer M, Buxbaum JD, Davis KL, Hollander E, Silverman JM, Hallmayer J, Lotspeich L, Sutcliffe JS, Haines JL, Folstein SE, Piven J, Wassink TH, Sheffield V, Geschwind DH, Bucan M, Brown WT, Cantor RM, Constantino JN, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Spence S, State M, Tanzi RE, Coon H, Dawson G, Devlin B, Estes A, Flodman P, Klei L, McMahon WM, Minshew N, Munson J, Korvatska E, Rodier PM, Schellenberg GD, Smith M, Spence MA, Stodgell C, Tepper PG, Wijsman EM, Yu CE, Roge B, Mantoulan C, Wittemeyer K, Poustka A, Felder B, Klauck SM, Schuster C, Poustka F, Bolte S, Feineis-Matthews S, Herbrecht E, Schmotzer G, Tsiantis J, Papanikolaou K, Maestrini E, Bacchelli E, Blasi F, Carone S, Toma C, Van Engeland H, de Jonge M, Kemner C, Koop F, Langemeijer M, Hijmans C, Staal WG, Baird G, Bolton PF, Rutter ML, Weisblatt E, Green J, Aldred C, Wilkinson JA, Pickles A, Le Couteur A, Berney T, McConachie H, Bailey AJ, Francis K, Honeyman G, Hutchinson A, Parr JR, Wallace S, Monaco AP, Barnby G, Kobayashi K, Lamb JA, Sousa I, Sykes N, Cook EH, Guter SJ, Leventhal BL, Salt J, Lord C, Corsello C, Hus V, Weeks DE, Volkmar F, Tauber M, Fombonne E, Shih A, Meyer KJ (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39, 319-328

Teschendorff AE, Caldas C (2008) A robust classifier of high predictive value to identify good prognosis patients in ER-negative breast cancer. Breast Cancer Research: BCR 10, R73-R73

Tsuang MT, Nossova N, Yager T, Tsuang M-M, Guo S-C, Shyu KG, Glatt SJ, Liew CC (2005) Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics 133B, 1-5

Vawter MP, Ferran E, Galke B, Cooper K, Bunney WE, Byerley W (2004) Microarray screening of lymphocyte gene expression differences in a multiplex schizophrenia pedigree. Schizophrenia Research 67, 41-52

Villa C, Venturelli E, Fenoglio C, Clerici F, Marcone A, Benussi L, Gallone S, Scalabrini D, Cortini F, Serpente M, Martinelli Boneschi F, Cappa S, Binetti G, Mariani C, Rainero I, Giordana MT, Bresolin N, Scarpini E, Galimberti D (2009) DCUN1D1 is a risk factor for frontotemporal lobar degeneration. European Journal of Neurology: The Official Journal of the European Federation of Neurological Societies 16, 870-873

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PMA, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SFA, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459, 528-533

Wei CJ, Li W, Chen J-F (2011) Normal and abnormal functions of adenosine receptors in the central nervous system revealed by genetic knockout studies. Biochimica Et Biophysica Acta 1808, 1358-1379

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO (2003) Individuality and variation in gene expression patterns in human blood. Proceedings of the National Academy of Sciences of the United States of America 100, 1896-1901

Yang HH, Hu N, Wang C, Ding T, Dunn BK, Goldstein AM, Taylor PR, Lee MP (2010) Influence of genetic background and tissue types on global DNA methylation patterns. PLoS ONE 5, e9355-e9355

CHAPTER 6

Conclusions and Future Directions

The aim of the studies described in this dissertation was to explore the technical aspects and biological relevance of microarray gene expression profiling of peripheral whole blood in regards to neural-based diseases. The studies were motivated by the technical difficulties associated with assaying peripheral whole blood via microarray, the difficulties associated with directly comparing blood and brain gene expression profiles, and the general lack of understanding in regards to the ability of blood-based gene expression profiling to capture biologically relevant neural-based disease profiles. The study of blood gene expression profiles is likely to be important to the development of expression-based biomarkers for neuropsychiatric, neurodegenerative, and neurodevelopmental diseases given the difficulties associated with collecting the large number of human brains necessary for clinical validation of such biomarkers. The chapters in this dissertation tackle various aspects associated with the processing and evaluation of microarray gene expression profiles of peripheral whole blood in mouse and humans. Here I provide an overview of the main findings as well as discuss limitations and future directions.

MAIN FINDINGS

The evaluation of a globin reduction method using microarray-based blood gene expression profiling in mouse highlighted the need for and the importance of removing or reducing globin transcripts in peripheral whole blood prior to microarray analysis (Chapter 2). Comparing blood gene expression profiles from before and after globin reduction, the removal of globin transcripts was found to improve detection sensitivity of low abundance transcripts thus significantly improving the ability to evaluate biological pathways and disease networks via whole blood. The need to

reduce globin transcripts in the analysis of anxiety-related phenotypes in particular was emphasized by the significant increase in the number of Schizophrenia network objects identified following globin reduction.

Given the positive effect of globin reduction on microarray-based gene expression profiling of mouse whole blood, globin reduced blood gene expression profiles were utilized in conjunction with neural tissue gene expression profiles in the evaluation of anxiety-related behavioral phenotypes in mice (Chapter 3). Several recent studies have compared blood and brain gene expression profiles with conflicting results as to the potential of blood as a surrogate tissue for blood (Cai et al. 2010, Davies et al. 2009, Jasinska et al. 2009). However, these studies only assess natural variation and heritability. Although naturally occurring variation and patterns may fail to be well correlated between blood and brain such as seen by Cai et al., it remains unclear whether genes and pathways associated with disease in blood will reflect those genes and pathways associated with disease in brain. Here we went beyond strain- and tissue-specific variation to evaluate genes and pathways associated with anxiety-related phenotypes in mice. We concluded blood gene expression profiles were able to capture only a small portion of the total trait-associated genes and enriched pathways identified in brain. Despite this finding, the genes and pathways associated with behavioral phenotypes in blood were highly enriched for biologically relevant genes and pathways suggesting blood is a viable surrogate tissue.

Next, blood gene expression profiles were assessed in the context of human samples. Human, peripheral whole blood is known to benefit from the reduction of globin transcripts similar to the effected reported here in the context of mice (Tian et

al. 2009, Field et al. 2007). The evaluation of a microarray target preparation method (WG-DASL) with the potential to eliminate the need for globin reduction confirmed the benefits of reducing globin transcripts (Chapter 4) suggesting the elimination of globin-specific probes from sample amplification is not sufficient to improve detection sensitivity.

The correlation between human lymphocyte and brain gene expression profiles associated with autism was assessed in a similar manner as behavioral traits in mice (Chapter 5). Approximately 4% of the genes dysregulated in the brain were dysregulated in lymphocytes and many of the same pathways dysregulated in brain were also significantly enriched in lymphocytes. We also found many of these genes (33%) were identified in previous genetic studies of autism. These results confirm the ability to identify neurobiologically relevant genes using blood-derived cells. Furthermore, lymphocyte gene expression profiles were able to classify disease state at a level similar to brain gene expression.

CONCLUSIONS AND FUTURE DIRECTIONS

In this work, I assessed the potential of blood as a surrogate tissue for the analysis of neural-based diseases. The correlation between blood and brain was ultimately assessed comparing significantly associated behavioral trait/disease genes and significantly enriched pathways and disease networks identified in blood and brain. We conclude whole blood or blood-derived cells reflect biologically relevant disease profiles despite capturing only a small fraction of the whole picture. However, these studies are not without their limitations, particularly in regards to sample size and microarray processing.

Microarray studies are especially prone to batch effects (i.e. time of sample collection, time of microarray processing, array manufacture date, etc.) and other technically-induced variation (i.e. RNA processing, scanning intensity, location of the probe on the array, etc.). This becomes exceptionally relevant when batch effects are correlated with the phenotype of interest. In the mouse gene expression profiles studied here, tissue type was highly correlated with the time of sample collection and microarray processing. Due to this correlation, correcting for batch (Johnson et al. 2007) would remove variation associated with both time of microarray processing as well as tissue type. The batch effects associated with tissue type may have also played a part in the difference in the number of significantly associated genes in blood as compared to brain (7%-35%) and hence the number of genes that overlap or the number of pathways capable of being significantly enriched in blood. The small biological differences associated with natural variations in inbred mouse strain behavioral phenotypes may largely have been overshadowed by differences associated with technical processing (Bryant et al. 2011).

Genome-wide transcriptional profiling of whole blood is also hampered by its heterogeneous nature. The effects of this heterogeneity maybe reflected in the larger overlap between blood and brain using lymphocytes as compared to globin-reduced whole blood. While brain tissues can be finely dissected to ensure a rather homogeneous cell population from the start, blood cells must first be collected and immediately fractionated or stored and subjected to further downstream processing at a later date or immortalized as cell lines. The effects of fractionation and storage on gene expression profiles have been well studied (Debey-Pascher et al. 2011, Debey et al. 2004). At the same time, various computational methods have been developed

in an attempt to identify cell populations (Bolen et al. 2011, Grigoryev et al. 2010). However, the most effective approach to dealing with different cell populations in whole blood maybe running a Complete Blood Count (CBC) test for each sample collected.

Sample size is also a problem given the variability associated with genome-wide transcriptional profiling and behavioral phenotype testing. Increasing the number of mice per tissue, behavioral phenotype, and strain should result in decreased variability due to sampling differences and technical variation thus improving the ability to detect smaller changes in gene expression and behavioral phenotypes across the different mouse strains. Our analyses may also be improved by limiting the study to mice on the extreme ends of the phenotype spectrum or by increasing the number of strains thus expanding the range of the behavioral phenotypes studied.

In the end, genome-wide transcriptome analyses will benefit from the development of more sensitive assays such as RNA sequencing (Ozsolak et al. 2011). RNA sequencing overcomes the limited dynamic range of microarray platforms and hence the contributing factor to the need for reducing globin transcripts in whole blood. As RNA sequencing continues to advance, whole blood gene expression studies will benefit from the ability to assess not only mRNA levels but also the ability to more accurately predict alternative splicing events. Alternative splicing is widespread in the brain (Lin et al. 2011, Boutz et al. 2007) and may be one of the reason blood gene expression profiles do not significantly overlap with brain gene expression profiles.

REFERENCES

Bolen CR, Uduman M, Kleinstein SH (2011) Cell subset prediction for blood genomic studies. BMC Bioinformatics 12, 258

Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M, Jr., Black DL (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev 21, 1636-1652

Bryant PA, Smyth GK, Robins-Browne R, Curtis N (2011) Technical variability is greater than biological variability in a microarray experiment but both are outweighed by changes induced by stimulation. PLoS ONE 6, e19556-e19556

Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R, Horvath S (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11, 589-589

Davies MN, Lawn S, Whatley S, Fernandes C, Williams RW, Schalkwyk LC (2009) To What Extent is Blood a Reasonable Surrogate for Brain in Gene Expression Studies: Estimation from Mouse Hippocampus and Spleen. Front Neurosci 3

Debey S, Schoenbeck U, Hellmich M, Gathof BS, Pillai R, Zander T, Schultze JL (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. The Pharmacogenomics Journal 4, 193-207

Debey-Pascher S, Hofmann A, Kreusch F, Schuler G, Schuler-Thurner B, Schultze JL, Staratschek-Jox A (2011) RNA-stabilized whole blood samples but not peripheral blood mononuclear cells can be stored for prolonged time periods prior to transcriptome analysis. The Journal of Molecular Diagnostics: JMD 13, 452-460

Field LA, Jordan RM, Hadix JA, Dunn MA, Shriver CD, Ellsworth RE, Ellsworth DL (2007) Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. Clinical Biochemistry 40, 499-502

Grigoryev YA, Kurian SM, Avnur Z, Borie D, Deng J, Campbell D, Sung J, Nikolcheva T, Quinn A, Schulman H, Peng SL, Schaffer R, Fisher J, Mondala T, Head S, Flechner SM, Kantor AB, Marsh C, Salomon DR (2010) Deconvoluting post-transplant immunity: cell subset-specific mapping reveals pathways for activation and expansion of memory T, monocytes and B cells. PLoS ONE 5, e13358

Jasinska AJ, Service S, Choi O-w, DeYoung J, Grujic O, Kong S-y, Jorgensen MJ, Bailey J, Breidenthal S, Fairbanks LA, Woods RP, Jentsch JD, Freimer NB (2009) Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits. Human Molecular Genetics 18, 4415-

4427

Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118-127

Lin L, Shen S, Jiang P, Sato S, Davidson BL, Xing Y (2010) Evolution of alternative splicing in primate brain transcriptomes. Hum Mol Genet 19, 2958-2973

Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12, 87-98

Tian Z, Palmer N, Schmid P, Yao H, Galdzicki M, Berger B, Wu E, Kohane IS (2009) A Practical Platform for Blood Biomarker Study by Using Global Gene Expression Profiling of Peripheral Whole Blood. PLoS ONE 4, e5157-e5157