in press
Biodemography and Social Biology

Genome-Wide Profiling of RNA from Dried Blood Spots: Convergence with Bioinformatic Results Derived from Whole Venous Blood and Peripheral Blood Mononuclear Cells

Thomas W. McDade[1,2,3], Kharah Ross[4], Ruby Fried[1], Jesusa M. G. Arevalo[5], Jeffrey Ma[6], Gregory E. Miller[2,7], and Steve W. Cole[6,8,9]

Author Affiliations
[1]Department of Anthropology, Northwestern University, Evanston, IL
[2]Cells to Society (C2S): The Center on Social Disparities and Health, Institute for Policy Research, Northwestern University, Evanston, IL
[3]Canadian Institute for Advance Research, Program in Child and Brain Development
[4]Psychology Department, University of California, Los Angeles, Los Angeles, CA
[5]Department of Medicine, Division of Hematology-Oncology, UCLA School of Medicine, Los Angeles, CA
[6]Cousins Center for Psychoneuroimmunology, Department of Psychiatry & Biobehavioral Sciences, 300 Medical Plaza, Los Angeles, CA
[7]Department of Psychology, Northwestern University, Evanston, IL
[8]Department of Psychiatry and Biobehavioral Sciences, University of California-Los Angeles, Los Angeles, California
[9]Department of Medicine, University of California-Los Angeles, Los Angeles, California

Acknowledgements
We thank the UCLA Neuroscience Genomics Core Laboratory for assay services. This project was supported by Grant Number 1R01HD074765 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and P30 AG017265 from the National Institute of Aging of the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Running Head: Genome-Wide RNA Profiling from Dried Blood Spots

Abstract

Genome-wide transcriptional profiling has emerged as a powerful tool for analyzing biological mechanisms underlying social gradients in health, but utilization in population-based studies has been hampered by logistical constraints and costs associated with venipuncture blood sampling. Dried blood spots (DBS) provide a minimally-invasive, low-cost alternative to venipuncture, and in this paper we evaluate how closely the substantive results from DBS transcriptional profiling correspond to those derived from parallel analyses of gold-standard venous blood samples (PAXgene whole blood and peripheral blood mononuclear cells; PBMC). Analyses focused on differences in gene expression between African-Americans and Caucasians in a community sample of 82 healthy adults (age 18-70 years; mean 35). Across 19,679 named gene transcripts, DBS-derived values correlated r = .85 with both PAXgene and PBMC values. Results from bioinformatics analyses of gene expression derived from DBS samples were concordant with PAXgene and PBMC samples in identifying increased Type I interferon signaling and up-regulated activity of monocytes and natural killer (NK) cells in African-Americans relative to Caucasian participants. These findings demonstrate the feasibility of DBS in field-based studies of gene expression, and encourage future studies of human transcriptome dynamics in larger, more representative samples than are possible with clinic- or lab-based research designs.

Key words: dried blood spots; biomeasures; genetics; biodemography

Introduction

The measurement of gene expression (messenger RNA; mRNA) in circulating blood allows investigators to study the pathways through which lifestyle and contextual factors impact physiological function and health (Finch et al. 2001; Weinstein et al. 2008). Venipuncture blood sampling is the current clinical standard for blood biomarker assessment, but the costs, participant burden, regulatory constraints, and logistics associated with venipuncture are major barriers to community-based research on the social/contextual determinants of health. Dried blood spots (DBS)—drops of whole blood collected on filter paper following a simple finger stick—represent a low cost, "field-friendly" alternative that allows investigators to collect physiological information from large numbers of participants in naturalistic settings, and to integrate this information with rich contextual data in ways not possible with clinic-based research designs (McDade et al. 2007; Mei et al. 2001). Several studies have shown that it is feasible to assay gene expression profiles in DBS samples (Ho et al. 2013; Khoo et al. 2011; Pappas et al. 2015; Resau et al. 2012; Slaughter et al. 2013; Wei et al. 2014). However, little is known about how well the gene expression profiles derived from DBS samples correspond to those from "gold standard" venipuncture blood samples.

Genome-wide transcriptional profiling (i.e., assaying all ~20,000 genes in the human genome) is particularly powerful when used in conjunction with bioinformatics techniques. The goal of these techniques is to extract biological themes based upon correlated changes in the expression of hundreds or thousands of genes. By doing so, bioinformatic analyses can provide insights into the underlying "genomic programs" that are activated by social conditions and the cellular and molecular signaling pathways that mediate these programs (Cole 2010, 2013, 2014; Miller, Chen & Cole 2009; Miller & Cole, 2010).

One example of the utility of bioinformatics is provided by a line of research analyzing the molecular pathways that mediate the effects of adverse social conditions on health outcomes. Over the past five years, studies have repeatedly identified a conserved transcriptional response to adversity (CTRA) that involves altered expression of three broad groups of genes: up-regulated expression of pro-

inflammatory genes, down-regulated expression of genes involved in innate antiviral responses (Type I interferon response genes), and down-regulated expression of genes involved in antibody synthesis (see review by Cole 2014). Bioinformatics analyses discovered this general pattern of gene regulation by assessing the transcription factors underlying observed changes in gene expression, such as increased activity of the pro-inflammatory transcription factor, nuclear factor kappa B (NF-κB), and decreased activity of interferon response factors (IRFs) (Cole et al. 2007, 2011; Fredrickson et al. 2013; Miller et al. 2008, 2009, 2014; Powell et al. 2013). Additional bioinformatics analyses identified monocytes and dendritic cells as the key cellular mediators of these effects within the overall pool of white blood cells (which contains a variety of other cell types such as B cells, CD4+ and CD8+ T cells, and Natural Killer cells) (Cole et al. 2011; Powell et al. 2013). The identification of these functional, regulatory, and cellular themes provided converging lines of insight into the biological mechanisms of the transcriptome profile, and ultimately allowed researchers to specify a set of 53 genes that could serve as canonical markers of the CTRA transcriptome profile (Cole et al. 2015; Fredrickson et al. 2013; 2015; Vedhara et al. 2015).

Collectively, these alterations in the genomic activity of immune cells are thought to explain why social adversity increases vulnerability to health problems that involve excessive inflammation (e.g., coronary heart disease) and compromised host resistance (e.g., upper respiratory infections) (Miller, Chen, & Cole, 2009). However, current CTRA research has been conducted only in the context of affluent western industrialized (so-called WEIRD; Henrich et al. 2010) populations that support the biomedical research infrastructure required for collection and processing of venipuncture blood samples. Substantive insights into the evolutionary origin and adaptive function of the CTRA genomic program would come from analyzing gene expression in other cultural and ecological conditions that capture a broader range of human variation, and that may be more representative of the human genome's ancestral environment.

The objective of this paper is to determine how closely bioinformatics results derived from DBS samples correspond to results from gold-standard venous blood samples that capture either whole blood or peripheral blood mononuclear cells (PBMC; the standard in immunology, involving centrifugation to

remove red blood cells and to produce samples containing lymphocytes, monocytes, and some residual granulocytes). As a framework for making these comparisons, we analyze expression of the CTRA transcriptome profile and related bioinformatic inferences of transcription factor activation and cellular origin across two groups of participants: self-identified African-Americans and Caucasians drawn from a community-based setting. In the United States there are widespread and persistent disparities in health by race (Meyer et al., 2013). Relative to Caucasians, African-Americans experience higher rates of morbidity and mortality in most disease categories (Braveman et al., 2010, Williams 2012), and there is preliminary evidence to suggest that alterations in immune cell function may underlie some of these disparities (Huang et al. 2011). The results of our study identify a consistent pattern of African-American/Caucasian differences in gene expression, and show that individual gene expression estimates and bioinformatic inferences derived from DBS RNA samples correspond well with those derived from venipuncture blood samples. These findings demonstrate the feasibility of using DBS as a minimally-invasive alternative to venous blood sampling for field-based studies of gene expression, and should encourage the application of DBS methods in larger, more representative samples than are possible with clinic- or lab-based research designs.

Methods

Sample

We recruited a convenience sample of 82 adults from the greater Chicago, Illinois area. Recruitment was done via word of mouth, flyers posted in community settings, and advertisements on Craigslist. To qualify for the study, participants had to be 18-70 years of age, fluent in English, and not pregnant. Participants received a $25 gift card upon completion of the study. The protocol was approved by Northwestern University's Institutional Review Board.

Protocol

After participants had given written informed consent, they remained seated at rest for approximately 10 minutes, after which a phlebotomist collected DBS samples and venous blood samples. DBS samples were collected by pricking the tip of the middle digit with a sterile disposable lancet, and then allowing drops of blood to wick onto filter paper cards (Whatman #903, GE Healthcare, Pascataway, NJ) (McDade et al. 2007). The cards were allowed to dry overnight at room temperature, and were then frozen at -20°C for batch processing. Venipuncture samples were collected from the antecubital fossa with a butterfly needle. Five mL of whole blood were first drawn into EDTA-coated vacutainer tubes. Immediately following collection, five drops of whole blood (50 uL each) were transferred by pipet to filter paper to make DBS samples. An additional 2.5 ml of blood was drawn into PAXgene Blood RNA tubes (Qiagen, Valencia, CA), and then frozen at -20°C until batch processing. For PBMC samples, 8 ml of venous blood were drawn into Cell Preparation Tubes (Becton, Dickinson and Company, Franklin Lakes, NJ), which were centrifuged for 20 minutes at 1800 x g. The PBMC layer was then removed by pipette, PMBCs were pelleted by centrifugation in a 1.5 ml microcentrifuge tube, supernatant fluid was cleared by aspiration, and the remaining cell pellet was lysed in RNA stabilization buffer (RLT Plus, Qiagen), homogenized (QiaShredder, Qiagen), and frozen at -80 °C until batch processing.

Following blood collection, participants were measured for weight and height to ascertain body mass index (BMI, kg/m^2), and then completed questionnaire measures of demographic characteristics, medical history, and lifestyle factors. Data on medical history was obtained with items from the Medical Condition Survey of the National Health and Nutrition Examination Study, and included presence of and treatment for major diseases (CDC 2009). Smoking and alcohol use were ascertained with validated self-reports, and quantified as cigarettes smoked per day and average number of alcoholic drinks per week. A drink was considered a bottle or can of beer, a glass of wine, or a shot of hard liquor (Miller et al. 2002).

RNA Extraction and Profiling

All samples were shipped on dry ice to the UCLA Social Genomics Core Laboratory for RNA extraction, quality assurance assays, and transcriptome profiling. Total RNA was extracted from

PAXgene RNA tubes and PBMC pellets in RLT using an automated nucleic acid extraction system (Qiagen QIAcube) with the manufacturer's standard protocol and reagents (PAXgene Blood RNA Kit IVD and RNeasy Mini QIAcube Kit, respectively), including DNAse treatment, yielding an 80 μL elution volume. RNA mass and integrity were tested using Nanodrop ND1000 and Agilent 2200 TapeStation instruments.

Total RNA was extracted from DBS samples produced from EDTA venous blood in order to conserve finger stick DBS samples for other validation purposes. However, a subset of N=10 matched EDTA DBS and finger stick DBS samples were analyzed to confirm concordance of gene expression estimates ($r = +.93$, $p < .0001$). Total RNA was extracted from DBS samples by cutting 2-4 blood spots from the 5-spot filter paper for each sample (using a separate razor blade for each sample to prevent contamination), depositing all filter paper spots from each sample into 370 μL of RLT in an RNAse-free sterile 1.5 ml microcentrifuge tube, incubating the tube for 30 min at 37° C with agitation (1000 rpm), transferring tube contents (including filter paper) into a QIAshredder column for 60 sec microcentrifugation at maximum speed, after which the 360 μL of remaining RLT was processed through the QIAcube nucleic acid extraction system using RNeasy Micro Kit reagents, the manufacturer's standard operating protocol (including DNAse treatment), and a 20 μL elution volume. No attempt was made to assess RNA mass or integrity in DBS samples because RNA quantity was expected to fall below the lower limit of detection for available assay systems.

All available DBS RNA and 40 ng of RNA from PAXgene and PBMC samples were submitted to cDNA amplification using the NuGEN Ovation PicoSL WTA System, and 100 ng of the resulting cDNA was fragmented and fluorescently labeled using the NuGEN Encore BiotinIL Module for Illumina Whole Genome Expression BeadChips. The resulting fluorescent target sample was assayed using Illumina Human HT-12 v4.0 BeadChips, following the manufacturer's standard protocol for cDNA hybridization, with scanning on an Illumina iScan instrument in the UCLA Neuroscience Genomics Core following the manufacturer's standard protocol. The raw data are deposited in Gene Expression Omnibus (Accession # GSE75511).

Statistical Analyses

Transcriptome data analysis followed methods employed and validated in our previous research. Briefly, raw gene expression data were quantile-normalized and $\log_2$ transformed for analysis. Preliminary analyses examined cross-sampling mode consistency of expression values across all 19,679 named gene transcripts assayed by the HT-12 v4 (i.e., excluding unnamed open reading frames and provisional *LOC*, *HS.*, and *KIAA* transcripts). Differential gene expression was assessed using standard linear statistical models including effects of age (years), gender (1/0), BMI (kg/m^2), history of heavy alcohol consumption (1/0), history of smoking (1/0), educational attainment (years) and a four-category race/ethnicity variable (1/0 indicators for African-American, Asian-American, and Other Minority, with Caucasian serving as the reference category). In secondary analyses, we also adjusted for relative proportions of monocytes, CD4+ and CD8+ T lymphocytes, B lymphocytes, and Natural Killer cells by including as covariates mRNAs encoding eight canonical markers of these cell types (*CD14*, *CD3D*, *CD3E*, *CD4*, *CD8A*, *CD19*, *FCGR3A*/CD16, *NCAM1*/CD56). Differentially expressed genes were defined by ≥1.20-fold difference between Caucasians and African-Americans. No gene-specific statistical testing or False Discovery Rate analyses were performed because previous research has found these procedures to yield less replicable results than does mapping differentially expressed genes based on point-estimates of fold-difference (Cole et al. 2003; Shi et al. 2008; Witten & Tibshirani 2007), especially when substantive interest lies in higher-order bioinformatics results derived from lists of differentially expressed genes (Fredrickson et al. 2013). As such, lists of differentially expressed genes reported here serve only as point-estimate inputs into higher-order bioinformatics analyses and should not be interpreted as statistically significant at the level of the individual gene transcript.

Three types of higher-order bioinformatics analysis were conducted in parallel across gene expression data derived from DBS, PBMC, and PAXgene (whole blood) samples. In each case, analyses tested whether African-American/Caucasian differences identified based on DBS samples corresponded to results derived from the PBMC "gold standard" and to results from PAXgene whole blood samples. The latter is arguably a more appropriate comparison point because the PAXgene system captures the

same total blood cell population as does DBS. In all analyses, standard errors for observed point estimates of bioinformatics statistics were derived from 200 cycles of bootstrap resampling of linear model residual vectors (accounting for potential correlation among residuals across transcripts) (Efron & Tibshirani 1993; Vedhara, et al. 2015).

The first bioinformatics analysis sought to determine whether African-Americans showed altered expression of the CTRA gene expression profile relative to Caucasians. Analyses used an a priori-defined contrast among differential gene expression estimates for 53 canonical CTRA indicator genes (Cole et al. 2015; Fredrickson et al. 2013; 2015; Vedhara et al. 2015), including 19 pro-inflammatory genes (*IL1A, IL1B, IL6, IL8, TNF,PTGS1, PTGS2, FOS, FOSB, FOSL1, FOSL2, JUN, JUNB, JUND,NFKB1, NFKB2, REL, RELA, RELB*), 28 genes involved in Type I interferon response (*GBP1, IFI16, IFI27, IFI27L1-2, IFI30, IFI35, IFI44, IFI44L, IFI6, IFIH1, IFIT1-3, IFIT5,IFIT1L, IFITM1-3, IFITM4P, IFITM5, IFNB1, IRF2, IRF7-8, MX1-2, OAS1-3, OASL*), and three genes involved in antibody synthesis (*IGJ, IGLL1, IGLL3*). Parameter estimate signs were reversed for the antiviral and antibody-related gene sets to reflect their inverse relationship to the overall CTRA profile (Cole et al. 2015; Fredrickson et al. 2013; 2015; Vedhara et al. 2015)

The second bioinformatics analysis sought to test the activity of *a priori* specified transcription control pathways as potential mechanisms of the observed African-American/Caucasian differences in gene expression using a 2-sample variant of the Transcription Element Listening System (TELiS; Cole et al., 2005). TELiS quantified the prevalence of specific transcription factor-binding motifs (TFBMs) for pro-inflammatory transcription factors (NF-κB, GATA, and AP-1), Type I interferon response factors (ISRE, STAT), and neuroendocrine-related transcription factors (CREB, GR) within promoters of all genes showing ≥1.20-fold up-regulation in African-Americans vs. Caucasians compared to all genes showing ≥1.20-fold down-regulation in African-Americans vs. Caucasians. We analyzed aggregate (log) prevalence ratios for TFBMs in up- vs. down-regulated promoters pooled across 9 alternative technical specifications involving parametric variations of promoter length (-300 bp, -600 bp, -1000 to +200 bp) and TFBM detection stringency (MatSim .80, .90, .95) (Cole et al. 2005).

The third bioinformatics analysis sought to identify specific cell types that contribute to the empirically observed differences in gene expression by applying Transcript Origin Analysis (TOA) (Cole et al., 2011) to all genes showing ≥1.20-fold difference in average expression between African-Americans and Caucasians. TOA tested for significant over-representation of genes preponderantly expressed by a specific subset of PBMCs (i.e., monocytes, dendritic cells, CD4+ T cells, CD8+ T cells, B cells, or Natural Killer cells) using cell-specific reference transcriptomes derived from previous research (Su et al. 2004).

Results

Characteristics of the 82 participants are described in Table 1. As can be seen, the sample was diverse with respect to age, gender, race/ethnic background, and educational attainment. As a group they were in relatively good health, with only three heavy smokers (3.7 percent), and a mean BMI in the slightly overweight range. Almost half the sample (46.3 percent) had BMI in the normal range (under 25); the others were overweight (BMI 25 to <30; 32.9 percent) or obese (BMI 30 or greater; 20.7 percent). There was no history of stroke, diabetes, or HIV/AIDS in the sample. One participant had a previous myocardial infarction (13 years prior), and four participants had a history of cancer (none in active treatment during the past year). One participant declined to provide data on race/ethnicity, resulting in a final analyzed sample of N = 81 in analyses of differences in gene expression across race/ethnicity.

Individual gene-level correspondence

In preliminary analyses, we assessed the correlation between individual gene expression values derived from DBS, PAXgene, and PBMC samples. Across the 19,679 named gene transcripts assayed in 82 participants, DBS-derived values correlated $r = +.85$ with PAXgene whole blood samples and $r = +.85$ with PBMC samples (both $p < .0001$). This correlation compared reasonably well with the $r = +.92$ correlation between the gold standard PAXgene and PBMC samples. For each sample type, we also estimated differences in expression of each gene across African-American and Caucasian participants,

controlling for age, gender, BMI, alcohol and cigarette use, and educational attainment. DBS-based estimates of differential gene expression correlated $r = +.48$ with those derived from PAXgene ($p < .0001$), which compares favorably with the $r = +.31$ ($p < .0001$) correlation between African-American vs. Caucasian gene expression differences estimated using PAXgene vs. PBMC. (Consistency of all difference estimates fell below the $r = .85$ consistency of single gene expression values due to the fact that the variance of a difference score is inherently greater than the variance of its constituent components.) Reasonable consistency at the level of individual gene expression values and differential expression estimates suggested that it would be viable to pursue this study's primary aim of estimating the consistency of higher-order bioinformatics results derived from DBS vs. PBMC and PAXgene samples.

Bioinformatics 1: CTRA gene expression

Our first comparison of high-level bioinformatics results assessed whether African-Americans showed different levels of CTRA gene expression than Caucasians, using an *a priori* defined contrast of 19 pro-inflammatory, 31 antiviral, and 3 antibody-related transcripts (Cole et al. 2015; Fredrickson et al. 2013; 2015; Vedhara et al. 2015). In analyses of DBS samples, African-Americans showed a 0.85-fold lower expression of the overall CTRA contrast (Table 2, column 1; $p = .001$), counter to our initial hypothesis of higher levels of CTRA expression in African-Americans. Separate analyses of the inflammatory, antiviral, and antibody-related components (Table 2, column 1, rows 2-4) found no significant differences in inflammatory or antibody-related gene expression, but a significant 1.27-fold up-regulation of Type I interferon antiviral gene expression in African-Americans relative to Caucasians ($p = .003$). (Note that Type I interferon responses are an inverse component of the CTRA profile.) Parallel analyses of PBMC and PAXgene samples yielded convergent results (Table 2, columns 2-3), with overall CTRA gene expression profiles showing 0.87-fold ($p < .001$) and 0.85-fold ($p < .001$) differential expression, respectively. Antiviral gene expression was up-regulated by 1.29-fold in PBMC ($p < .001$) and by 1.28-fold in PAXgene samples ($p = .003$). Similar results emerged in ancillary analyses

that additionally controlled for gene transcripts indicating the relative prevalence of major leukocyte subsets (Table 2, rows 5-8).

Bioinformatics 2: Transcription factor activity

Our second comparison of high-level bioinformatics results assessed whether African-Americans and Caucasians showed different levels of activity in seven major transcription control pathways that have been implicated in social environmental influences on gene regulation, including three pathways involved in pro-inflammatory gene expression (AP-1, NF-κB, and GATA factors), two pathways involved in interferon-related antiviral signaling (ISRE and STAT family factors), and two pathways involved in neuroendocrine signaling (CREB and glucocorticoid receptor [GR] signaling). Table 3 (column 1) reports results from TELiS analysis of 291 genes identified as up-regulated ≥ 1.20-fold and 540 genes identified as down-regulated ≥ 1.20-fold in African-Americans relative to Caucasians based on analyses of DBS RNA. Results indicated relative activation of five of the seven analyzed pathways in African-Americans relative to Caucasians, including two pro-inflammatory transcription factors (AP-1: 1.27-fold, $p = .019$; GATA: 1.48-fold, $p < .001$), both interferon responsive factors (ISRE: 2.22-fold, $p = .008$; STAT: 1.40-fold, $p = .015$), and one of the two neuroendocrine pathways (CREB: 1.46-fold, $p = .023$). In parallel analyses of 211 genes identified as up-regulated and 223 genes identified as down-regulated based on PAXgene whole blood samples, results concurred for four of the five transcription control pathways found to be significant in DBS-based analyses (Table 3, column 2: AP-1, GATA, ISRE, and STAT). In parallel analyses of 170 genes identified as up-regulated and 76 genes identified as down-regulated based on PBMC samples, results concurred for two of the five transcription control pathways found to be significant in DBS-based analyses (Table 3, column 3: ISRE and STAT; CREB was indicated significant, but in the opposite direction).

In a follow-up set of analyses, we added statistical controls for cellular heterogeneity. Each class of white blood cells—granulocytes, monocytes, and lymphocytes—has distinct immunologic functions, enabled in part by different gene expression patterns. There is considerable person-to-person differences

in the abundance of these cell classes, so their relative balance is important to consider as a potential explanation for any observed transcriptional differences. When covariates reflecting cellular heterogeneity were added to the models, the overall pattern of consistency between RNA sampling modes reported above was maintained (i.e., DBS vs. PAXgene vs. PBMC). However, the covariates did affect which specific transcription control pathways yielded statistically significant indications of difference across African-American and Caucasian participants. Leukocyte subset-adjusted analyses based on DBS RNA (Table 3, column 1, rows 8-14) indicated relative activation of three of the seven targeted pathways in African-Americans, including two interferon-related pathways (ISRE: 2.67-fold, $p < .001$; STAT: 1.47-fold, $p = .003$) and one neuroendocrine pathway (GR: 2.20-fold, $p = .002$). However, these covariates rendered non-significant the indications of group differences in inflammatory transcription factor activity. Parallel subset-adjusted analyses based on PAXgene RNA (Table 4, column 2, rows 8-14) verified all three of the significant adjusted results from DBS. However, none of these three subset-adjusted results were found in PBMC analyses. Instead, PBMC analyses indicated disparities in activation of three pro-inflammatory pathways that were not identified by analyses of either DBS or PAXgene RNA (AP-1, NF-κB, and GATA).

To summarize, the analyses of TF activity revealed substantial convergence between DBS and PAXgene results, even when adjustments for cellular heterogeneity were performed. There was much less correspondence between DBS and PBMC at the level of TF activity. As we explain in the Discussion, these patterns could reflect the similar cellular composition of DBS and PAXGene samples relative to PBMC and/or differences in specimen processing duration that act to delay cell lysis and thereby facilitate additional gene expression dynamics in PBMC.

Bioinformatics 3: Cellular origins

Our third comparison of high-level bioinformatics results assessed the cellular sources of differences in gene expression using Transcript Origin Analysis (Cole et al. 2011). These analyses made use of reference data from isolated monocytes, dendritic cells, CD4+ T cells, CD8+ T cell, B cells, and

Natural Killer cells (Su et al., 2004). Results from DBS samples (Table 4, column 1) implicated monocytes and Natural Killer cells as predominate cellular sources of differences between African-Americans and Caucasians in gene expression (all $p < .001$). The same conclusions emerged from analyses of PAXgene samples (Table 4, column 2) and PBMC (Table 4, column 3). Results from PAXgene and PBMC additionally suggested that dendritic cells contribute to the observed transcriptional differences across groups. Similarly consistent results also emerged in analyses that adjusted for mRNA markers of leukocyte subset prevalence (Table 4, rows 7-12).

Discussion

This study demonstrates that genome-wide transcriptional profiling of RNA derived from DBS samples can yield bioinformatic inferences that correspond well to those derived from gold-standard venipuncture blood samples. These results establish the feasibility of using DBS as a minimally-invasive alternative to venous blood sampling for field-based studies of gene expression, particularly when research interests focus on higher-order bioinformatic analyses of the general molecular and cellular processes that drive broad patterns of variation in gene expression. By capitalizing on the statistical advantages of aggregating across many individual transcripts, these bioinformatics analyses revealed common abstract biological themes that were generally evident across specimen types. These themes included significant differences between African-American and Caucasian adults in the CTRA profile, patterns of transcription factor activation (e.g., TELiS analyses of inflammatory, antiviral, and neuroendocrine signaling pathways), and the types of cells that contribute to such disparities (e.g., TOA implication of monocytes and Natural Killer cells). The convergence of these findings with results from venous blood suggests that DBS methods will be a useful tool for expanding social genomics research to sampling contexts in which clinic-based assessments are infeasible. This would include large, population-based cohorts, and field-studies in diverse cultural and ecological settings.

Although bioinformatics analyses yielded broadly similar results across DBS and venipuncture blood samples, there were areas of relatively stronger vs. weaker convergence that merit consideration.

One dimension of variation involved the specific bioinformatic parameters analyzed, with good consistency across sampling modalities for the 53-gene CTRA contrast (and its inflammatory, antiviral, and antibody-related components) and TOA inferences regarding cell type. Results were more variable for TELiS inferences of transcription factor activity, with some transcription control pathways showing highly consistent results (e.g., interferon/antiviral and inflammatory signaling pathways) and others showing less uniform findings (e.g., neuroendocrine signaling). Nevertheless, the overall pattern of transcription factor inferences was broadly consistent (e.g., 5 out of 7 comparisons yielded consistent findings for the most biologically valid comparison of DBS vs. PAXgene whole blood samples).

The sampling modality used as reference point also affected the consistency of results, with DBS results corresponding more closely to PAXgene samples than to PBMC samples. This asymmetry was expected, due to the fact that the PAXgene whole blood sample matches the cellular composition of DBS more closely than does the PBMC sample (which is depleted of red blood cells and most granulocytes). DBS and PAXgene samples also result in rapid cell lysis following sample collection (stopping any subsequent RNA transcription), whereas blood drawn for PBMC isolation may continue to undergo transcriptional dynamics during transport and laboratory processing (Baechler et al. 2004; Debey et al. 2004; Fan & Hegde 2005; Lyons et al. 2007; Rainn et al. 2002; Tanner et al. 2002). Both factors would lead to greater transcriptome disparity between PBMC and DBS than between PAXgene and DBS. This is supported by the fact that PAXgene and PBMC also showed notable disparities in some bioinformatic characteristics (particularly TELiS indications of transcription factor activity).

It is also notable that several instances of inconsistency arose because PAXgene or PBMC samples yielded significant results that were not evident in DBS. This may be explained by the greater noise in DBS gene expression estimates resulting from their notably smaller initial RNA content (around $1/100^{th}$ that of venipuncture samples) and additional stochastic variation introduced by enzymatic amplification of cDNA. Both sources of noise would result in a consequent decline in statistical power. This implies that bioinformatic analyses of DBS RNA may be more likely to incur conservative false negative errors than erroneous declarations of substantive effect (false positive errors). Given the

greater sampling variability of DBS RNA assessments, somewhat larger samples would be required to maintain statistical power at levels comparable to PBMC. For example, given the slightly lower reliability of CTRA assessment using DBS vs. PBMC (Cronbach's alpha = .88 vs .90), sample sizes would need to increase by approximately 5% to maintain equivalent statistical power (assuming a simple 2-group comparison targeting a small effect size of d = .2-.3).

A limitation of our validation is the use of DBS samples derived from venous whole blood applied to filter paper, rather than finger stick whole blood. There is no *a priori* reason to expect different patterns of results across these samples, and our comparison of 10 matched venous and finger stick whole blood samples revealed no bias, and very high level of agreement. We are therefore confident that our results generalize to finger stick DBS samples.

All of the comparative results examined here were based on analyses of differential gene expression across African-American and Caucasian participants, and controlled for differences in gender, BMI, alcohol and cigarette use, and educational attainment. This context limits the scope of the present study's conclusions; analyses of other substantive effects may well indicate greater consistency (e.g., if biological effect sizes are larger than those assessed here) or greater inconsistency (e.g., if biological effect sizes are smaller than those observed here, or if they occur on different biological dimensions from those selected for comparative testing). However, the analysis of group differences also represents a substantively important dimension of human transcriptome regulation. Discrimination and other forms of social inequity place African-Americans at increased risk for a wide variety of adverse health outcomes relative to Caucasians (Braveman et al. 2010; Meyer et al. 2013; Williams et al. 2012), and the findings from this study join others (Huang et al. 2011) in identifying potential biological pathways that may mediate these outcomes. Notably, we found no indication of significantly elevated pro-inflammatory signaling in African-Americans vs. Caucasians (e.g., CTRA and TELiS NF-κB). This may stem from our control for inflammation-related risk factors such as BMI and SES (educational attainment) that are differentially distributed in the African-American vs. Caucasian participants in our sample. It may also

reflect idiosyncrasies of this study, as this was not a population-representative sample. Future research using more broadly representative samples of participants across socio-demographic groups will be required to verify and extend these results.

The most striking observed difference involved a marked up-regulation of interferon signaling in African-Americans relative to Caucasians (i.e., the general CTRA contrast and its interferon component, and TELiS indications regarding interferon responsive transcription factors). TOA bioinformatics implicated monocytes and Natural Killer cells as cellular sources of group differences in gene expression, but statistical control for mRNA markers of monocyte and Natural Killer cell prevalence failed to eradicate the indicated differences in interferon activity. This pattern of findings suggests a marked state of Type I interferon-related cellular activation in the blood transcriptome of this study's African-American participants, which would be consistent with TELiS indications of increased AP-1 and GATA factor signaling (both of which mediate general immunologic activation beyond the more focally pro-inflammatory transcription factor, NF-κB). The robustness and origin of these effects remain to be determined in future research, and the present bioinformatics findings (consistent across three different RNA sampling modalities) establish a clear set of molecular and cellular targets for future analyses. Nevertheless, it is important to underscore that the present results were derived from group differences in a specific environmental context, and there is no guarantee that similar findings would be observed in other social or cultural contexts. It is also worth noting that interferon signaling has long been observed to constitute one of the most prominent dimensions of naturally occurring individual differences in human blood cell gene expression (Eady et al. 2005; Radich et al. 2004; Whitney et al. 2003).

The low cost and technical burden of DBS sampling facilitates the collection of blood from large numbers of participants in diverse community-based settings, and it opens up new lines of research into the pathways through which social, behavioral, and other contextual factors "get under the skin" to shape physiological function and health (Finch et al. 2001; McDade et al. 2007; Weinstein et al. 2008). Genome-wide transcriptional profiling generates insight into critical intra-cellular dynamics that regulate these processes, and it provides unprecedented opportunities for unbiased discovery as well as the

interrogation of candidate biological pathways. At a per sample cost of approximately $150, transcriptome profiling may be out of reach for large studies, but costs are projected to decrease over the next few years and broader applications in a wide variety of field-based settings may thus become feasible. Our aim is to bring these approaches together to promote innovative research at the biosocial interface that bridges concepts and methods from the population and biological sciences.

References

Baechler, E. C., F. M. Batliwalla, G. Karypis, P. M. Gaffney, K. Moser, W. A. Ortmann, K. J. Espe, S Balasubramanian, KM Hughes, JP Chan, A Begovich, S-YP Chang, PK Gregerson, and TW Behrens. 2004. Expression Levels for Many Genes in Human Peripheral Blood Cells Are Highly Sensitive to Ex Vivo Incubation. *Genes and Immunity* 5 (5): 347–53. doi:10.1038/sj.gene.6364098.

Braveman PA., C. Cubbin, S. Egerter, D.R. Williams, and E. Pamuk. Socioeconomic disparities in health in the United States: what the patterns tell us. *Am J Public Health*. 2010;100 Suppl 1 S186-S196.

Centers for Disease Control and Prevention (CDC). 2009. *National Health and Nutrition Examination Survey Questionnaire*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Cole, S.W. 2013. Social Regulation of Human Gene Expression: Mechanisms and Implications for Public Health. *American Journal of Public Health* 103 (Suppl 1): S84–92. doi:10.2105/AJPH.2012.301183.

Cole, S.W. 2014. Human Social Genomics. *PLoS Genetics* 10 (8). doi:10.1371/journal.pgen.1004601.

Cole, S.W. 2010. Elevating the Perspective on Human Stress Genomics. *Psychoneuroendocrinology* 35 (7): 955–62. doi:10.1016/j.psyneuen.2010.06.008.

Cole, S.W., Z. Galic, and J.A. Zack. 2003. Controlling False-Negative Errors in Microarray Differential Expression Analysis: A PRIM Approach. *Bioinformatics (Oxford, England)* 19 (14): 1808–16.

Cole S.W., L.C. Hawkley, J.M. Arevalo, C.S. Sung, R.M. Rose, and J.T. Cacioppo. 2007. Social regulation of gene expression: Inflammation and the human transcriptional response to loneliness. *Genome Biology*. 8 r89.

Cole, S.W., L.C. Hawkley, J.M.G. Arevalo, and J.T. Cacioppo. 2011. Transcript Origin Analysis Identifies Antigen-Presenting Cells as Primary Targets of Socially Regulated Gene Expression in Leukocytes. *Proceedings of the National Academy of Sciences of the United States of America* 108 (7): 3080–85. doi:10.1073/pnas.1014218108.

Cole, S.W., M.E. Levine, J.M.G. Arevalo, J. Ma, D.R. Weir, and E.M. Crimmins. 2015. Loneliness, Eudaimonia, and the Human Conserved Transcriptional Response to Adversity. *Psychoneuroendocrinology* 62 (December): 11–17. doi:10.1016/j.psyneuen.2015.07.001.

Cole, S.W., W. Yan, Z. Galic, J. Arevalo, and J.A. Zack. 2005. Expression-Based Monitoring of Transcription Factor Activity: The TELiS Database. *Bioinformatics (Oxford, England)* 21 (6): 803–10. doi:10.1093/bioinformatics/bti038.

Debey, S., U. Schoenbeck, M. Hellmich, B. S. Gathof, R. Pillai, T. Zander, and J. L. Schultze. 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 (3): 193–207. doi:10.1038/sj.tpj.6500240.

Eady, J.J., G.M. Wortley, Y.M. Wormstone, J.C. Hughes, S.B. Astley, R.J. Foxall, J.F. Doleman, and R.M. Elliott. 2005. Variation in Gene Expression Profiles of Peripheral Blood Mononuclear Cells from Healthy Volunteers. *Physiological Genomics* 22 (3): 402–11. doi:10.1152/physiolgenomics.00080.2005.

Efron, B. and R.J. Tibshirani. 1993. *An introduction to the bootstrap*. New York, Chapman & Hall.

Fan H, Hegde PS. The transcriptome in blood: challenges and solutions for robust expression profiling. Curr Mol Med. 2005 Feb;5(1):3-10. Review. PubMed PMID: 15720265.

Finch, C.E., J.W. Vaupel, K.G. Kinsella, eds. 2001. *National Research Council (U.S.). Committee on Population: Cells and surveys-Should biological measures be included in social science research?* Washington, D.C.: National Academies Press.

Fredrickson B.L., K.M. Grewen, K.A. Coffey, S.B. Algoe, A.M. Firestine, J.M.G. Arevalo, J. Ma, and S.W. Cole. A functional genomic perspective on human well-being. *Proc Natl Acad Sci U S A.* 2013;110 (33):13684-13689.

Fredrickson, B.L., K.M. Grewen, S.B. Algoe, A.M. Firestine, J.M.G. Arevalo, J. Ma, and S.W. Cole. 2015. Psychological Well-Being and the Human Conserved Transcriptional Response to Adversity. *PLoS ONE* 10 (3). doi:10.1371/journal.pone.0121839.

Henrich, J., S.J. Heine, and A. Norenzayan. 2010. The Weirdest People in the World? *Behavioral and Brain Sciences* 33 (2-3): 61–83. doi:10.1017/S0140525X0999152X.

Ho, N.T., K. Furge, W. Fu, J. Busik, S.K. Khoo, Q. Lu, M. Lenski, J. Wirth, E. Hurvitz, N. Dodge, J. Resau, and N. Paneth. 2013. Gene Expression in Archived Newborn Blood Spots Distinguishes Infants Who Will Later Develop Cerebral Palsy from Matched Controls. *Pediatric Research* 73 (4 0 1): 450–56. doi:10.1038/pr.2012.200.

Huang, C.C., D.M. Lloyd-Jones, X. Guo, N.M. Rajamannan, S. Lin, P. Du, Q. Huang, L. Hou and K. Liu. 2011. Gene expression variation between African Americans and Whites is associated with coronary artery calcification: The multiethnic study of atherosclerosis. *Physiol Genomics*. 43 (13):836-843.

Khoo, S.K., K. Dykema, N.M. Vadlapatla, D. LaHaie, S. Valle, D. Satterthwaite, S.A. Ramirez, J.A. Carruthers, P.T. Haak, and J.H. Resau. 2011. Acquiring Genome-Wide Gene Expression Profiles in Guthrie Card Blood Spots Using Microarrays. *Pathology International* 61 (1): 1–6. doi:10.1111/j.1440-1827.2010.02611.x.

Lyons, P.A., M. Koukoulaki, A. Hatton, K. Doggett, H.B. Woffendin, A.N. Chaudhry, and K.G.C. Smith. 2007. Microarray Analysis of Human Leucocyte Subsets: The Advantages of Positive Selection and Rapid Purification. *BMC Genomics* 8 (March): 64. doi:10.1186/1471-2164-8-64.

McDade, T.W., S. Williams, and J.J. Snodgrass. 2007. What a Drop Can Do: Dried Blood Spots as a Minimally Invasive Method for Integrating Biomarkers into Population-Based Research. *Demography* 44 (4): 899–925.

Mei, J.V., J.R. Alexander, B.W. Adam, and W.H. Hannon. 2001. Use of filter paper for the collection and analysis of human whole blood specimens. *Journal of Nutrition* 131: 1631S-6S.

Meyer, P.A, P.W. Yoon, R.B. Kaufmann, and Centers for Disease Control and Prevention. 2013. Introduction: CDC Health Disparities and Inequalities Report - United States, 2013. *MMWR Surveill Summ*. 62 Suppl 3 3-5.

Miller, G. E., E. Chen, and S.W. Cole. 2009. Health Psychology: Developing Biologically Plausible Models Linking the Social World and Physical Health. *Annual Review of Psychology* 60 (1): 501–24. doi:10.1146/annurev.psych.60.110707.163551.

Miller G.E., C.A. Stetler, R.M. Carney, K.E. Freedland, and W.A. Banks. Clinical depression and inflammatory risk markers for coronary heart disease. *Am J Cardiol*. 2002;90 1279-1283.

Miller G.E., E. Chen, J. Sze, T. Marin, J.M.G. Arevalo, R. Doll, R. Ma, and S.W. Cole. 2008. A genomic fingerprint of chronic stress in humans: Blunted glucocorticoid and increased NF-κB signaling. *Biol Psychiatry*. 64:266-272.

Miller G.E., E. Chen E, A. Fok A, H. Walker, A. Lim, E.F. Nicholls, S. Cole, and M.S. Kobor. 2009. Low early-life social class leaves a biological residue manifest by decreased glucocorticoid and increased pro-inflammatory signaling. *Proc Natl Acad Sci U S A*. 106:14716-14721.

Miller G.E., M.L. Murphy, R. Cashman, R. Ma, J. Ma, J.M.G. Arevalo, M.S. Kobor, and S.W. Cole. 2014. Greater inflammatory activity and blunted glucocorticoid signaling in monocytes of chronically stressed caregivers. *Brain Behav Immun*. 41: 191-199.

Miller, G. E., and S.W. Cole. 2010. Functional genomic approaches in behavioral medicine research. *In A. Steptoe (Ed) Handbook of Behavioral Medicine: Methods and Applications*. p 443-454. New York, Springer.

Pappas, A., T. Chaiworapongsa, R. Romero, S.J. Korzeniewski, J.C. Cortez, G. Bhatti, N. Gomez-Lopez, S.S. Hassan, S. Shankaran, and A.L. Tarca. 2015. Transcriptomics of Maternal and Fetal Membranes Can Discriminate between Gestational-Age Matched Preterm Neonates with and without Cognitive Impairment Diagnosed at 18–24 Months. *PLoS ONE* 10 (3). doi:10.1371/journal.pone.0118573.

Powell N.D., E.K. Sloan, M.T. Bailey, J.M.G Arevalo, G.E. Miller, E. Chen, M.S. Kobor, B.F. Reader, J.F. Sheridan, and S.W. Cole. Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via beta-adrenergic induction of myelopoiesis. *Proc Natl Acad Sci U S A*. 2013;110 (41):16574-16579.

Radich, J.P., M. Mao, S. Stepaniants, M. Biery, J. Castle, T. Ward, G. Schimmack, S. Kobayashi, M. Carleton, J. Lampe, and P.S. Linsley. 2004. Individual-Specific Variation of Gene Expression in Peripheral Blood Leukocytes. *Genomics* 83 (6): 980–88. doi:10.1016/j.ygeno.2003.12.013.

Rainen, L., U. Oelmueller, S. Jurgensen, R. Wyrich, C. Ballas, J. Schram, C. Herdman, D. Bankaitis-Davis, N. Nicholls, D. Trollinger, and V. Tryon. 2002. Stabilization of mRNA Expression in Whole Blood Samples. *Clinical Chemistry* 48 (11): 1883–90.

Resau, J.H., N.T. Ho, K. Dykema, M.S. Faber, J.V. Busik, R.Z. Nickolov, K.A. Furge, N.Paneth, S. Jewell, and S.K. Khoo. 2012. Evaluation of Sex-Specific Gene Expression in Archived Dried Blood Spots (DBS). *International Journal of Molecular Sciences* 13 (8): 9599–9608. doi:10.3390/ijms13089599.

Shi, L., W.D. Jones, R.V. Jensen, S.C. Harris, R.G. Perkins, F.M. Goodsaid, L. Guo, L.J. Croner, C. Boysen, H. Fang, F. Qian, S. Amur, W. Bao, C.C. Barbacioru, V. Bertholet, X.M. Cao, T. Chu. P.J. Collins, X. Fan, F.W. Frueh, J.C. Fuscoe, X. Guo, J. Han. D. Herman, H. Hong, E.S. Kawasaki, Q. Li, Y. Luo, Y. Ma, N. Mei, R.L. Peterson, R.K. Puri, R. Shippy, Z. Su, Y.A. Sun, H. Sun, B. Thorn. Y. Turpaz, C. Wang, S.J. Wang, J.A. Warrington, J. C. Willey, J. Wu, Q. Xie, L. Zhang, L. Zhang, S. Zhong, R.D. Wolfinger, and W. Tong. 2008. The Balance of Reproducibility, Sensitivity, and Specificity of Lists of Differentially Expressed Genes in Microarray Studies. *BMC Bioinformatics* 9 (Suppl 9): S10. doi:10.1186/1471-2105-9-S9-S10.

Slaughter, J., C. Wei, S.J. Korzeniewski, Q. Lu, J.S. Beck, S.L. Khoo, A. Brovont, J. Maurer, D. Martin, M. Lenski, and N. Paneth. 2013. High Correlations in Gene Expression between Paired Umbilical Cord Blood and Neonatal Blood of Healthy Newborns on Guthrie Cards. *The Journal of Maternal-Fetal & Neonatal Medicine : The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians* 26 (18): 1765–67. doi:10.3109/14767058.2013.804050.

Su, A.I., T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, and J.B. Hogenesch. 2004. A Gene Atlas of the Mouse and

Human Protein-Encoding Transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* 101 (16): 6062–67. doi:10.1073/pnas.0400782101.

Tanner, M. A., L. S. Berk, D. L. Felten, A. D. Blidy, S. L. Bit, and D. W. Ruff. 2002. Substantial Changes in Gene Expression Level due to the Storage Temperature and Storage Duration of Human Whole Blood. *Clinical & Laboratory Haematology* 24 (6): 337–41. doi:10.1046/j.1365-2257.2002.00474.x.

Vedhara, K., S. Gill, L. Eldesouky, B.K. Campbell, J.M.G. Arevalo, J. Ma, and S.W. Cole. 2015. Personality and Gene Expression: Do Individual Differences Exist in the Leukocyte Transcriptome? *Psychoneuroendocrinology* 52 (February): 72–82. doi:10.1016/j.psyneuen.2014.10.028.

Wei, Ch., Q. Lu, S.K. Khoo, M. Lenski, R. Fichorova, A. Leviton, and N. Paneth. 2014. Comparison of Frozen and Unfrozen Blood Spots for Gene Expression Studies. *The Journal of Pediatrics* 164 (1): 189–91.e1. doi:10.1016/j.jpeds.2013.09.025.

Weinstein, M., J.W. Vaupel, J.W., and K.W. Wachter, eds. 2008. *National Research Council (U.S.) Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys: Biosocial surveys*. Washinton, D.C. National Academies Press.

Whitney, A.R., M. Diehn, S.J. Popper, A.A. Alizadeh, J.C. Boldrick, D.A. Relman, and P.O. Brown. 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 (4): 1896–1901. doi:10.1073/pnas.252784499.

Williams DR. 2012. Miles to go before we sleep: racial inequities in health. *J Health Soc Behav*. 53 (3):279-295.

Witten, D. M., and R. Tibshirani. 2007. *A comparison of fold-change and the t-statistic for microarray data analysis*. Stanford, Stanford University: 1-17.

Tables

Table 1. Characteristics of the study sample (n=82).

Characteristics	**N (%) or Mean (*SD*)**
Female gender	52 (63.4%)
Age (years)	34.8 (13.7)
Race/ethnicity: White	52 (62.2%)
African-American	16 (19.8%)
Asian	7 (8.6%)
Other or multiple	6 (7.4%)
Education	
High school degree or less	23 (28.1%)
Associates degree	7 (8.5%)
Bachelor's degree or more	52 (63.4%)
Full-time employment	45 (54.9%)
Body Mass Index (kg/m^2)	26.4 (5.5)
Heavy smoker ($\geq$ 10 cigarettes per day)	3 (3.7%)
Heavy drinker ($\geq$ 10 alcoholic drinks per week)	11 (13.4%)

Table 2. Linear model parameter estimate (SE) differences in CTRA contrast scores for African-American participants relative to Caucasian participants.

	RNA sample*		
	DBS	**PAXgene**	**PBMC**
Primary analysis			
CTRA contrast	-0.226 (0.065) p = 0.001	-0.231 (0.060) p < 0.001	-0.198 (0.048) p < 0.001
Inflammation	-0.039 (0.050) p = 0.450	-0.044 (0.029) p = 0.143	+0.086 (0.063) p = 0.188
Interferon	+0.347 (0.106) p = 0.003	+0.340 (0.104) p = 0.003	+0.368 (0.071) p < 0.001
Antibody	+0.160 (0.095) p = 0.235	+0.293 (0.122) p = 0.138	+0.238 (0.101) p = 0.143
Adjusted for leukocyte subsets			
CTRA contrast	-0.285 (0.063) p < 0.001	-0.197 (0.048) p < 0.001	-0.149 (0.050) p = 0.004
Inflammation	-0.039 (0.045) p = 0.404	-0.037 (0.022) p = 0.114	+0.092 (0.060) p = 0.140
Interferon	+0.446 (0.097) p < 0.001	+0.290 (0.079) p = 0.001	+0.291 (0.071) p < 0.001
Antibody	+0.180 (0.097) p = 0.205	+0.250 (0.104) p = 0.137	+0.203 (0.093) p = 0.161

Table 3. Linear model parameter estimate (SE) for log2 ratio of TFBM prevalence in promoters of genes that are relatively up-regulated vs. down-regulated in African-American participants relative to Caucasian participants.

	RNA sample*		
	DBS	**PAXgene**	**PBMC**
Primary analysis			
AP-1	+0.350 (0.148) p = 0.019	+0.368 (0.184) p = 0.047	+0.135 (0.324) p = 0.678
NF-kB	+0.624 (0.352) p = 0.078	-0.155 (0.332) p = 0.641	+1.715 (0.430) p < 0.001
GATA	+0.570 (0.142) p < 0.001	+0.393 (0.162) p = 0.016	+0.485 (0.323) p = 0.134
ISRE	+1.149 (0.428) p = 0.008	+1.592 (0.459) p = 0.001	+1.017 (0.414) p = 0.015
STAT	+0.481 (0.197) p = 0.015	+0.781 (0.248) p = 0.002	+0.765 (0.352) p = 0.031
CREB	+0.542 (0.237) p = 0.023	-0.386 (0.342) p = 0.260	-0.989 (0.445) p = 0.028
GR	-0.243 (0.290) p = 0.403	+0.818 (0.354) p = 0.022	+0.200 (0.506) p = 0.694
Adjusted for leukocyte subsets			
AP-1	+0.234 (0.159) p = 0.144	+0.300 (0.203) p = 0.140	+1.183 (0.261) p < 0.001
NF-kB	+0.257 (0.382) p = 0.502	-0.158 (0.328) p = 0.630	+0.694 (0.336) p = 0.040
GATA	+0.281 (0.154) p = 0.069	+0.186 (0.185) p = 0.318	+0.885 (0.292) p = 0.003
ISRE	+1.419 (0.348) p < 0.001	+1.77 (0.437) p < 0.001	+0.781 (0.445) p = 0.081
STAT	+0.557 (0.183) p = 0.003	+0.729 (0.272) p = 0.008	+0.435 (0.296) p = 0.143
CREB	+0.082 (0.227) p = 0.719	-0.606 (0.322) p = 0.062	-0.178 (0.461) p = 0.700
GR	+1.140 (0.354) p = 0.002	+0.788 (0.368) p = 0.034	+0.295 (0.541) p = 0.586

Table 4. Linear model parameter estimate (SE) for TOA diagnosticity scores (which are inherently 1-tailed) for genes differentially expressed in African-American participants relative to Caucasian participants.

	RNA sample*		
	DBS	**PAXgene**	**PBMC**
Primary analysis			
Monocyte	+1.786 (0.173) p < 0.001	+1.909 (0.475) p < 0.001	+2.007 (0.485) p < 0.001
Dendritic cell	-0.032 (0.083) p = 0.651	0.318 (0.132) p = 0.008	+0.385 (0.135) p = 0.002
Natural killer cell	+1.408 (0.295) p < 0.001	+2.386 (0.681) p < 0.001	+4.421 (0.952) p < 0.001
CD4+ T cell	-0.403 (0.034) p = 1.000	-0.347 (0.056) p = 1.000	-0.249 (0.078) p = 0.999
CD8+ T cell	-0.202 (0.032) p = 1.000	-0.284 (0.054) p = 1.000	-0.188 (0.073) p = 0.995
B cell	-0.145 (0.051) p = 0.998	-0.123 (0.061) p = 0.977	+0.038 (0.093) p = 0.341
Adjusted for leukocyte subsets			
Monocyte	+1.088 (0.142) p < 0.001	+2.019 (0.469) p < 0.001	+1.360 (0.279) p < 0.001
Dendritic cell	+0.085 (0.087) p = 0.164	+0.352 (0.140) p = 0.006	+0.161 (0.144) p = 0.131
Natural killer cell	+1.811 (0.212) p < 0.001	+2.370 (0.487) p < 0.001	+3.842 (0.836) p < 0.001
CD4+ T cell	-0.381 (0.034) p = 1.000	-0.317 (0.061) p = 1.000	-0.317 (0.058) p = 1.000
CD8+ T cell	-0.180 (0.032) p = 1.000	-0.285 (0.048) p = 1.000	-0.293 (0.070) p = 1.000
B cell	-0.154 (0.048) p = 0.999	-0.097 (0.054) p = 0.963	+0.073 (0.083) p = 0.190