Immune signatures and disorder-specific patterns in a cross-disorder gene expression analysis


Background
Recent studies point to overlap between neuropsychiatric disorders in symptomatology and genetic aetiology.

Aims
To systematically investigate genomics overlap between childhood and adult attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD) and major depressive disorder (MDD).

Method
Analysis of whole-genome blood gene expression and genetic risk scores of 318 individuals. Participants included individuals affected with adult ADHD (n = 93), childhood ADHD (n = 177), MDD (n = 63), ASD (n = 51), childhood dual diagnosis of ADHD–ASD (n = 16) and healthy controls (n = 78).

Results
Weighted gene co-expression analysis results reveal disorder-specific signatures for childhood ADHD and MDD, with moderate overlap between MDD and schizophrenia, bipolar disorder and ADHD and low, but significant, overlap between schizophrenia and ASD. These results suggest that some shared molecular genetic factors underlie a significant proportion of the risk for development of several psychiatric disorders. To date, however, it is not known whether these genomic overlaps are reflected in overlapping patterns of gene expression in tissues. In this study we therefore explored whole blood gene expression across adult and childhood ADHD, ASD, MDD and healthy controls, using weighted gene co-expression analysis to search for patterns of correlated gene expression between disorders as well as disorder-specific gene expression signatures. In addition, we generate polygenic risk scores to assess overlap between disorders on a genome-wide genetic-risk level. We next investigate whether differences in polygenic risk scores are reflected in disorder-related gene expression profiles and whether they can be used to tease apart genetic and environmental influences on gene expression.

Method
Participants
Participants were from four different projects and included individuals affected with adult ADHD, childhood ADHD, MDD, ASD, childhood dual diagnosis ADHD–ASD and healthy controls (Table 1). Because of the age difference and its possible confounding effect on gene expression within the ADHD diagnostic group we decided to split the ADHD samples into a childhood (mean age 10 years) and adult (mean age 32 years) ADHD groups. Participant characteristics and inclusion/exclusion
criteria used by each project are included in online supplement DS1. Briefly, we included: (a) 63 people with MDD and 57 controls from the Depression Case-Control (DeCC) study, a large case–control study that recruited unrelated patients from three clinical sites in the UK; (b) 93 adults with ADHD attending a National adult ADHD out-patient clinic in London, UK; (c) 16 individuals with ASD–ADHD, 7 with ASD, 17 with childhood ADHD and 7 controls from the Biomarkers for Childhood Neuropsychiatric Disorders (BioNed) project; and (d) 44 people with ASD and 14 controls from the Autism Interventions (AIMS) project. The following phenotypic information was available within each project and subsequently used for the current cross-disorder analyses: age, gender, diagnosis, date of collection, ethnicity, psychoactive medication use and, for all projects except DeCC, comorbidity of other psychiatric disorders (another psychiatric diagnosis was an exclusion criterion in the DeCC study). Each project has ethical approval and full informed consent for each participant (details in online supplement DS1). Whole blood samples were collected using PAXGene tubes for RNA and EDTA for DNA. All RNA samples were processed within one batch to generate whole-genome gene expression data.

Gene expression data preprocessing

Whole-genome gene expression data of a total of 424 individuals were generated using the Illumina HT-12.v4 BeadChips at the SGD/P/BRC BioResource Illumina core lab according to the manufacturer’s protocol. We rigorously quality controlled and preprocessed the data using a standard pipeline (https://github.com/snewhouse/BRC_MH_Bioinformatics), excluding sample and probe outliers, after which robust spline normalisation and log² transformation were applied.12 After excluding samples based on low-quality expression profiles we included only participants with full phenotype data (age, gender, diagnosis, date of collection, ethnicity, psychoactive medication use per individual, and RNA integrity number (RIN) and RNA concentration per sample). This left 318 participants and 5638 probes for analysis (Table 1). To minimise project collection and sample handling batch effects, the data were corrected for three unknown variables (Table 1). To minimise project collection and sample handling batch effects, the data were corrected for three unknown variables (Table 1). To minimise project collection and sample handling batch effects, the data were corrected for three unknown variables (Table 1). To minimise project collection and sample handling batch effects, the data were corrected for three unknown variables (Table 1). To minimise project collection and sample handling batch effects, the data were corrected for three unknown variables (Table 1).

Corrected gene expression data were analysed using weighted gene co-expression analysis (WGCNA).15,16 We constructed a signed weighted co-expression network based on the matrix of pairwise Pearson correlation coefficients, which were raised to a fixed power (β = 12) by the criteria described by Zhang & Horvath.16 Soft-thresholding results in a 5638 × 5638 dimensional weighted adjacency matrix containing pairwise connection strengths. Subsequently, a topological overlap measure is calculated based on the number of shared neighbours. Modules were then defined as branches of a hierarchical clustering tree using a dissimilarity measure (1–topological overlap). Each module is subsequently assigned a colour. To define a representative module expression profile for each module, we summarised the (standardised) gene expression profiles of the module by their first principal component. This statistic is referred to as the module eigengene: it can be thought of as an average gene expression value for all probes in a module per sample. In order to identify hub genes, we calculated a connectivity measure (‘module membership’) per probe by correlating expression values with the relevant module eigengene. Probes with high module membership are defined as ‘hub genes’ of their module.

Module eigengenes and phenotypes of interest

The module eigengenes were correlated to phenotype of interest using a linear model.17 Participants were assigned a main diagnosis of a particular disorder (controls, ASD, MDD, adult ADHD, childhood ADHD, ADHD–ASD) v. all other participants. We investigated the effect of possible covariates on the association of module eigengenes with our phenotypes of interest. This led us to correct for gender, RIN and RNA concentration. Medication use defined as dichotomous measures of use of antidepressants (n = 73), stimulants (n = 49), antipsychotics (n = 7), benzodiazepines (n = 5) and mood stabilisers (n = 5) were investigated as a possible covariate. Considering the nature of the sample collection of the childhood ADHD–ASD and childhood ADHD samples, analyses were repeated also including age as covariate, allowing us to tease out diagnosis and age-specific effects. We used a Bonferroni threshold for significance (the number of tests was calculated as the number of modules × number of diagnoses × 2).

Characterisation of modules of interest

Modules of interest were tested for enrichment of blood cell type lists using the userListEnrichment function in WGCNA package with five as minimum number of genes in a pathway. This function compares the number of overlapping genes to the maximal possible overlap and applies Bonferroni correction. Enrichment analyses were performed in WebGestalt using Wikipathways database.18 In addition, we performed connectivity mapping on selected modules to investigate overlap with drug-induced gene expression changes through the LINCS/CMap database (http://apps.lincscloud.org).19 We entered the gene content of modules of interest as upregulated, resulting in a list of compounds of which the application to cell lines results in a similar gene expression pattern.

Table 1 Sample characteristics

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n</th>
<th>% male</th>
<th>Age (s.d.)</th>
<th>RIN (s.d.)</th>
<th>% medication-free</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>78</td>
<td>59</td>
<td>46 (15)</td>
<td>8.3 (0.4)</td>
<td>97</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>51</td>
<td>90</td>
<td>29 (11)</td>
<td>8.5 (0.3)</td>
<td>61</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>63</td>
<td>33</td>
<td>47 (9)</td>
<td>8.5 (0.4)</td>
<td>29</td>
</tr>
<tr>
<td>Adult attention-deficit hyperactivity disorder</td>
<td>93</td>
<td>77</td>
<td>32 (12)</td>
<td>8.5 (0.5)</td>
<td>63</td>
</tr>
<tr>
<td>Childhood attention-deficit hyperactivity disorder</td>
<td>17</td>
<td>100</td>
<td>10 (2)</td>
<td>8.6 (0.4)</td>
<td>29</td>
</tr>
<tr>
<td>Autism spectrum disorder–attention-deficit hyperactivity disorder</td>
<td>16</td>
<td>100</td>
<td>11 (2)</td>
<td>8.7 (0.3)</td>
<td>69</td>
</tr>
</tbody>
</table>

RIN, RNA integrity number.
Genotype information and previous genome-wide association study (GWAS) findings
For most individuals genotype data were available and after quality control and imputation polymorphic risk score analysis and genetic pathway analysis was carried out to investigate the relationship of genome-wide risk to disorder-specific gene expression findings in this study (see online supplement, Supplements DS2–4 and Figs DS1 and DS2).

Results
The WGCNA on 318 participants and 5638 probes resulted in seven gene co-expression modules, ranging from 2077 probes in the turquoise module to 80 probes in the red module. The grey module contains 675 probes not belonging to any other module, representing background noise. The network dendrogram is given in online Fig. DS3 and all probes and corresponding module eigengenes = 42 tests per model, therefore 84 tests in total.

The green and red immune modules are inversely correlated to psychiatric disorders
The green (186 probes) and red module (80 probes) eigenvalues estimates per individual correlated negatively with MDD and positively with adult ADHD status (Table 2, Fig. 1). Even though the MDD sample is female-dominated and the adult ADHD sample male-dominated, gender was included as a covariate and therefore does not fully account for this effect. Results also remain significant after additional correction for age. Of the 93 individuals with ADHD, 7 had comorbid MDD. However, these individuals still resembled the pattern of the adult ADHD group and Figs DS1 and DS2).

Turquoise module represents overlapping signatures between psychiatric disorders
The turquoise module (2077 probes) is positively correlated with both MDD and childhood ADHD status (Table 2). This effect does not remain significant for childhood ADHD after correction for age. Enrichment analyses reveal significant categories to be cytoplasmic ribosomal proteins (69 genes, adjusted

<table>
<thead>
<tr>
<th>Module</th>
<th>Probes, n</th>
<th>Controls</th>
<th>Major depressive disorder</th>
<th>Adult attention-deficit hyperactivity disorder</th>
<th>Childhood attention-deficit hyperactivity disorder</th>
<th>Autism spectrum disorder</th>
<th>Autism spectrum disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turquoise</td>
<td>2077</td>
<td>ns</td>
<td>4.4 (1.4 \times 10^{-7}) b</td>
<td>ns</td>
<td>4.0 (5.8 \times 10^{-7}) b</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Red</td>
<td>80</td>
<td>ns</td>
<td>-4.3 (1.5 \times 10^{-7}) a</td>
<td>4.3 (2.2 \times 10^{-7}) a</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Green</td>
<td>186</td>
<td>ns</td>
<td>-5.8 (8.8 \times 10^{-8})  a</td>
<td>4.8 (1.6 \times 10^{-7}) b</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Blue</td>
<td>1672</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>-4.0 (6.7 \times 10^{-7}) c</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

ns, not significant.

a. Module eigengenes were tested for association with phenotypes of interest including gender, RNA integrity number (RIN) and RNA concentration as covariates.
b. Positive associations.
c. Negative associations.

204
https://doi.org/10.1192/bjp.bp.115.175471 Published online by Cambridge University Press
Cross-disorder gene expression analysis

connectivity is $r = -0.48$, $P < 2.2 \times 10^{-16}$ for childhood ADHD and $r = -0.48$, $P < 2.2 \times 10^{-16}$ for MDD. Major hub genes in this module are ZRANB2, TMEM126B, RPL15, PCNP and SRP9. The genes with highest MDD gene significance are LOC1001291, ANXA1, RNF13, GMFG and H3F3B. For childhood ADHD these are SUMO3, RPL22, UTP3, POLE3 and CCDC50.

Disorder-specific signature of the blue module

The blue module (1672) is negatively correlated to childhood ADHD status (Table 2). This effect does not remain significant after correction for age. Enrichment analyses reveal the following significant Wikipathways: integrin-mediated cell adhesion (26 genes, adjusted $P = 8 \times 10^{-4}$), focal adhesion (35 genes, adjusted $P = 1 \times 10^{-3}$), prostate cancer (21 genes, adjusted $P = 6 \times 10^{-7}$) and IL17 signalling pathway (13 genes, adjusted $P = 0.04$). The correlation between childhood ADHD gene significance and connectivity is $r = -0.5$, $P < 2.2 \times 10^{-16}$. Blue hub genes are WAS, MOBKL2A, C15orf39, G6PD and GNAI2. Genes with highest negative correlation to childhood ADHD disease status are TMUB2, STAT5B, DENND3, CA4, FCGR3B.

Relationship gene expression signatures and previous GWAS findings

We did not find enrichment of GWAS signal or association of the module eigengenes with polygenic risk scores for the modules of interest (online supplements DS2–4 and Figs DS1 and DS2).

Discussion

Main findings

This study of whole-blood gene expression of several psychiatric disorders aimed to reveal overlapping gene expression patterns and disorder-specific signatures. The WGCNA on 318 participants revealed seven gene co-expression modules. Of these, two small modules are inversely related to MDD and adult ADHD, the large turquoise module is associated with both MDD and childhood ADHD and finally, the blue module shows a disorder-specific signature for childhood ADHD. No significant results were found for the ASD, ADHD–ASD and control groups. Even though there is some evidence for increased immune-related comorbidities in ASD,20,21 we do not find an immune gene expression signature for these groups, as we do for MDD and adult ADHD.

Interpretation and comparison with findings from other studies

Two small modules, red and green, are negatively correlated to MDD status, but positively correlated to adult ADHD status. GWAS on both disorders have not yielded genome-wide significant hits to date.22,23 Previous literature about genetic overlap between MDD and ADHD has yielded conflicting results, reporting a genetic correlation between MDD and ADHD,11 but no significant overlap in polygenic risk scores.10 A reason for these
discrepancies could be the lack of distinction between adult ADHD and childhood ADHD in previous studies, which could add to phenotypic heterogeneity. The group with childhood ADHD may include individuals whose condition will become less severe with age, whereas the adult group with ADHD contains individuals whose condition is chronic. Another reason why we chose to analyse participants with childhood and adult ADHD in separate groups is because of anticipated biological heterogeneity between children and adults in the context of peripheral gene expression. An explanation for our findings would be gender differences in prevalence of MDD and adult ADHD, however, analyses were corrected for gender and stratified analyses yielded the same results (data not shown). We also performed a correlation with the module eigengenes and an indicator of current state of depression (Beck Depression Inventory24) within participants, which did not show a significant association (data not shown).

The green module shows enrichment for a cell signalling category and harbours some interesting highly connected hub genes that also show high gene significance, most notably, YY1, WDRL2 and AAK1. YY1 is a transcription factor involved in many processes including transforming growth factor beta (TGF-β) signalling, but has also shown to be active in histone modification.25 In addition WDRL2 (WD repeat domain 82) is part of SET1A/SET1B histone H3K4 methyltransferase complexes,26 which genetic pathway analyses has shown to be strongly enriched for association in MDD, bipolar disorder and schizophrenia.27 This implication of epigenetic processes provides a mechanism by which environmental influences can exert their influence on gene expression and thereby contribute to the pathology of neuropsychiatric disorders.28,29 The green hub gene adaptor-associated kinase 1 (AAK1) is of interest because it is a positive regulator of the Notch pathway.30 This pathway is traditionally implicated in cell-fate determination during development, but has important function in tissue homeostasis and neuronal plasticity later in life. In addition, it has been implicated to play a role in immune functioning.31,32

Even though the immune system has not been implicated in adult ADHD before, there are many studies on its association with MDD although the relationship is still somewhat controversial.34 Expression differences in cytokines have been shown to differentiate patients with MDD from ones with bipolar disorder and controls35 and studies have shown regulation of the serotonin receptor through cytokines and neurotrophins.36,37 Importantly, cytokines and polymorphisms in interleukin genes have been shown to predict antidepressant treatment response.38,39 The majority of patients with MDD (70%) and a few (13%) of the participants with adult ADHD in the current study were on antidepressant medication although medication use did not correlate significantly with any gene expression module. However, connectivity mapping revealed that the upregulation of genes in red and green modules as seen in our adult ADHD group coincides with those seen in response to application of a number of tricyclic antidepressants, indicated for the treatment of depression and ADHD with comorbid depression. Results also contained some anti-inflammatory drugs, converging with the module enrichments for immune system genes. This could support the hypothesis that anti-inflammatory drugs such as non-steroidal anti-inflammatory drugs (NSAIDs) might have a role in the treatment of MDD and our data suggests, also ADHD. Literature on the effects of NSAIDs in patients with MDD is, however, mixed.40,41 Our results suggest potential drug repositioning opportunities for NSAIDs for both MDD and ADHD.

Whereas MDD and adult ADHD showed opposite patterns of gene expression changes (the green and red modules), we found an overlapping gene expression signature between MDD and childhood ADHD, in the form of the turquoise module. One possible interpretation of this finding is that childhood ADHD increases risk for developing MDD in later life.42 In order to fully untangle this, we would need childhood diagnoses of ADHD in the MDD cohort, but this was not available.

Finally, we find a disorder-specific signature of childhood ADHD in the blue module, which is highly enriched for cell signalling genes. There are two hub genes located on the X-chromosome; G6PD and WAS. Even though we corrected for gender in our analyses, the childhood ADHD group consisted solely of males, which might explain downregulation of X-chromosome genes, although the blue module is not enriched for genes on sex chromosomes. The significance of the association of childhood ADHD with the turquoise and blue modules disappears when adding age as a covariate in the analyses, but in the absence of a healthy child control group, age and childhood ADHD cohort membership are highly correlated. However, the fact that the childhood dual diagnosis ADHD–ASD group does not show a correlation with these modules suggests that age cannot fully explain these results.

We did not find enrichment of GWAS signal or association of the module eigengenes with polygenic risk scores for the modules of interest (online supplements DS2–4 and Figs DS1 and DS2). This could indicate that the differences in gene expression are driven by environmental rather than genetic factors or, perhaps, that polygenic scores are not yet strong enough in disorders such as MDD and ADHD. One possible known environmental influence on gene expression is smoking. Even though we did not have access to smoking behaviour for all participants, a Fisher’s exact test of enrichment of smoking-related genes43 did not reveal a significant enrichment for the modules of interest (data not shown). Likewise, a sample-handling or collection effect is unlikely because of initial batch correction and the fact that the healthy controls were for the most part from the same project as the participants with MDD and yet do not show an effect for the relevant modules. In addition, the lack of genetic association could also be the result of the initial GWAS results being underpowered to detect variants associated with MDD, ADHD and ASD.

**Directions for further study**

Future research could extend this study to include a broader range of psychiatric disorders, such as psychotic disorders (schizophrenia) and anxiety disorders (obsessive–compulsive disorder, generalised anxiety disorder), in order to better understand the genomic correlates of different syndromes. However, our results have several limitations including power to detect effects in relatively small samples and the reliance upon cross-sectional study designs. Our findings in adult ADHD and MDD will require replication and assessment in different study designs to assess potential therapeutic applications. Also, in this study we examined gene expression in blood but it will be important to determine whether the pattern of results holds true for brain tissue. In ASD, for instance, changes in the expression of a number of genes has been reported to be altered in post-mortem studies.44

In conclusion, in a study of gene expression in peripheral blood of patients with psychiatric disorders and healthy controls, we identified both cross-disorder and disorder-specific signatures for adult ADHD and MDD. With the caveats discussed above, they suggest new pathways contributing to distinct pathophysiology in psychiatric disorders and shed light on potential shared genomic risk factors.
We thank the participants and staff at the various recruitment centres.

C.E.W. is funded through European Union’s Seventh Framework Programme via Marie Skłodowska-Curie Action, co-financed by the Junta de Andalucía and the European Commission under Talents Postdoctoral agreement number 262726.

First received 22 May 2015, final revision 10 Sep 2015, accepted 27 Oct 2015

Acknowledgements

We thank the participants and staff at the various recruitment centres.

References


