

Genome-wide analysis of blood gene expression in migraine implicates immune-inflammatory pathways

Cephalalgia

0(0) 1–12

© International Headache Society 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0333102416686769

journals.sagepub.com/home/cep



Zachary F Gerring¹, Joseph E Powell^{2,3}, Grant W Montgomery²
and Dale R Nyholt¹

Abstract

Background: Typical migraine is a frequent, debilitating and painful headache disorder with an estimated heritability of about 50%. Although genome-wide association (GWA) studies have identified over 40 single nucleotide polymorphisms associated with migraine, further research is required to determine their biological role in migraine susceptibility. Therefore, we performed a study of genome-wide gene expression in a large sample of 83 migraine cases and 83 non-migraine controls to determine whether altered expression levels of genes and pathways could provide insights into the biological mechanisms underlying migraine.

Methods: We assessed whole blood gene expression data for 17994 expression probes measured using IlluminaHT-12 v4.0 BeadChips. Differential expression was assessed using multivariable logistic regression. Gene expression probes with a nominal p value < 0.05 were classified as differentially expressed. We identified modules of co-regulated genes and tested them for enrichment of differentially expressed genes and functional pathways using a false discovery rate < 0.05 .

Results: Association analyses between migraine and probe expression levels, adjusted for age and gender, revealed an excess of small p values, but there was no significant single-probe association after correction for multiple testing. Network analysis of pooled expression data identified 10 modules of co-expressed genes. One module harboured a significant number of differentially expressed genes and was strongly enriched with immune-inflammatory pathways, including multiple pathways expressed in microglial cells.

Conclusions: These data suggest immune-inflammatory pathways play an important role in the pathogenesis, manifestation, and/or progression of migraine in some patients. Furthermore, gene-expression associations are measurable in whole blood, suggesting the analysis of blood gene expression can inform our understanding of the biological mechanisms underlying migraine, identify biomarkers, and facilitate the discovery of novel pathways and thus determine new targets for drug therapy.

Keywords

Migraine, gene, expression, blood, inflammation

Date received: 24 August 2016; revised: 17 October 2016; 25 November 2016; accepted: 30 November 2016

Introduction

Migraine is a complex genetic disorder marked by episodic attacks of moderate to severe headache lasting between 4 and 72 hours, often unilateral and pulsating, and associated with nausea, photophobia, and phonophobia. A large minority of attacks are also inaugurated by visual, auditory, or other sensory hallucinations, and in some instances neurological deficits that affect speech and ideation (1). Migraine has a lifetime prevalence of 15–20% and most commonly affects people aged from 25 to 50, during their most

¹Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

²Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

³The Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

Corresponding author:

Dale R Nyholt, Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia.

Email: d.nyholt@qut.edu.au

productive years of life (2,3). As such, migraine not only has a significant impact on the quality of life of individual sufferers, but also represents a major economic and societal burden.

There is a diversity of evidence on the aetiology and pathophysiology of migraine. While some investigators have proposed migraine is initiated and sustained in the central nervous system, most consider the disorder to be neurovascular – a term that couples metabolic changes in the brain with dysfunctional intracranial and extracranial blood vessels. Heritable genetic factors explain 40–60% of the individual differences in developing migraine (4–7). Large genome-wide association (GWA) studies have identified over 40 common single nucleotide polymorphism (SNP) loci associated with the disorder, several of which have previous associations with vascular disease, and implicated regions show enrichment for genes expressed in vascular and smooth muscle tissues (8). Although these results provide important genomic support for vascular dysfunction being a primary mechanism underlying migraine, their biological role in migraine susceptibility is largely unknown. Analysis of other molecular data such as gene expression may provide insights into the biological mechanisms underlying migraine.

Gene expression is modified by genomic and environmental factors and has facilitated the discovery of molecular pathways for common neurological disorders (9,10). A recent gene expression study in the periosteum of 18 chronic migraine cases and seven non-migraine controls found differential expression of genes involved in the inflammatory response (11). These data provide some of the first evidence for an extracranial vascular mechanism in migraine and reveal a potential pathway for the development of targeted therapies and diagnostic biomarkers. However, the study only assessed inflammatory genes and included a small sample of highly selected clinic-based subjects, so neither the genome-wide significance nor the generalisability of their findings to the general population can be established.

These limitations may be overcome with the analysis of genome-wide gene expression in the whole blood of migraine subjects (12). Blood expresses many of the genes involved in the pathophysiology of complex neurological traits, and can be collected from a large number of individuals in a population-based setting. Although caveats exist (13), several small blood gene expression studies have shown clear separation of expression profiles based on migraine status (14), migraine subtype (15), and treatment response (16), suggesting blood gene expression signatures may provide valuable information in the assessment of migraine activity and risk. We therefore examined gene expression measured on 17994 probes from the whole blood

of 166 unrelated individuals. Both the number of subjects studied and expression probes measured represent a large improvement on the existing literature, and provides increased power to identify expression patterns underlying migraine.

Methods

Brisbane Systems Genetics Study

Participants were recruited from the Brisbane Systems Genetics Study (BSGS), a genetic study of 962 individuals from 314 families of Northern European descent (17). Families consisted of adolescent monozygotic and dizygotic twin pairs, their non-twin siblings, and their parents. Informed written consent was obtained from each participant, and the study was approved by the Human Research Ethics Committee (HREC) of the QIMR Berghofer (QIMRB) Medical Research Institute.

RNA preparation

Whole blood samples were collected from participants and processed within 24 hours of collection. Whole blood was collected directly into PAXgeneTM Blood RNA Tubes (QIAGEN, Valencia, CA). Total RNA was extracted from PAXgeneTM tubes using the PAXgeneTM Blood RNA Kit (QIAGEN, Valencia, CA). RNA from all samples was run on an Agilent Bioanalyzer to assess quality and to estimate RNA concentrations. RNA was converted to cDNA, and amplified and purified using the Ambion Illumina TotalPrep RNA Amplification Kit (Ambion).

Gene expression quantification

Expression profiles were generated by hybridising 750 ng of fragmented biotinylated cRNA (generated from 500 ng of total RNA) to Illumina HumanHT-12 v4.0 Beadchips for 18 h at 58°C, according to the Illumina whole-genome gene expression direct hybridisation assay guide (Illumina Inc, San Diego, USA). Beadchips were then washed, stained and subsequently scanned using an Illumina Bead Array Reader to obtain fluorescence intensities. A Latin square design was used to randomise the samples on the chips and chip positions.

Data processing and normalisation

Preprocessing and normalisation was previously described by Powell et al. (18). Pre-processing of data generated by the Illumina Bead Array Reader was done using Illumina software, Genome Studio (Illumina Inc.,

Table 1. Demographic information for patients with gene expression data.

Dataset	Number (females)		Mean age (Std Dev)		
	Cases	Controls	Case	Control	Total
Male adults	15	15	47.88 (9.33)	46.96 (5.49)	47.42 (7.54)
Female adults	23	23	46.12 (5.72)	45.25 (4.61)	45.69 (5.16)
Male adolescents	17	17	14.37 (2.2)	13.84 (2.25)	14.11 (2.2)
Female adolescents	28	28	14.22 (2.3)	14.8 (2.54)	14.51 (2.42)
All subjects	83	83	29.17 (17.11)	28.85 (16.24)	29.01 (16.63)

San Diego). Background noise detected from negative control beads was subtracted from raw expression values for each transcript. Data were then filtered for gene transcripts that were present in at least 50% of samples at $p < 0.05$ according to the global-error threshold calculated by Genome Studio's cross-gene error model. Filtering based on this criterion removes genes that are either not expressed or only expressed in a low number of samples. After removing 6322 putative and/or poorly characterised genes, 17994 well-characterised detected probes remained for analysis, which corresponds to 13355 RefSeq genes. Probe signal intensities were stabilised using a quantile transformation procedure (19) and normalised with a linear mixed model to allow comparisons across chips and genes (20). The dataset is available at the Gene Expression Omnibus under the accession code GSE53195.

Migraine phenotype

The ID MigraineTM (21) screening questionnaire was used to identify migraine cases. The questionnaire contains three items related to the presence of photophobia, nausea, and health-related disability. A respondent scores positive to migraine if they endorse two out of the three items. The diagnostic performance of the questionnaire has been validated in a clinical setting (21), and has an estimated sensitivity of 0.84 (95% confidence interval (CI): 0.75–0.90) and specificity of 0.76 (95% CI: 0.69–0.83). The accuracy of ID migraineTM is similar across sex, age, presence of other co-morbid headaches, or existing diagnostic status (22).

Migraine groups

A total of 38 unrelated adults (aged 34–74 years) and 90 unrelated adolescents (aged 10–20 years) screened positive for migraine. Migraine cases were stratified by age – that is, whether an individual was recruited into the BSGS as an adolescent twin (or non-twin sibling) or as the parent of a twin – and gender. Each case was age and gender matched with an unrelated individual from

the BSGS study who screened negative for migraine, thereby generating four independent, non-overlapping, analytical cohorts: male adults ($n = 30$), female adults ($n = 46$), male adolescents ($n = 34$), and female adolescents ($n = 56$) (Table 1). Patient matching by age and sex increases statistical efficiency of case-control analyses (23,24), and the unrelated structure between the cohorts enabled the unbiased assessment of gene expression between non-overlapping datasets.

Association of gene expression levels with migraine

Genome-wide differential gene expression between migraine and control groups was performed using generalised linear models for each of the 17994 gene expression probes, adjusted for age and gender. An ANOVA function was used to test the change in model deviance obtained by adding gene expression as a covariate. To capture a reasonable number of genes for downstream analyses, a nominal significance threshold $p < 0.05$ was used. Ranked lists of differentially expressed genes were compared across independent cohorts using a one-sided binomial test. The global distribution of differential gene expression was examined by generating Q-Q plots of p values. Q-Q plots were constructed by ranking the p values from smallest to largest (the 'order statistics'), and plotting them against their expected values under the null hypothesis of no association (sampled from the known chi-squared distribution). Deviations above the line of equality (drawn in white) indicate an excess of smaller p values. To aid interpretation, we have also calculated 95% confidence envelopes (shaded grey), formed by calculating, for each order statistic, the 2.5th and 97.5th centiles of the distribution of the order statistic under random sampling and the null hypothesis. The inflation factor (λ), defined as the ratio of the median observed chi-squared statistic divided by the test statistic of the expected median (0.4549), was also calculated to quantify the excess of smaller p values; where the expected λ value is 1 under the null hypothesis of no differential expression. To assess the potential impact of

correlated gene expression on the migraine association test statistics distribution, principal component analysis followed by varimax rotation of the probe-expression correlation matrix was used to extract the 3316 least correlated ('independent') probes that captured the majority of the variation in gene expression (i.e., loaded highest onto the largest 3316 principal components). The number of 'independent' probes (3316) to select was determined using the approach of Li and Ji (25).

Analysis of migraine-related genes in biological processes

We first collated up-to-date pathways and gene sets from Reactome (date accessed: 04/04/2016; number of pathways=674) (26), the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (04/04/2016; n=291) (27), Biocarta (07/04/2016; n=217) (28), and the Protein Interaction Database (PID) (07/04/2016; n=196) (29), as well as sets of genes expressed predominantly in glial (03/20/2016; n=71) (30) and neuronal cells (03/20/2016; n=18) (31,32). The highly expressed, cell specific (HECS) gene expression database was used to assess enrichment of cell types (03/20/2016; n=84) (33). We compared the distribution of *p* values from the differential expression analyses within each pathway against the *p* values for all other expressed genes using the Wilcoxon rank sum test. A false discovery rate (FDR < 0.05) was used to adjust for multiple testing (34), accounting for the number of assessed pathways.

Identification of gene expression networks (modules)

Gene co-expression modules were constructed using the weighted gene co-expression network analysis (WGCNA) package in R (35). Missing data were removed and a single probe with the highest mean expression per gene was selected to be included in the analysis using the collapseRows function in WGCNA, resulting in 5472 unique RefSeq genes. An unsigned pairwise correlation matrix – using Pearson's product moment correlation coefficient – was generated for all genes and transformed into an adjacency matrix, where nodes correspond to genes and edges to the connection strength between genes. Each adjacency was raised to a power of 5 to place greater weight on large nodal relationships, and was normalised using a topological overlap function. Hierarchical clustering was done using average linkage, with one minus the topological overlap matrix as the distance measure. The hierarchical cluster tree was cut into modules using the dynamic tree cut algorithm (36),

with a minimum module size of 60 genes. We amalgamated modules if the correlation between their eigen-genes – defined as the first principal component of their expression values – was greater or equal to 0.8. Functional annotation of each gene module was performed by comparing the enrichment of Reactome pathways within a module compared to all other genes using Fisher's exact test.

Gene expression networks associated with differentially expressed genes

Enrichment of differentially expressed genes within a module was determined using a one-sided binomial test that calculated whether the proportion of differentially expressed genes in a given module is higher than expected based on the total proportion of differentially expressed genes. Differentially expressed genes may have similar expression patterns due to their co-expression, rather than a true biological effect. Therefore, we performed a second co-expression adjusted binomial test in which we included only the number of effectively independent genes in the calculation using principal component analysis. That is, the number of effectively independent genes was estimated utilizing the matSpD program (<http://neurogenetics.qimrberghofer.edu.au/matSpD/>) based on the method of Li and Ji (25). In this way, we were able to control the Type I error rate in the presence of correlated expression values. All statistical analyses were performed using the computing environment R (37).

Results

Construction of independent migraine cohorts

We identified four independent cohorts of migraine case and control subjects from the Brisbane Systems Genetics Study (BSGS). A total of 38 unrelated adults (aged 37–74 years) and 45 unrelated adolescents (aged 10–20 years) screened positive for migraine using the clinically validated ID migraineTM screening tool (21). We stratified migraine cases by age – that is, whether an individual was recruited into the BSGS as an adolescent twin (or non-twin sibling) or as a parent of a twin – and gender. Each case was then age and gender matched with an unrelated individual from the BSGS study who screened negative for migraine. This resulted in four independent, non-overlapping, datasets: male adults (n=30), female adults (n=46), male adolescents (n=34), and female adolescents (n=56) (Table 1). This analytical approach enabled the assessment of gene expression underlying migraine within multiple independent datasets.

Association between migraine status and gene expression

To assess differential expression between the migraine case and non-migraine control groups, we fitted generalised linear models for each of the 17994 probes within each dataset, thereby generating four independent lists of migraine association p values. No single probe achieved genome-wide significance ($p < 2.8 \times 10^{-6}$, Bonferroni correction for testing 17994 independent probes [i.e., $p < 0.05/17994$]). To examine the comparability of each dataset, we intersected top-ranked probes and observed a clear trend in overlap of differential expression (Figure 1). For example, of the 866 differentially expressed probes in adult females, 81 probes (9%) were also differentially expressed in adult males – significantly more than expected by chance (binomial test, $p = 0.0234$).

We therefore combined all datasets ($n = 166$ individuals, comprising 83 migraine cases and 83 non-migraine controls) and tested for an association between migraine status and gene expression using the same

regression model used for the individual datasets, but with the addition of three dummy variables included as covariates to designate samples belonging to one of the four independent datasets (male adults, female adults, male adolescents, female adolescents). A total of 12 probes were differentially expressed at the recommended false discovery rate (FDR) threshold < 0.1 (10), and a Q-Q plot showed an excess of migraine differential expression test with small p values (i.e., $p_{DE} < 0.05$ [$\log_{10}(P_{DE}) > 1.3$]) above the 95% confidence interval (shaded area) of the identity line (Figure 2), suggesting many of the differentially expressed genes represent true association signals of small effect. To assess whether correlated probe expression induced the observed excess of small p values, we performed principal component analysis to extract independent probes that capture the majority (75%) of variation. The resulting Q-Q plot and λ of 1.102 still indicated an excess of small p values, suggesting real biological effects underlie the observed associations (online Supplementary Material, Figure S1).

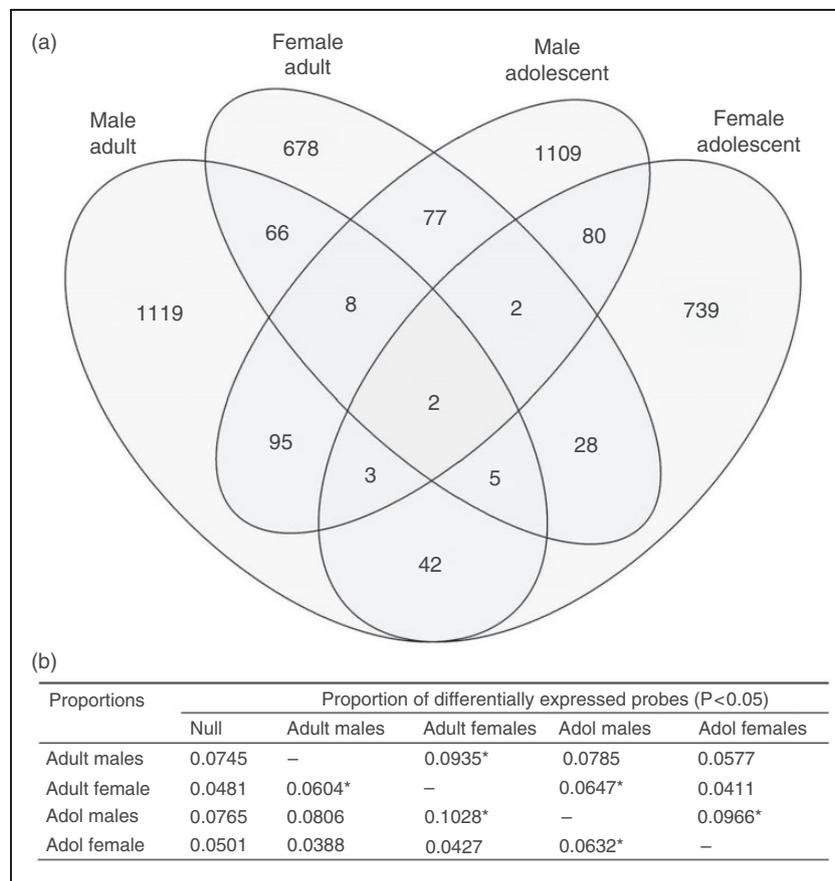


Figure 1. (a) Venn diagram of differentially expressed probes ($p < 0.05$) across four independent migraine groups. (b) Proportion of probe overlap for all pairwise comparisons.

Notes: The null expectation equals the observed proportion of differentially expressed probes for a given dataset. *significant overlap ($p < 0.05$).

Involvement of differentially expressed genes in biological pathways

For biological pathway annotation, we restricted the combined dataset to non-missing probes and selected a single probe with the highest mean expression across all subjects as the representative gene probe (5472 RefSeq genes). We used the Wilcoxon rank sum statistic to compare all ranked p values from the differential

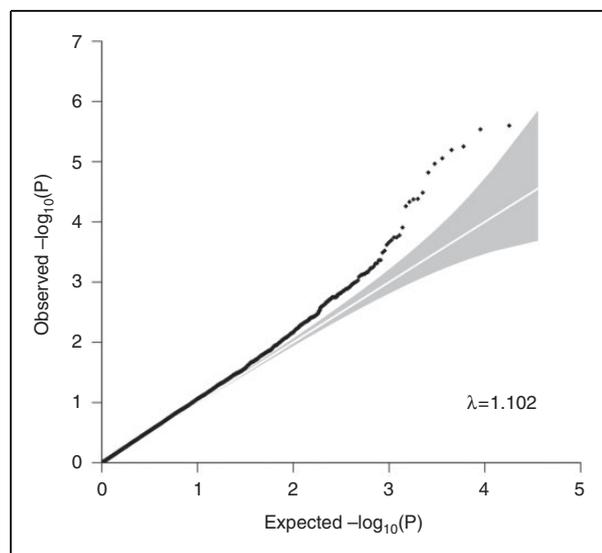


Figure 2. Q-Q plot of $-\log_{10} p$ values for differential expression between migraine cases and non-migraine controls ($n = 166$).

Notes: λ , inflation factor. The grey shaded area represents the 95% confidence interval for the expected distribution of p values under the null hypothesis of no association between migraine disease status and gene expression.

expression analysis within a given pathway, with all remaining p values outside the pathway. The ranked list of p values was enriched with six canonical pathways ($FDR < 0.05$), including Reactome ‘immune system’, ‘interferon signalling’ and ‘cytokine signalling in immune system’ (Table 2). We also tested for the enrichment of pathways expressed in microglial cells – the resident immune cells of the central nervous system – and found enrichment of genes involved in ‘cell death and apoptosis’.

Gene expression networks (modules) associated with differentially expressed genes

It is well-known that genes interact with one another in highly organised networks. This level of organisation can be estimated using pairwise gene correlations and clustering techniques. We applied Weighted Gene Co-expression Network Analysis (WGCNA) to the combined expression data and identified 10 gene co-expression modules. Using a binomial test, three modules, nominally labelled M1, M4 and M9, showed enrichment ($p_{\text{binomial}} < 0.05$) of differentially expressed genes ($p_{\text{DE}} < 0.05$) (Table 3). To ensure the observed module enrichments were not solely due to correlated gene expression, a second round of binomial tests, utilizing independent (i.e. uncorrelated) genes selected via principal component analysis was performed. Using the subset of independent genes, modules M1 and M4 still showed enrichment of genes differentially expressed between migraine cases and non-migraine controls. Module M1 showed the highest enrichment of differentially expressed genes (all genes $p_{\text{binomial}} = 4.17 \times 10^{-9}$; independent genes $p_{\text{binomial}} = 0.0003$), and was down-regulated in migraine cases compared to controls.

Table 2. Involvement of differentially expressed genes in biological pathways.

Pathway	Genes		Test for enrichment	
	Total	DE (%)	p value	FDR
<i>REACTOME</i>				
Immune system	457	36 (7.88)	8.51×10^{-5}	0.0270
SRP dependent membrane targeting	95	3 (3.16)	0.0001	0.0270
Interferon signalling	97	11 (11.34)	0.0002	0.0270
Cytokine signalling in immune system	158	18 (11.39)	0.0002	0.0309
Antiviral mechanism by IFN stimulated genes	49	5 (10.20)	0.0003	0.0309
Peptide Chain Elongation	75	2 (2.67)	0.0005	0.0427
<i>Microglia</i>				
Cell death and apoptosis	36	6 (16.67)	0.0008	0.0155

Notes: DE, differentially expressed; FDR, false-discovery rate; IFN, interferon, SRP, signal recognition particle. The test for pathway enrichment utilised the Wilcoxon rank sum test to measure the association between pathway membership and differential gene expression.

Table 3. Gene expression networks associated with differentially expressed genes.

Gene co-expression module		Test for enrichment of differentially expressed genes					
		All genes			Independent genes		
Label	Function	Total	DE (%)	p_{binomial}	Total	DE (%)	p_{binomial}
M1	Immune system	216	33 (15)	4.17×10^{-9}	117	20 (17)	0.0003
M2	Electron transport	1221	36 (3)	0.9994	523	31 (6)	0.9005
M3	B cell activation	769	38 (5)	0.4349	325	26 (8)	0.3354
M4	RNA metabolism	404	36 (9)	0.0003	174	23 (13)	0.0040
M5	Unspecified	244	16 (7)	0.1252	194	15 (8)	0.4411
M6	External plasma membrane	82	4 (5)	0.5534	44	4 (9)	0.3989
M7	Signal transduction	161	4 (2)	0.9514	80	4 (5)	0.8419
M8	Viral response	81	2 (2)	0.9035	49	2 (4)	0.8801
M9	Oxygen transporter activity	357	27 (8)	0.0133	176	17 (10)	0.1419
M10	RNA processing	1937	65 (3)	0.9991	945	49 (5)	0.9959

Notes: DE, differentially expressed. p_{binomial} is the p value for the binomial test comparing the observed proportion of DE genes to that expected by chance.

Table 4. Most highly-enriched pathways and gene sets in the migraine associated immune system module (M1).

Functional category	Total genes	Tested genes	Genes DE (%)	FDR
<i>Gene ontology biological process</i>				
Response to type I interferon	53	27	6 (22)	4.81×10^{-21}
Defence response to virus	66	38	10 (26)	4.81×10^{-21}
Type I interferon signalling pathway	33	26	6 (23)	2.82×10^{-19}
<i>REACTOME</i>				
Interferon signalling	159	32	8 (25)	3.46×10^{-18}
Interferon alpha/beta signalling	64	22	6 (27)	8.05×10^{-18}
Cytokine signalling in immune system	270	36	9 (25)	2.08×10^{-15}
<i>Microglia</i>				
Immune system	117	17	4 (24)	3.12×10^{-7}
Signal transduction	96	12	2 (17)	7.14×10^{-5}
Protein metabolism	76	10	1 (10)	0.0007
<i>HECS</i>				
CD56 NK cells	1054	59	12 (19)	5.49×10^{-9}
CD14 monocytes	888	53	11 (21)	2.62×10^{-8}
CD4 T cells	758	47	11 (23)	2.62×10^{-8}

Notes: 'Total genes' refers to the total number of genes assigned to the pathway; 'tested genes' refers to the number of total genes assigned to the pathway that were tested for differential expression; Genes DE, the number (and %) of genes differentially expressed; FDR, the significance of differential expression measured by the false discovery rate; HECS, highly expressed, cell specific; NK, natural killer. See Supplementary Table 1 for a full list of highly differentially expressed pathways and gene sets in the immune system module.

Pathway-based analyses using all 216 genes in the migraine associated 'immune system' module (M1) identified several immune-inflammatory functional categories and pathways, including Gene Ontology Biological Process 'response to type I interferon' (FDR = 4.81×10^{-21}), Reactome 'interferon signalling' (FDR = 3.46×10^{-18}), and microglia 'immune system'

(FDR = 3.12×10^{-7}), as well as enrichment of highly expressed, cell specific genes (HECS) in natural killer (NK) cells and monocytes (Table 4; online Supplementary Material, Table S1). Top-ranked pathways in module M4 included Gene Ontology Biological Process 'histone H4-K5 acetylation' (FDR = 0.791), and Reactome 'TCR signalling (FDR = 0.533)' and

‘cytosolic tRNA aminoacylation’ (FDR = 0.819). However, no single functional category or pathway survived correction for multiple comparisons (FDR < 0.05).

Discussion

We analysed migraine genome-wide gene expression data with the aim of identifying gene networks and pathways involved in migraine pathophysiology. Gene expression data measured in whole blood samples collected from 166 individuals from four independent cohorts showed global evidence for genes differentially expressed between migraine cases and non-migraine controls. A weighted gene co-expression network analysis (WGCNA) identified 10 gene co-expression modules. One module harboured a significant number of differentially expressed genes and was strongly enriched with immune and inflammatory pathways, including multiple pathways expressed in microglial cells.

The pathophysiology of migraine is restricted to a limited number of organ systems or tissues. Therefore, the choice of tissue for gene expression studies is an important consideration, particularly for neurological traits given the inaccessibility of nervous tissue from living individuals. Several recent studies of neurological disorders have used whole blood or its cellular constituents to measure changes in gene expression (10,38), citing biological and technical advantages offered by the tissue. From a biological perspective, blood is ubiquitous and transcriptionally active, and contains a diverse array of leukocytes that possess the receptors and pathways required to maintain homeostasis and mount a coordinated immune response to foreign substances and tissue damage. Technical advantages include the ability to collect blood samples from large population-based cohorts, rapidly preserve RNA upon extraction, and easily quantify cellular heterogeneity (12). Matched comparisons of human blood and brain tissues suggest the vast majority of genes expressed in the brain are also expressed in blood (39), and it has been shown that genes with similar detection levels between the tissues have highly correlated expression levels (40). Thus, whole blood is likely to provide biologically meaningful information for neurological traits such as migraine.

Despite these advantages, several important caveats of blood genomic profiling in complex neurological traits should be noted. First, blood gene expression is confounded by multiple biological and technical factors that represent a substantial proportion of expression variation and may lead to spurious associations (41). These sources of variation can be minimised with appropriate normalisation strategies to correct for inconsistencies between arrays, as well as the strict

control of sample collection and laboratory conditions (42). Second, many genes are differentially expressed between tissues (39), so the transcriptional profile of whole blood may not capture pathogenic changes in another tissue. Third, although pilot data from the Genotype Tissue Expression (GTEx) study indicates more than half of the genetic control of gene expression is common across different tissues and shows largely consistent effect directions (43), there are instances where genetic effects differ across tissues (44). This highlights the importance of a multi-tissue approach and the use of genomic compendia such as GTEx to help characterise expression variation observed in blood.

A modular-based approach based on gene co-expression patterns was used to identify groups of functionally related genes associated with migraine. This approach enabled us to reduce the dimensionality of genome-wide gene expression data without the loss of important biological information, and thereby alleviate the multiple testing burden associated with traditional gene-based methods. Similar methods have been successful for other neurological disorders, including post-traumatic stress syndrome (45), major depressive disorder (10), and psychosis (46). However, these studies ignored important methodological issues in the modular analysis of transcriptional data. Notably, the use of traditional 2×2 table methods such as the chi-squared test to test for over-representation of biologically relevant genes (for example, differentially expressed genes or previously implicated candidate genes) is not statistically appropriate in a modular paradigm. That is, the underlying assumption of these tests – that the gene expression estimates are independent – is not met, due to the high correlation among genes within a module (compared to all remaining genes), and thus may result in biased test statistics with an inflated Type I error. Therefore, methods that include only the number of independent genes in a dataset, or adjust for correlated expression using, for example, permutation (47) or bootstrap resampling (48), should be applied in modular analysis to maintain control of the Type I error rate. In this study, we estimated the effective number of independent genes within each module using principal component analysis to provide unbiased estimates of association in these highly correlated data.

One module (module M1) was down-regulated in migraine cases and significantly enriched with differentially expressed genes involved in immune-inflammatory pathways. Particularly strong enrichments were observed in immune pathways of microglial cells, the blood-derived immune cells of the central nervous system. Inflammatory pathways play an important role in migraine pathophysiology, evidenced by the continued use of anti-inflammatory drugs for acute

migraine relief (49). The precise role of inflammation in migraine is unclear, however a putative mechanism begins with the release of cellular stressors from the brain parenchyma, blood vessel walls, or both, which triggers the release of vasoactive neuropeptides near sensory fibres that innervate the meninges. These neuropeptides may subsequently bind and activate peripheral nociceptors at the level of the trigeminovascular system, leading to vasodilation, localised inflammation, and prolonged head pain (50). This mechanism is supported by the recent successful clinical trial of prophylactic treatment designed to block the release of a potent vasoactive peptide from the trigeminal nerve afferents in the early stages of an attack, thereby potentially increasing the activation threshold of the trigeminovascular system and its brain stem connections (51). These lines of evidence suggest experimental approaches that focus on immune-inflammatory genes, particularly those highly expressed in microglial cells, may identify more effective treatments for migraine and inform the development of biomarkers to monitor disease activity and progression.

Our results largely corroborate the conclusions of Perry *et al.* (11), who performed a targeted gene expression study of 540 inflammatory genes in the periosteum of 18 patients with chronic migraine and self-reported cranial muscle tenderness and seven controls. Of the 540 genes investigated, 36 were differentially expressed between migraine cases and controls and formed a gene network associated with activation of immune and inflammatory processes. We find nominal evidence of association for six out of 36 genes (*NFKB1Z*, *TNFSF10*, *TNFAIP3*, *CXCR4*, *ABC1*, and *NFIL3*) reported as differentially expressed, and our gene network is similarly enriched with immune system pathways. Our study expands upon the targeted analysis of inflammatory genes in a small group of highly selected patients performed by Perry *et al.*, by assessing genome-wide gene expression in the whole blood of 166 subjects. It has also been shown that blood gene expression profiles can differentiate chronic migraine and acute migraine from healthy controls (14), chronic migraine without medication overuse from patients with medication overuse (16), and menstrual-related migraine from non-menstrual-related migraine (15). These studies used small patient samples and older microarray technologies, limiting direct comparisons with our study. Nonetheless, taken together with our results and those of Perry *et al.*, these data provide further support for blood genomic studies of migraine.

In summary, we performed the largest known gene expression study of migraine and found multiple immune-inflammatory pathways underlying the disorder. Our results complement a recent expression study of migraine, and suggest immune processes play

an important role in migraine pathophysiology. However, this study should be interpreted in the context of several potential limitations.

First, we relied on the ID migraineTM tool to identify migraine cases and non-migraine controls, with no relation between last migraine attack and time of blood sampling. The ID migraineTM tool is a clinically validated screening instrument with an estimated sensitivity of 81% and specificity of 75%, and has been shown to positively identify 93–98% of people who were subsequently diagnosed with migraine by headache experts using International Headache Society (IHS) criteria (22). Although the developers of the ID migraineTM screening tool validated its utility in men and women aged 18 to 55 years old, subsequent studies have demonstrated its utility in adolescents (52,53). However, the age of onset for migraine with aura peaks between the ages of 12 and 13 in females, and on or before ages 4–5 in males, while the age of onset for migraine without aura peaks between the ages of 14 and 15 in females, and between ages 10 and 11 in males (54). Given the mean age of the male and female adolescent samples in our study is over 14 (Table 1), we expect that the majority of adolescents at risk of migraine would have experienced a migraine episode. In addition, the controls were selected as not having a family history of migraine. Nonetheless, the possibility remains for some controls being later diagnosed with migraine. Regardless, the consequence of controls being undiagnosed migraineurs would be to simply reduce our power to identify gene expression differences between migraine cases and controls (i.e., as it would make gene expression in controls more similar to cases). Ideally, future expression studies should confirm diagnoses using a semi-structured interview and IHS diagnostic criteria, and attempt to match sample collection with respect to the last migraine attack.

Second, the observed deviation of probe level test statistics from the null distribution may be due to correlation between probe expression values. However, the Q-Q plot of p values from the subset of 3316 independent probes identified via principal component analysis also indicated an excess of small p values (online Supplementary Material, Figure S1).

Third, despite the relatively large sample size compared to previous gene expression studies of migraine, our study did not identify individual genes reaching study-wide significance. This indicates that, while our study size is able to demonstrate global differential expression and implicate pathways, it lacked power to implicate individual genes associated with migraine. Indeed, assuming a standard deviation of 0.7 for the gene intensity measures on a \log_2 scale, using a two-sided test based on the conservative Bonferroni-adjusted p value threshold $p < 2.8 \times 10^{-6}$, we would

require 438 migraine cases and 438 non-migraine controls to have 80% power to detect a 1.2-fold difference in gene expression levels.

Finally, we could not evaluate the impact of potential (unobserved) confounding effects on gene expression, such as medication use and cell type proportions. Studies of large, well phenotyped samples aimed at identifying genes differentially expressed between migraine cases and non-migraine controls will overcome such limitations. We also advocate an approach

where gene expression data is integrated with other – omic data such as SNP genotype, DNA methylation, and metabolomic data measured in whole blood; this can then be compared with results from other tissues, such as the recent gene expression study in periosteum tissue published in the *Annals of Neurology* (11). The use of such diverse molecular data will improve our ability to dissect novel pathogenic pathways underlying migraine and ultimately help target new preventive therapies.

Article highlights

- Whole blood is an accessible and reliable biological substrate for genomic studies of brain-related disorders and may therefore be used to interrogate biological pathways underlying migraine.
- We assessed 17994 gene expression probes measured from the whole blood of 83 migraine cases and 83 age and gender-matched non-migraine controls with European ancestry.
- Association analyses between migraine and probe expression levels, adjusted for age and gender, revealed an excess of small p values.
- Network-based analyses identified 10 gene co-expression modules, one of which contained a higher than expected number of genes differentially expressed between migraine cases and controls after correction for correlated gene expression, and was significantly enriched with immune function and inflammatory signaling pathways.
- Blood gene expression studies of migraine can improve our understanding of biological mechanisms underlying migraine and may thereby inform the development of diagnostic biomarkers and new targets for drug therapy.

Acknowledgements

We acknowledge the participation of the individuals sampled in the Brisbane Systems Genetics Study. Dale R Nyholt (613674) and Grant W Montgomery (619667, 1078399) were supported by an Australian National Health and Medical Research Council (NHMRC) Research Fellowship. Joseph E Powell is supported by an NHMRC Career Development Fellowship (APP1107599). Research reported in this publication was supported by NHMRC project grants (APP1046880, APP1075175, APP1083405) and by the European Union's Seventh Framework programme (2007–2013) under grant agreement no. 602633 (EUROHEADPAIN).

Ethics approval

Individuals present in this study were recruited as part of the Brisbane Systems Genetics Study (BSGS). This study was approved by the QIMR Berghofer Medical Research Institute Human Research Ethics Committee (HREC) Approval number: P589 and the QUT HREC (Administrative Review approval number 1500000115 (as per QIMR Berghofer Medical Research Institute, Approval number: P589). All participants gave informed written consent.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia* 2013; 33: 629–808.
2. Vos T, Barber RM, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; 386: 743–800.
3. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2163–2196.
4. Mulder EJ, Van Baal C, Gaist D, et al. Genetic and environmental influences on migraine: A twin study across six countries. *Twin Res* 2003; 6: 422–431.
5. Ulrich V, Gervil M, Kyvik KO, et al. Evidence of a genetic factor in migraine with aura: A population-based Danish twin study. *Ann Neurol* 1999; 45: 242–246.
6. Gervil M, Ulrich V, Kaprio J, et al. The relative role of genetic and environmental factors in migraine without aura. *Neurology* 1999; 5: 995–999.

7. Russell MB and Olesen J. Increased familial risk and evidence of genetic factor in migraine. *Br Med J* 1995; 311: 541–544.
8. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* 2016; 48: 856–866.
9. Mostafavi S, Battle A, Zhu X, et al. Type I interferon signaling genes in recurrent major depression: Increased expression detected by whole-blood RNA sequencing. *Mol Psychiatry* 2014; 19: 1267–1274.
10. Jansen R, Penninx BWJH, Madar V, et al. Gene expression in major depressive disorder. *Mol Psychiatry* 2015; 21: 339–347.
11. Perry CJ, Blake P, Buettner C, et al. Upregulation of inflammatory gene transcripts in periosteum of chronic migraineurs: Implications for extracranial origin of headache. *Ann Neurol* 2016; 79: 1000–1013.
12. Gerring Z, Rodriguez-Acevedo AJ, Powell JE, et al. Blood gene expression studies in migraine: Potential and caveats. *Cephalgia* 2016; 36: 669–678.
13. Tanriverdi K and Freedman JE. Blood and cardiovascular disease: The promise and limitations of gene expression analysis. *Circ Cardiovasc Genet* 2008; 1: 7–9.
14. Hershey AD, Tang Y, Powers SW, et al. Genomic abnormalities in patients with migraine and chronic migraine: Preliminary blood gene expression suggests platelet abnormalities. *Headache* 2004; 44: 994–1004.
15. Hershey A, Horn P, Kabbouche M, et al. Genomic expression patterns in menstrual-related migraine in adolescents. *Headache* 2012; 52: 68–79.
16. Hershey AD, Burdine D, Kabbouche MA, et al. Genomic expression patterns in medication overuse headaches. *Cephalgia* 2011; 31: 161–171.
17. Powell JE, Henders AK, McRae AF, et al. Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res* 2012; 22: 456–466.
18. Powell JE, Henders AK, McRae AF, et al. Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet* 2013; 9: e1003502.
19. Smyth GK and Speed T. Normalization of cDNA microarray data. *Methods* 2003; 31: 265–273.
20. Goldinger A, Henders AK, McRae AF, et al. Genetic and nongenetic variation revealed for the principal components of human gene expression. *Genetics* 2013; 195: 1117–1128.
21. Lipton RB. A self-administered screener for migraine in primary care: The ID migraine™ validation study. *Neurology* 2003; 61: 375–382.
22. Cousins G, Hijazze S, Van De Laar FA, et al. Diagnostic accuracy of the ID migraine: A systematic review and meta-analysis. *Headache* 2011; 51: 1140–1148.
23. Kerr KF. Comments on the analysis of unbalanced microarray data. *Bioinformatics* 2009; 25: 2035–2041.
24. Rothman K, Greenland S and Lash T. *Modern epidemiology*, 3rd edn. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008, pp.134–135.
25. Li J and Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; 95: 221–227.
26. Joshi-Tope G, Gillespie M, Vastrik I, et al. Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res* 2005; 33: D428–432.
27. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008; 36: D480–484.
28. Kong SW, Pu WT and Park PJ. A multivariate approach for integrating genome-wide expression data and biological knowledge. *Bioinformatics* 2006; 22: 2373–2380.
29. Schaefer CF, Anthony K, Krupa S, et al. PID: The Pathway Interaction Database. *Nucleic Acids Res* 2009; 37: D674–679.
30. Goudriaan A, De Leeuw C, Ripke S, et al. Specific glial functions contribute to schizophrenia susceptibility. *Schizophr Bull* 2014; 40: 925–935.
31. Ruano D, Abecasis GR, Glaser B, et al. Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am J Hum Genet* 2010; 86: 113–125.
32. Lips ES, Cornelisse LN, Toonen RF, et al. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol Psychiatry* 2012; 17: 996–1006.
33. Shoemaker JE, Lopes TJS, Ghosh S, et al. CTen: A web-based platform for identifying enriched cell types from heterogeneous microarray data. *BMC Genomics* 2012; 13: 1–11.
34. Storey JD and Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; 100: 9440–9445.
35. Langfelder P and Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 1–13.
36. Langfelder P, Zhang B and Horvath S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* 2008; 24: 719–720.
37. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available from: <http://www.r-project.org/> (accessed 4 April 2016).
38. van Eijk KR, de Jong S, Stregman E, et al. Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood. *Eur J Hum Genet* 2015; 23: 1106–1110.
39. Melé M, Ferreira PG, Reverter F, et al. The human transcriptome across tissues and individuals. *Science* 2015; 348: 660–665.
40. Rollins B, Martin MV, Morgan L, et al. Analysis of whole genome biomarker expression in blood and brain. *Am J Med Genet Part B Neuropsychiatr Genet* 2010; 153: 919–936.
41. Spielman RS and Cheung VG. Reply to “On the design and analysis of gene expression studies in human populations”. *Nat Genet* 2007; 39: 808–809.
42. Allison DB, Cui X, Page GP, et al. Microarray data analysis: From disarray to consolidation and consensus. *Nat Rev Genet* 2006; 7: 55–65.

43. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* (80-) 2015; 348: 648–660.
44. McKenzie M, Henders AK, Caracella A, et al. Overlap of expression Quantitative Trait Loci (eQTL) in human brain and blood. *BMC Med Genomics* 2014; 7: 1–11.
45. Breen MS, Maihofer AX, Glatt SJ, et al. Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry* 2015; 20: 1538–1545.
46. Chen C, Cheng L, Grennan K, et al. Two gene co-expression modules differentiate psychotics and controls. *Mol Psychiatry* 2013; 18: 1308–1314.
47. Goeman JJ and Bühlmann P. Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics* 2007; 23: 980–987.
48. Barry WT, Nobel AB and Wright FA. Significance analysis of functional categories in gene expression studies: A structured permutation approach. *Bioinformatics* 2005; 21: 1943–1949.
49. Tfelt-Hansen P. Triptans vs other drugs for acute migraine. Are there differences in efficacy? A comment. *Headache* 2008; 48: 601–605.
50. Pietrobon D and Striessnig J. Neurobiology of migraine. *Nat Rev Neurosci* 2003; 4: 386–398.
51. Dodick DW, Goadsby PJ, Spierings ELH, et al. Safety and efficacy of LY2951742, a monoclonal antibody to calcitonin gene-related peptide, for the prevention of migraine: A phase 2, randomised, double-blind, placebo-controlled study. *Lancet Neurol* 2016; 13: 885–892.
52. Zarifoğlu M, Karli N and Taşkapılıoğlu Ö. Can ID Migraine™ be used as a screening test for adolescent migraine? *Cephalalgia* 2008; 28: 65–71.
53. Oztora S, Korkmaz O, Dagdeviren N, et al. Migraine headaches among university students using id migraine test as a screening tool. *BMC Neurol* 2011; 11: 1–5.
54. Stewart WF, Linet MS, Celentano DD, et al. Age- and sex-specific incidence rates of migraine with and without visual aura. *Am J Epidemiol* 1991; 134: 1111–1120.