

Seasonal changes in gene expression represent cell-type composition in whole blood

Simone De Jong¹, Marjolein Neeleman², Jurjen J. Luykx^{3,4}, Maarten J. Ten Berg⁵, Eric Strengman⁶, Hanneke H. Den Breeijen^{7,8}, Leon C. Stijvers⁵, Jacobine E. Buizer-Voskamp², Steven C. Bakker², René S. Kahn², Steve Horvath^{9,10}, Wouter W. Van Solinge⁵ and Roel A. Ophoff^{1,2,9,*}

¹Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA, ²Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands, ³Neurogenetics Unit, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands, ⁴Department of Psychiatry, ZNA hospitals, Antwerp, Belgium, ⁵Department of Clinical Chemistry and Haematology, Division of Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands, ⁶Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands, ⁷Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands, ⁸Department of Clinical Pharmacy, University Medical Center Utrecht, The Netherlands, ⁹Department of Human Genetics, David Geffen School of Medicine and ¹⁰Department of Biostatistics, School of Public Health, University of California, Los Angeles, CA 90095, USA

Received July 1, 2013; Revised October 30, 2013; Accepted December 23, 2013

Seasonal patterns in behavior and biological parameters are widespread. Here, we examined seasonal changes in whole blood gene expression profiles of 233 healthy subjects. Using weighted gene co-expression network analysis, we identified three co-expression modules showing circannual patterns. Enrichment analysis suggested that this signal stems primarily from red blood cells and blood platelets. Indeed, a large clinical database with 51 142 observations of blood cell counts over 3 years confirmed a corresponding seasonal pattern of counts of red blood cells, reticulocytes and platelets. We found no direct evidence that these changes are linked to genes known to be key players in regulating immune function or circadian rhythm. It is likely, however, that these seasonal changes in cell counts and gene expression profiles in whole blood represent biological and clinical relevant phenomena. Moreover, our findings highlight possible confounding factors relevant to the study of gene expression profiles in subjects collected at geographical locations with disparaging seasonality patterns.

INTRODUCTION

Seasonal patterns in human behavior are widespread and influence mood, sleeping and eating patterns in the general population (1–3). Season-dependent differences have been observed in several diseases. There is latitude-dependent seasonal variation in influenza and respiratory syncytial virus epidemics (4), migraine (5) and incidence and fatality of cardiovascular and cerebrovascular diseases (6,7).

Hematological parameters could be underlying these fluctuations. Infectious diseases affect primarily white blood cells (WBC) (8), and it was recently suggested that seasonality in cardio and cerebrovascular disease could be due to variation in

red blood cell health and composition (9,10). Indeed, several hematological parameters including platelet count, neutrophils and mean corpuscular volume (MCV) show seasonal variation with peak values in summer or winter (11–13). Aforementioned small studies addressed seasonality in hematological parameters but with inconsistent and conflicting results. A meta-analysis of circannual patterns in hematocrit levels of 12 793 individuals spanning 18 studies showed that these inconsistencies are likely due to differences in blood volume, age, sex and other within-subject factors (14).

Psychiatric disorders also exhibit seasonal variation. Most notably, mood disorders symptomatology and suicide may be characterized by recurrent patterns throughout the year (15–18).

*To whom correspondence should be addressed at: Psychiatry and Human Genetics, Center for Neurobehavioral Genetics, 695 Charles E. Young Drive South, 4335B Gonda, PO Box 951761, Los Angeles, CA 90095-1761, USA. Tel: 1 3107949602; Fax: 1 3107949613; Email: ROphoff@mednet.ucla.edu

It has been suggested that these seasonal fluctuations in behavior are regulated by the serotonergic system in the brain (19,20). Accordingly, we recently demonstrated a circannual pattern of serotonin metabolite levels in cerebrospinal fluid with a peak in spring (21). Moreover, it has been reported that blood platelet serotonin concentrations show seasonal variation, but with disparaging patterns in controls, depressed patients and schizophrenia subtypes (22).

Besides clinical implications, seasonal variation in hematological parameters is of practical importance for molecular studies involving gene expression since it has been shown that different blood cell types have distinct gene expression profiles (23–25). Differential seasonal variation in gene expression profiles of particular cell types will affect the outcome of such studies when using whole blood.

To address seasonality in gene expression systematically, we investigated circannual patterns of gene expression in whole blood of 233 healthy controls using a systems-biology approach. We subsequently examined circannual variation in blood cell composition of 51 142 subjects collected during a time span of 3 years. We demonstrate for the first time that seasonal patterns in gene expression of whole blood do exist and are correlated with blood cell composition.

RESULTS

Weighted gene co-expression network reconstruction

After quality control of gene expression data, 233 of the 240 samples remained (75% male; mean age 39 years). Detection filtering left 25 361 gene expression probes for analysis. Weighted gene co-expression analysis (WGCNA) resulted in 17 modules, ranging in size from 158 probes in M17 module to 3183 probes in M1. The background module MB contains 7391 probes that do not belong to any of the other modules. This module represents background noise. Figure 1 depicts the network dendrogram.

Gene co-expression modules showing a seasonal effect

Significance threshold was set to $P = 0.05/(17 \times 6) = 4.9 \times 10^{-4}$ (Bonferroni correction for number of cosine models \times co-expression modules tested). The strongest associations were found for the cosine model with the most significant peaks in August and September. We report correlations and P -values for a seasonal pattern with a peak in August (and trough in February) throughout the article (Supplementary Material, Information 6). M1 (3183 probes) was found to be negatively associated with this seasonal pattern, implying a drop in expression values in August and a peak in February ($r = -0.29$, $P = 5 \times 10^{-6}$). M5 (1664 probes) and M16 (215 probes) modules were positively associated, with a peak in expression values in August and a drop in February ($r = 0.34$, $P = 2 \times 10^{-7}$ for M5 and $r = 0.33$, $P = 2 \times 10^{-7}$ for M16). These modules were not associated with age, sex or season of birth. Results are depicted in the barplot in Figure 2A, D and G.

Gene content of season modules suggests association with blood cell composition

The content of M1, M5 and M15 is listed in Supplementary Material, Information 7. The functional annotation tool by DAVID Bioinformatics resources (<http://david.abcc.ncifcrf>.

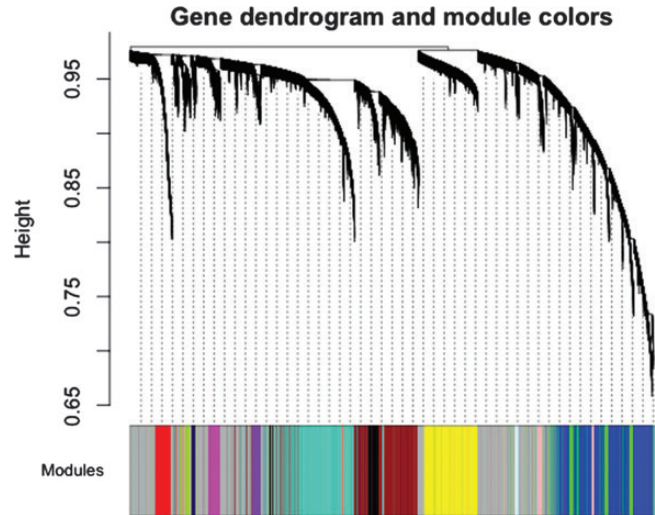


Figure 1. Network construction identifies 17 distinct modules of co-expressed genes in whole blood of 233 healthy human subjects. The dendrogram was produced by average linkage hierarchical clustering of genes using topological overlap. Modules of co-expressed genes were assigned numbers corresponding to the branches indicated by the horizontal bar beneath the dendrogram. Modules are designated by numbers throughout this article: turquoise = M1, blue = M2, brown = M3, yellow = M4, green = M5, red = M6, black = M7, pink = M8, magenta = M9, purple = M10, greenyellow = M11, tan = M12, salmon = M13, cyan = M14, midnightblue = M15, lightcyan = M16, grey60 = M17 and grey background module = MB.

gov, accessed November 2012) did not yield significant functional enrichment results for M1 and M16. M5 (1458 known genes) was only slightly enriched for ‘acetylation’ compared with background genes (genes detected above detection threshold in current gene expression dataset) with Benjamini & Hochberg corrected $P = 5 \times 10^{-2}$.

To characterize the gene co-expression modules regarding cell types present in blood, we utilized the userListEnrichment function incorporated in the WGCNA package (26). This function holds 11 lists representing markers for different blood cell types (23,24,25,27). Significance threshold was set to $P = 0.05/(3 \times 11) = 0.002$ (Bonferroni correction for modules \times blood list categories). We found M5 to be highly enriched for red blood cell markers ($P = 1.0 \times 10^{-33}$) and to a lesser extent for blood platelets ($P = 2.4 \times 10^{-6}$) and time of blood draw ($P = 2.8 \times 10^{-5}$). M16 also shows enrichment for red blood cell markers ($P = 2.7 \times 10^{-6}$). Finally, M1 is enriched for markers related to red blood cell size distribution ($P = 5.7 \times 10^{-9}$), red blood cell markers ($P = 7.2 \times 10^{-5}$) and blood platelet markers ($P = 2.0 \times 10^{-4}$).

Upon closer inspection we identified many ferritin genes in the modules of interest. Ferritin proteins consist of two subunits, H and L, encoded by separate genes (28). Our gene expression dataset contains 12 probes for ferritin genes, of which 11 end up in the season-related gene networks. M1 contains two probes for *FTL* (ferritin, light polypeptide). M5 contains *FTH1* (ferritin, heavy polypeptide 1) in addition to probes for pseudogenes for *FTH1*; *FTHL2*, *FTHL8*, *FTHL11* and two probes for *FTHL12*. M16 also contains probes for *FTH1* pseudogenes; *FTHL2*, *FTHL11* and *FTHL3P*. Ferritin genes are involved in iron transport, as well as hemoglobin. Red blood cells are the main cell types involved in this process, next to platelets (28).

Of the 11 probes for hemoglobin genes available in our dataset, 7 appear in M1: *HBB*, *HBD*, *HBM*, *HBQ1*, *HBZ* and 2 probes for *HBA2*. Both ferritin and hemoglobin genes are reliably expressed in red blood cells and blood platelets (25,29).

Blood cell composition over time confirms seasonal effects

Our results led us to believe that seasonal changes in gene expression could be driven by seasonal variation in blood cell composition. To test this hypothesis, we obtained a large dataset consisting of 51 142 observations of blood cell counts across 3 years. We examined seasonal changes in indices for platelets (platelet count and mean platelet volume), red blood cells (RBC; total RBC, proportion of reticulocytes) and WBC (total WBC, percentages of neutrophils, eosinophils, basophils, lymphocytes and monocytes). Results are displayed in Figure 2B, E and H.

Cross-sectional analysis

The blood cell count data are from different individuals than the ones in our original expression dataset. Therefore, we calculated the mean module eigengenes per month for M1, M5 and M16. Subsequently, we averaged the cell counts for each type per month, over 3 years and performed a one-sided ($P < 0.1$) correlation test based on expectations from the previous enrichment results. Results are displayed in Figure 2C, F and I.

Both M5 and M16 are enriched for red blood cell markers. Indeed, both module eigengenes correlated positively with the proportion of reticulocytes over RBC (M16: $r = 0.82$, $P = 0.001$, M5: $r = 0.74$, $P = 0.006$). Although visual inspection of the barplot in Figure 2E does suggest a seasonal pattern in total RBC, the correlation is not significant for either module (M16: $r = 0.26$, $P = 0.42$; M5: $r = 0.29$, $P = 0.36$). Both modules are correlated negatively with platelet count (M16: $r = -0.60$, $P = 0.04$; M5: $r = -0.52$, $P = 0.08$). Finally, there was a negative correlation with percentage of monocytes over WBC (M16: $r = -0.51$, $P = 0.09$; M5: $r = -0.55$, $P = 0.06$). M1 (enriched for platelet markers) was indeed positively correlated with platelet count ($r = 0.57$, $P = 0.05$) and negatively to mean platelet volume ($r = -0.57$, $P = 0.05$). Finally, the M1 module eigengene was negatively correlated to proportion of reticulocytes over RBC ($r = -0.69$, $P = 0.01$).

To confirm a similar seasonal pattern, we performed a linear regression with cosine function peaking in August with above-mentioned blood cell parameters. We selected 2010 since this year contains the most individual measurements. Indeed, we find that mean platelet volume (MPV) ($t = 13.2$, $P = 1.2 \times 10^{-39}$, difference peak versus trough 0.19×10^{-15} l), RBC ($t = 7.9$, $P = 2.0 \times 10^{-15}$, difference peak versus trough 0.08×10^{-12} l) and proportion of reticulocytes over RBC ($t = 5.4$, $P = 5.3 \times 10^{-8}$, difference peak versus trough 1.52) are positively correlated with the cosine function. In addition, percentage of monocytes ($t = -3.2$, $P = 1.4 \times 10^{-3}$, difference peak versus trough 0.27%) and platelet count ($t = -12.0$, $P = 4.3 \times 10^{-33}$, difference peak versus trough 0.02×10^{-9} l) are negatively correlated to a seasonal pattern with a peak in August.

Blood cell counts are known to differ between males and females and change over age. All except one parameter were significantly different between sexes (Supplementary Material, Information 8). However, stratified analysis for each sex separately

and for two age bins (18–40 and 40–65) resulted in the same seasonal pattern in blood cell counts (Supplementary Material, Information 9).

Longitudinal analysis

To investigate a seasonal pattern within individuals over time, we performed longitudinal analyses. We considered a first season (months 12–5) and second season of the year (months 6–11). We selected individuals with a maximum of two measurements per year and at least one in each season. This resulted in 2459 individuals. A Wilcoxon rank-sum test to test for differences between the two seasons was applied to RBC, proportion of reticulocytes, blood platelets, MPV and percentage of monocytes. Bonferroni correction for multiple testing resulted in a significance threshold of $P = 0.05/5 = 0.01$. In line with cross-sectional results, RBC ($P = 0.001$, difference of 0.03×10^{-12} l), proportion of reticulocytes ($P = 2.2 \times 10^{-6}$, difference of 0.06 proportion over reticulocytes) and MPV ($P = 6.0 \times 10^{-14}$, difference of 0.07×10^{-15} l) showed higher values in the second season and the blood platelet count ($P = 4.2 \times 10^{-15}$, difference of 0.07×10^{-9} l) the reverse pattern. Percentage of monocytes did not reach a significant difference between seasons.

DISCUSSION

In this study, we observed a circannual pattern of gene expression in whole blood of a healthy control population using co-expression network analysis. The gene co-expression modules that show seasonal patterns are enriched for particular blood cell types, mainly RBC and blood platelets. We subsequently confirmed that the observed variation of expression profiles are likely to stem from seasonal fluctuation in whole blood cell composition.

Previous studies have addressed seasonality in hematological parameters, but results regarding the peaks and troughs are inconsistent (11–13). Although these inconsistencies are thought to result from factors like age and sex (14), in our large sample we consistently find the same seasonal patterns for men and women as well as for the different age groups. Our results suggest that the fluctuations in whole blood gene expression coincide with particular cell-type counts, which does not exclude the possibility of a different seasonal pattern for remaining cell types. The M1 gene co-expression module is upregulated in the first half of the year and downregulated in the second, with the most significant trough in August. This module is enriched for RBC markers and platelets, which is confirmed by corresponding seasonal fluctuation in platelets in a large blood cell counts dataset. The percentage of monocytes also shows this pattern. Two other gene co-expression modules (M5 and M16) show under-expression in the first half of the year and overexpression in the second. These modules are enriched for RBC markers and this pattern is correlated with the seasonal fluctuation in proportion of reticulocytes.

Although WBC are most transcriptionally active, gene expression differences related to different blood cell-type counts have been reported before (23). Reticulocytes are immature red blood cells and contain residual RNA. This could explain, why the proportion of reticulocytes gives a stronger seasonal signal than the total count of RBC in our analyses. Although RBC discard their nucleus during maturation, they are still

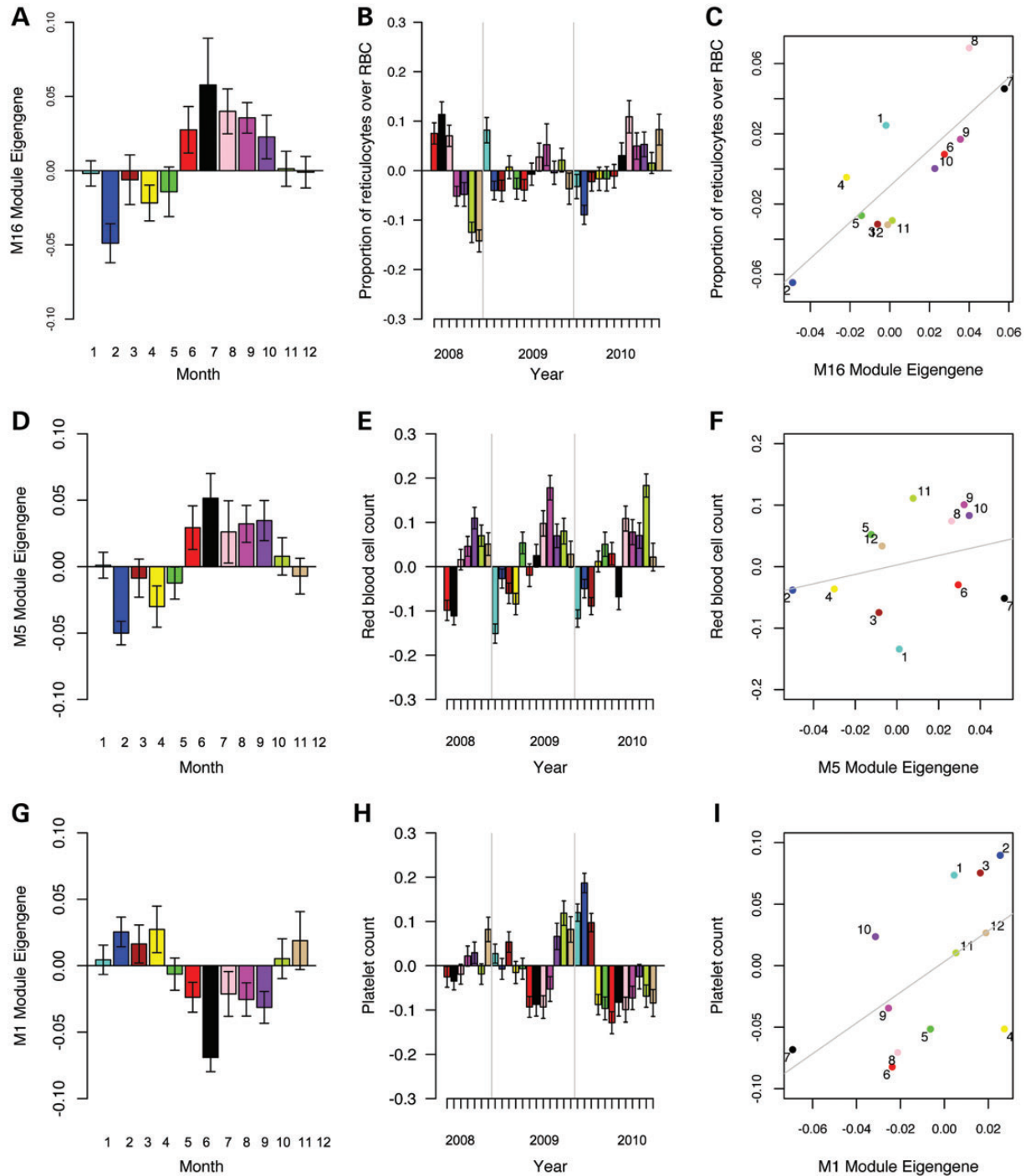


Figure 2. Seasonal patterns in gene co-expression modules and blood cell counts. The barplot in (A) depicts the module eigengene values of M16 (215 probes) per month. The barplots in (B) represent standardized red blood cell counts per month for 2008 (months 6–12), 2009 and 2010, respectively. The scatterplot in (C) has standardized red blood cell counts per month on the y-axis versus M16 module eigengene values per month on the x-axis. Colors for each month correspond in all three panels. The barplot in (D) depicts the module eigengene values of M5 (1664 probes) per month. The barplots in (E) represent standardized platelet counts per month for 2008 (months 6–12), 2009 and 2010, respectively. The scatterplot in (F) has standardized platelet counts per month on the y-axis versus M5 module eigengene values per month on the x-axis. Colors for each month correspond in all three panels. The barplot in (G) depicts the module eigengene values of M1 (3183 probes) per month. The barplots in (H) represent standardized proportion of reticulocytes over RBC per month for 2008 (months 6–12), 2009 and 2010, respectively. The scatterplot in (I) has standardized proportion of reticulocytes over RBC per month on the y-axis versus M1 eigengene values per month on the x-axis. Colors for each month correspond in all three panels.

found to express some 1000 transcripts (25). Blood platelets are also anuclear but express around roughly 2000 transcripts, of which the majority is mitochondrial (29). A recent study even

demonstrates a complex transcriptional landscape in blood platelets (30). The fact that in our study mean platelet volume and platelet count show inverse correlations indicates that, in

contrast to RBC, the smaller mature platelets are responsible for the seasonal peak in gene expression. The lesser number of transcripts and the resulting lower variation might be a reason for a subtle seasonal gene expression signal to be picked up from reticulocytes and blood platelets and not WBC.

Although we find important genes involved in regulating circadian rhythm, *CLOCK*, *PER3* and *RORA* in modules related to season, we do not detect any formal evidence for enrichment of circadian clock genes in any of the modules associated with seasonality (31). Clock genes might have a tight transcriptional pattern related to day and night rhythmicity only and be less important for seasonal cycling. In addition, functional enrichment analyses do not implicate immune-related pathways in the seasonal modules, which suggests that the observed patterns of gene expression changes are unrelated to seasonal exposure of human influenza viruses or other infectious diseases. However, the cell-type markers and ferritin genes point to seasonal fluctuations in iron homeostasis. Ferritin is involved in cellular and organismal iron homeostasis through iron sequestration. Iron is necessary for transport, binding and release of oxygen and disruption of this process is implicated in pathogenesis of disease; e.g. hemochromatosis and diseases characterized by inflammation, infection, injury and repair (such as rheumatoid arthritis, vascular and neurodegenerative diseases) (28). Incidence and fatality of cardio- and cerebrovascular diseases have consistently shown a seasonal pattern with a peak in winter (6,7) possibly linked to the morphology and composition of RBC (9,10), which coincides with the results of the current study. It is unlikely that individuals with seasonal disease confound the current expression sample, since they were screened for cardiovascular, pulmonary and psychiatric disease before inclusion. The involvement of monocytes in iron recycling explains the seasonal fluctuation of monocytes corresponding to that of the M1 blood platelet module (28). It has been hypothesized that hemodilution due to warmer weather decreases the hematocrit values in humans, although measurements are sensitive to within-subject variation (14). In addition, food intake could be involved, since people tend to eat less and lighter food in warmer weather (11). Since subjects in the gene expression part of the study were included after overnight fasting, short-term influence of food intake on gene expression is unlikely.

These results once again highlight the importance of tissue and cell type specificity of gene expression profiles. In addition, it adds a temporal dimension to interpreting laboratory results of whole blood expression data regardless of disease. Researchers should be careful when sampling blood from cases and controls in different seasons. Our study focused on whole blood and observed seasonal differences in gene expression patterns linked to variation in the distribution of cell counts. Other types of tissue might be less temporally reactive, assuming that the patterns of seasonal transcriptional changes are primarily a reflection of circannual changes in the population of cell types in whole blood. Further study is needed to test whether seasonal patterns in gene expression can also be observed in other tissues or within (single) cell types.

One field of research that has to rely on tissue other than that primarily affected (the brain) is that of neurogenetics. It is beyond the scope of this paper to address whether the detected seasonal pattern in blood differs between neuropsychiatric cases and controls, although existing literature hints at a

relationship between season on the one hand and disease epidemiology and possibly disturbed seasonal rhythms on the other hand in cases (15–17). Intuitively, case control differences in neuropsychiatric disorders might not necessarily be expected in properties of RBC or blood platelets. However, serotonin is carried by blood platelets and has been shown to exhibit seasonal fluctuation (21,22). A recent study even finds a direct relationship between platelet count and suicidal thoughts (32). In addition, a common variant in the neurogranin (*NRGN*) gene is associated with schizophrenia (33) and was identified as an intracellular platelet expressed gene highlighted by Gnatenko and colleagues (29). In accordance with this, this gene is assigned to the M1 platelet module in our gene expression results. Finally, medication can influence blood cell counts and several psychoactive drugs, e.g. clozapine, may incur hematological toxicity (34). With regard to healthy individuals, it would be of interest to examine seasonal fluctuations or the absence thereof in geographic locations that have the reverse or no seasonal pattern.

In conclusion, our results indicate a seasonal effect in whole blood gene expression, which is likely caused by fluctuations in whole blood cell composition. This phenomenon may have clinical relevance for incidence and severity of disease but can also be a potential confounding factor in genetic research utilizing whole blood samples.

MATERIALS AND METHODS

Subjects

Two hundred and forty neurological healthy volunteers were recruited as part of a previously published study (20,35) at outpatient pre-operative screening services in four hospitals in The Netherlands, from August 2008 until June 2011. Inclusion criteria were 18–60 years of age and four grandparents born in The Netherlands or other North-Western European countries. Exclusion was based on presence of psychotic or major neurological disorders any use of psychotropic medication (by personal telephone interview). Subjects had fasted at least 6 h before sample collection. Whole blood RNA was collected in PAXgene tubes. All participants gave written informed consent. Medical Research Ethics Committee of the University Medical Center Utrecht, The Netherlands and all local ethics committees approved this study.

Genome-wide expression profiling

Gene expression data for this sample set have been described before (35). In short, for isolation and purification of mRNA from whole blood, the PAXgene extraction kit (Qiagen) was used. Genome-wide RNA expression profiling was obtained with the Illumina HumanRef-12 version 3 arrays at the UCLA Neuroscience Genomics Core. Raw data extraction and background correction were performed using GenomeStudio. The Lumi package in R was used for robust spline normalization and variance stabilizing transformation of the gene expression data (36,37). A probe detection filter was applied to select those probes that were expressed in at least one sample, with the detection *P*-value as generated by GenomeStudio to $P < 0.01$. Our gene expression dataset consisted of three batches

based on Illumina serial number, normalization between these batches was performed with ComBat (38).

Cosine calculation

For every subject, month of blood draw was recorded (Table 1). A biannual pattern, representing season was calculated using the cosine function (21) for July through December (remaining months represent the opposite correlation by definition). As an example, for a peak in July and trough in January this would be $\cos((2 \times \pi/12) \times (m + 5))$ with m being the month # for each month of blood draw with addition of 5 to position the peak in July. Formulae for all months are given in Supplementary Material, Information 1. Cosine outcome was taken as a phenotype in subsequent analyses.

Weighted gene co-expression analysis

Gene expression data were analyzed using WGCNA (39–41). Weighted co-expression networks are based on the matrix of pairwise Pearson correlation coefficients that were constructed using the soft thresholding approach with a $\beta = 6$. This method entails raising each correlation to a fixed power by the criterion described by Zhang and Horvath (41). Soft thresholding resulted in a weighted adjacency matrix containing signed pairwise connection strengths.

We define co-expression modules as branches of a hierarchical clustering tree. Specifically, we used average linkage hierarchical clustering with a dissimilarity measure derived from the topological overlap matrix to define a cluster tree. This topological overlap measure is calculated based on the number of shared neighbors (41,42). A dendrogram is produced by hierarchical clustering of $1 - \text{topological overlap}$; branches of the tree are cut using a dynamic tree cut algorithm to define modules. For branch cutting (module detection), we used the dynamic branch-cutting algorithm implemented in the dynamicTreeCut and WGCNA R library (40,43). Each module (or branch) is subsequently assigned a numeric label based on module size, which is visualized underneath the cluster tree.

Table 1. Number of subjects per month

Month collection	Expression	UPOD 2008	UPOD 2009	UPOD 2010
January	39	(36)	2360	2785
February	29	(57)	2030	2276
March	24	(89)	1808	2469
April	18	(61)	1661	1708
May	17	(57)	1592	1369
June	19	1996	1547	1475
July	6	2598	1399	1096
August	9	1865	1204	1270
September	17	1964	1263	1293
October	19	1678	1293	1246
November	21	1586	1292	1366
December	15	1387	1162	1104
Total	233	13 074	18 611	19 457

The table contains the number of subjects collected per month, for the expression data and each of the 3 years in the UPOD. Brackets indicate months that have not been included due to limited subjects.

To define a representative module expression profile for each module, we summarized the (standardized) gene expression profiles of the module by their first principal component. This statistic is referred to as the module eigengene. The module eigengene can be thought of as an average gene expression value for all probes in a module per sample. Lists of blood cell-type-specific markers were used as input for the userListEnrichment function incorporated in the WGCNA package. This compares the number of overlapping genes to the maximal possible overlap (26).

Blood composition measurements

Data from the Utrecht Patient Oriented Database (UPOD) were used. UPOD is an infrastructure of relational databases comprising data on patient characteristics, hospital discharge diagnoses, medical procedures, medication orders and laboratory tests for all patients treated at the University Medical Center Utrecht (UMC Utrecht) since 2004. The UMC Utrecht is a 1042-bed academic teaching hospital in the Netherlands, with ~28 000 clinical and 15 000 day care hospitalizations and 334 000 outpatient visits annually. UPOD data acquisition and management are in accordance with current regulations concerning privacy and ethics. The structure and content of UPOD have been described in more detail elsewhere (44). In addition to clinical laboratory data from the laboratory information system, UPOD contains a specific database with hematology data on automated blood cell analyses performed with Abbott Cell-Dyn Sapphire (Abbott Diagnostics, Santa Clara, CA, USA) automated blood cell analyzers used at UMC Utrecht. A feature of this hematology analyzer is that all parameters of the complete blood count are measured irrespective of the requested parameter (45).

We obtained laboratory measurements of patients in 2008 ($n = 174\,706$), 2009 ($n = 166\,133$) and 2010 ($n = 177\,015$). For each measure, we obtained information about age, sex, date of blood draw and the department for which blood was drawn. All departments in the University Medical Center Utrecht measure blood cell counts and, therefore, the data are not biased by hematology department referral. Blood cell parameters included red blood cell count (RBC 10^{-12} l of blood), mean corpuscular volume (MCV 10^{-15} l), platelet count (10^{-9} l of blood), mean platelet volume (MPV 10^{-15} l) and reticulocytes (10^{-15} l). Reticulocytes were taken as proportion of total RBC. Parameters also included leukocytes (white blood cells, WBC), neutrophils, basophils, eosinophils, lymphocytes and monocytes (all in 10^{-9} l of blood). The different types of WBC were converted to a percentage of total WBC.

We performed rigorous quality control in order to select the most complete dataset for the healthiest adult subjects. An overview of all quality control steps and the number of included individuals is outlined in Supplementary Material, Information 2. We disregarded departments with means outside the normal range and $>2 \times \text{SD}$ (list of included/excluded departments and normal values in Supplementary Material, Informations 3 and 4). We also excluded months 1–5 of the year 2008 due to too few observations ($n < 100$) per month. This resulted in a dataset with $n = 13\,074$ for 2008, $n = 18\,611$ for 2009 and $n = 19\,457$ for 2010. Principal component analysis revealed no systematic effect of admitting department (Supplementary Material, Information 5). For each year, 43% of the individuals were male and the mean age was 44 years old. The number of

subjects per month for each year is listed in Table 1. Cross-sectional analysis was performed on this dataset by correlating the mean module eigengene values per month with mean standardized blood cell values for each month over 3 years. Data were standardized (centered by subtracting column mean from each column and scaled by dividing each column by its sample SD). To confirm a similar seasonal pattern, we performed a linear regression with cosine function peaking in August on unstandardized blood cell counts.

For the longitudinal analyses, we divided the year into two seasons, based on seasonal pattern in gene expression: the first season (months 12–5) and second season of the year (months 6–11). Quality control procedures were applied as described above, except for allowing multiple observations per individual. To prevent selecting for the least healthy subjects, we only included individuals with a maximum of two measurements per year, and with at least one measurement per season. This implies that the number of measurements per subject ranges from 3 to 6. Values per season were averaged over the 3 years. This selection resulted in 2459 individuals (50% male and mean age of 45 years). The difference of the second half versus the first half of the year was tested using a Wilcoxon rank-sum test using the MASS package for R (46).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the participants and staff at the various recruitment centres and Esther Janson and Joe DeYoung for technical assistance.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by National Institutes of Health funding R01DA028526 and R01NS058980 to R.A.O.

REFERENCES

- Harris, S. and Dawson-Hughes, B. (1993) Seasonal mood changes in 250 normal women. *Psychiatry Res.*, **49**, 77–87.
- Okawa, M., Shirakawa, S., Uchiyama, M., Oguri, M., Kohsaka, M., Mishima, K., Sakamoto, K., Inoue, H., Kamei, K. and Takahashi, K. (1996) Seasonal variation of mood and behaviour in a healthy middle-aged population in Japan. *Acta Psychiatr. Scand.*, **94**, 211–216.
- Chotai, J., Smedh, K., Johansson, C., Nilsson, L.G. and Adolffson, R. (2004) An epidemiological study on gender differences in self-reported seasonal changes in mood and behaviour in a general population of northern Sweden. *Nord. J. Psychiatry*, **58**, 429–437.
- Bloom-Feshbach, K., Alonso, W.J., Charu, V., Tamerius, J., Simonsen, L., Miller, M.A. and Viboud, C. (2013) Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PLoS ONE*, **8**, e54445.
- Alstadhaug, K.B., Salvesen, R. and Bekkelund, S.I. (2005) Seasonal variation in migraine. *Cephalalgia*, **25**, 811–816.
- Manfredini, R., Fabbian, F., Pala, M., Tiseo, R., De Giorgi, A., Manfredini, F., Malagoni, A.M., Signani, F., Andreati, C., Boari, B. *et al.* (2011) Seasonal and weekly patterns of occurrence of acute cardiovascular diseases: does a gender difference exist? *J. Womens Health*, **20**, 1663–1668.
- Khan, F.A., Engstrom, G., Jerntorp, I., Pessah-Rasmussen, H. and Janzon, L. (2005) Seasonal patterns of incidence and case fatality of stroke in Malmo, Sweden: the STROMA study. *Neuroepidemiology*, **24**, 26–31.
- Shen, H., Li, B., Bai, B., Hou, J., Xu, Z., Zhao, M., Mao, P., Liu, A., Li, J., Guo, T. *et al.* (2013) Laboratory features throughout the disease course of influenza A (H1N1) virus infection. *Clin. Lab.*, **59**, 337–342.
- Kurt, M., Tamboga, I.H., Buyukkaya, E., Karakas, M.F., Akcay, A.B. and Sen, N. (2013) Relation of Red Cell Distribution Width With CHA2DS2-VASc Score in Patients With Nonvalvular Atrial Fibrillation. *Clin. Appl. Thrombosis/Hemostasis*. [Epub ahead of print].
- Gaillard, C.A. and Schiffelers, R.M. (2013) Red blood cell: barometer of cardiovascular health? *Cardiovasc. Res.*, **98**, 3–4.
- Yanai, M., Satomura, A., Uehara, Y., Murakawa, M., Takeuchi, M. and Kumasaka, K. (2008) Circannual rhythm of laboratory test parameters among chronic haemodialysis patients. *Blood Purif.*, **26**, 196–203.
- Touitou, Y., Touitou, C., Bogdan, A., Reinberg, A., Auzéby, A., Beck, H. and Guillet, P. (1986) Differences between young and elderly subjects in seasonal and circadian variations of total plasma proteins and blood volume as reflected by hemoglobin, hematocrit, and erythrocyte counts. *Clin. Chem.*, **32**, 801–804.
- Frohlich, M., Sund, M., Russ, S., Hoffmeister, A., Fischer, H.G., Hombach, V. and Koenig, W. (1997) Seasonal variations of rheological and hemostatic parameters and acute-phase reactants in young, healthy subjects. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 2692–2697.
- Thirup, P. (2003) Haematocrit: within-subject and seasonal variation. *Sports Med.*, **33**, 231–243.
- Faedda, G.L., Tondo, L., Teicher, M.H., Baldessarini, R.J., Gelbard, H.A. and Floris, G.F. (1993) Seasonal mood disorders. Patterns of seasonal recurrence in mania and depression. *Arch. Gen. Psychiatry*, **50**, 17–23.
- Friedman, E., Gyulai, L., Bhargava, M., Landen, M., Wisniewski, S., Foris, J., Ostacher, M., Medina, R. and Thase, M. (2006) Seasonal changes in clinical status in bipolar disorder: a prospective study in 1000 STEP-BD patients. *Acta Psychiatr. Scand.*, **113**, 510–517.
- Ajdacic-Gross, V., Bopp, M., Ring, M., Gutzwiller, F. and Rossler, W. (2010) Seasonality in suicide – a review and search of new concepts for explaining the heterogeneous phenomena. *Social Sci. Med.*, **71**, 657–666.
- Heerlein, A., Valeria, C. and Medina, B. (2006) Seasonal variation in suicidal deaths in Chile: its relationship to latitude. *Psychopathology*, **39**, 75–79.
- Deisenhammer, E.A. (2003) Weather and suicide: the present state of knowledge on the association of meteorological factors with suicidal behaviour. *Acta Psychiatr. Scand.*, **108**, 402–409.
- Luykx, J.J., Bakker, S.C., van Geloven, N., Eijkemans, M.J., Horvath, S., Lentjes, E., Boks, M.P., Strengman, E., DeYoung, J., Buizer-Voskamp, J.E. *et al.* (2013) Seasonal variation of serotonin turnover in human cerebrospinal fluid, depressive symptoms and the role of the 5-HTTLPR. *Transl. Psychiatry*, **3**, e311.
- Luykx, J.J., Bakker, S.C., Lentjes, E., Boks, M.P., van Geloven, N., Eijkemans, M.J., Janson, E., Strengman, E., de Lepper, A.M., Westenberg, H. *et al.* (2012) Season of sampling and season of birth influence serotonin metabolite levels in human cerebrospinal fluid. *PLoS ONE*, **7**, e30497.
- Jakovljevic, M., Muck-Seler, D., Pivac, N., Ljubicic, D., Bujas, M. and Dodig, G. (1997) Seasonal influence on platelet 5-HT levels in patients with recurrent major depression and schizophrenia. *Biol. Psychiatry*, **41**, 1028–1034.
- Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A. and Brown, P.O. (2003) Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. USA*, **100**, 1896–1901.
- Gnatenko, D.V., Dunn, J.J., Schwedes, J. and Bahou, W.F. (2009) Transcript profiling of human platelets using microarray and serial analysis of gene expression (SAGE). *Methods Mol. Biol.*, **496**, 245–272.
- Kabanova, S., Kleinbongard, P., Volkmer, J., Andree, B., Kelm, M. and Jax, T.W. (2009) Gene expression analysis of human red blood cells. *Int. J. Med. Sci.*, **6**, 156–159.
- Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R. and Horvath, S. (2011) Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics*, **12**, 322.
- Gnatenko, D.V., Zhu, W., Xu, X., Samuel, E.T., Monaghan, M., Zarrabi, M.H., Kim, C., Dhundale, A. and Bahou, W.F. (2010) Class prediction models of thrombocytosis using genetic biomarkers. *Blood*, **115**, 7–14.

28. Torti, F.M. and Torti, S.V. (2002) Regulation of ferritin genes and protein. *Blood*, **99**, 3505–3516.
29. Gnatenko, D.V., Dunn, J.J., McCorkle, S.R., Weissmann, D., Perrotta, P.L. and Bahou, W.F. (2003) Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood*, **101**, 2285–2293.
30. Bray, P.F., McKenzie, S.E., Edelstein, L.C., Nagalla, S., Delgrosso, K., Ertel, A., Kupper, J., Jing, Y., Londin, E., Loher, P. *et al.* (2013) The complex transcriptional landscape of the anucleate human platelet. *BMC Genomics*, **14**, 1.
31. Moller-Levet, C.S., Archer, S.N., Bucca, G., Laing, E.E., Slak, A., Kabiljo, R., Lo, J.C., Santhi, N., von Schantz, M., Smith, C.P. and Dijk, D.J. (2013) Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl. Acad. Sci. USA*, **110**, E1132–E1141.
32. Ragolsky, M., Shimon, H., Shalev, H., Weizman, A. and Rubin, E. (2013) Suicidal thoughts are associated with platelet counts in adolescent inpatients. *J. Child Adolesc. Psychopharmacol.*, **23**, 49–53.
33. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B. *et al.* (2009) Common variants conferring risk of schizophrenia. *Nature*, **460**, 744–747.
34. Nooijen, P.M., Carvalho, F. and Flanagan, R.J. (2011) Haematological toxicity of clozapine and some other drugs used in psychiatry. *Hum. Psychopharmacol.*, **26**, 112–119.
35. Luykx, J.J., Bakker, S.C., Lentjes, E., Neeleman, M., Strengman, E., Mentink, L., DeYoung, J., de Jong, S., Sul, J.H., Eskin, E. *et al.* (2013) Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid. *Mol. Psychiatry*. [Epub ahead of print].
36. Du, P., Kibbe, W.A. and Lin, S.M. (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24**, 1547–1548.
37. Lin, S.M., Du, P., Huber, W. and Kibbe, W.A. (2008) Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.*, **36**, e11.
38. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
39. Horvath, S. (2011) *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer, New York.
40. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
41. Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, **4**, Article17.
42. Yip, A.M. and Horvath, S. (2007) Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, **8**, 22.
43. Langfelder, P., Zhang, B. and Horvath, S. (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, **24**, 719–720.
44. ten Berg, M.J., Huisman, A., van den Bemt, P.M., Schobben, A.F., Egberts, A.C. and van Solinge, W.W. (2007) Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin. Chem. Lab. Med.*, **45**, 13–19.
45. Muller, R., Mellors, I., Johannessen, B., Aarsand, A.K., Kiefer, P., Hardy, J., Kendall, R. and Scott, C.S. (2006) European multi-center evaluation of the Abbott cell-Dyn sapphire hematology analyzer. *Lab. Hematol.*, **12**, 15–31.
46. Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S*. Springer, New York.
47. Amr, M. and Volpe, F.M. (2012) Seasonal influences on admissions for mood disorders and schizophrenia in a teaching psychiatric hospital in Egypt. *J. Affect. Disord.*, **137**, 56–60.
48. de Jong, S., Boks, M.P., Fuller, T.F., Strengman, E., Janson, E., de Kovel, C.G., Ori, A.P., Vi, N., Mulder, F., Blom, J.D. *et al.* (2012) A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PLoS ONE*, **7**, e39498.
49. Kurian, S.M., Le-Niculescu, H., Patel, S.D., Bertram, D., Davis, J., Dike, C., Yehyaw, N., Lysaker, P., Dustin, J., Caligiuri, M. *et al.* (2011) Identification of blood biomarkers for psychosis using convergent functional genomics. *Mol. Psychiatry*, **16**, 37–58.
50. Tsuang, M.T., Nossova, N., Yager, T., Tsuang, M.M., Guo, S.C., Shyu, K.G., Glatt, S.J. and Liew, C.C. (2005) Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **133B**, 1–5.
51. Altar, C.A., Vawter, M.P. and Ginsberg, S.D. (2009) Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology*, **34**, 18–54.
52. Booij, B.B., Lindahl, T., Wetterberg, P., Skaane, N.V., Saebo, S., Feten, G., Rye, P.D., Kristiansen, L.I., Hagen, N., Jensen, M. *et al.* (2011) A gene expression pattern in blood for the early detection of Alzheimer's disease. *J. Alzheimer's Dis.*, **23**, 109–119.
53. Saris, C.G., Horvath, S., van Vught, P.W., van Es, M.A., Blauw, H.M., Fuller, T.F., Langfelder, P., DeYoung, J., Wokke, J.H., Veldink, J.H. *et al.* (2009) Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics*, **10**, 405.