Different gene sets contribute to different symptom dimensions of depression and anxiety
van Veen, Tineke; Goeman, Jelle J.; Monajemi, Ramin; Wardenaar, Klaas J.; Hartman, Catharina A.; Snieder, Harold; Nolte, Ilja M.; Penninx, Brenda W. J. H.; Zitman, Frans G.

Published in:
American Journal of Medical Genetics. Part B: Neuropsychiatric Genetics

DOI:
10.1002/ajmg.b.32058

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2012

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Different Gene Sets Contribute to Different Symptom Dimensions of Depression and Anxiety

Tineke van Veen,1* Jelle J. Goeman,2 Ramin Monajemi,2 Klaas J. Wardenaar,1 Catharina A. Hartman,4 Harold Snieder,3 Ilja M. Nolte,3 Brenda W.J.H. Penninx,1,4,5 and Frans G. Zitman1

1Department of Psychiatry, Leiden University Medical Centre, Leiden, The Netherlands
2Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, Leiden, The Netherlands
3Unit of Genetic Epidemiology & Bioinformatics, Department of Epidemiology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
4Department of Psychiatry, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
5Department of Psychiatry, VU University Medical Centre, Amsterdam, The Netherlands

Manuscript Received: 12 January 2012; Manuscript Accepted: 19 April 2012

Although many genetic association studies have been carried out, it remains unclear which genes contribute to depression. This may be due to heterogeneity of the DSM-IV category of depression. Specific symptom-dimensions provide a more homogenous phenotype. Furthermore, as effects of individual genes are small, analysis of genetic data at the pathway-level provides more power to detect associations and yield valuable biological insight. In 1,398 individuals with a Major Depressive Disorder, the symptom dimensions of the tripartite model of anxiety and depression, General Distress, Anhedonic Depression, and Anxious Arousal, were measured with the Mood and Anxiety Symptoms Questionnaire (30-item Dutch adaptation; MASQ-D30). Association of these symptom dimensions with candidate gene sets and gene sets from two public pathway databases was tested using the Global test. One pathway was associated with General Distress, and concerned molecules expressed in the endoplasmatic reticulum lumen. Seven pathways were associated with Anhedonic Depression. Important themes were neurodevelopment, neurodegeneration, and cytoskeleton. Furthermore, three gene sets associated with Anxious Arousal regarded development, morphology, and genetic recombination. The individual pathways explained up to 1.7% of the variance. These data demonstrate mechanisms that influence the specific dimensions. Moreover, they show the value of using dimensional phenotypes on one hand and gene sets on the other hand.

Key words: GWAS; pathway; tripartite model; MDD

INTRODUCTION

Major depressive disorder (MDD) is highly prevalent, with a lifetime prevalence of 15% [Weissman et al., 1993]. Patients with MDD vary widely with respect to their symptom profiles. If this is not taken into account, symptom-specific risk factors will remain undetected. Further, extensive comorbidity between depressive and anxiety disorders exists, suggesting that part of their pathogenesis is shared [Mineka et al., 1998; Vollebergh et al., 2001; Kendler et al., 2003]. These problems can be overcome by addressing the etiology of underlying, homogeneous symptom dimensions. The tripartite model of anxiety and depression is a well-validated dimensional model [Clark and Watson, 1991]. The dimension of General Distress (i.e., negative affect) contains symptoms common to both depression and anxiety, including irritability, hopelessness and guilt. The dimension of Anhedonic Depression (i.e., lack of positive affect) refers to lack of energy and enthusiasm, and is most specific to depression. The dimension of Anxious Arousal (i.e., somatic arousal) is most specific to anxiety, and consists of symptoms of somatic tension and hyperarousal.

How to Cite this Article:
The coverage of these dimensions is quite specific and likely influenced by fewer genes than the broad phenotype of MDD. Furthermore, the dimensions are continuous and therefore provide more statistical power to detect small genetic effects [MacCallum et al., 2002]. A recent simulation study indeed demonstrated that much power is gained when the multidimensionality of the phenotype is taken into account, such that genetic effects specific to certain phenotype dimensions can be detected [van der Sluis et al., 2010]. Thus, it is not very surprising that genome-wide association studies (GWAS) for MDD yielded no significant associations [Sullivan et al., 2009; Muglia et al., 2010; Wray et al., 2010; Shi et al., 2011; Shyn et al., 2011]. Further, studies of candidate genes for MDD have not revealed major disease loci [Bosker et al., 2011]. Neuroticism is a personality trait closely associated with MDD, largely due to shared genetic risk factors [Hettema et al., 2006]. Although neuroticism provides a less complex phenotype (strongly related to General Distress) [Clark et al., 1994], the three GWAS for neuroticism yielded no significant findings [Shifman et al., 2008; van den Oord et al., 2008; Calboli et al., 2010].

In complex diseases, each gene has a small effect, but unfavorable combinations of genes in a pathway may affect the output of the pathway. This may affect downstream phenotypes. Thus, analyses should be carried out at the pathway-level [Comings, 1998]. Whereas gene set analyses have been widely used in the field of microarrays, analyses of predefined gene sets in genetic association studies are very recent [Hooper-van Veen et al., 2006; Wang et al., 2007; Holmans et al., 2009; Ruano et al., 2010]. Replication of genetic associations with phenotypes was shown to be easier at the pathway level than at the gene or SNP levels [Luo et al., 2010]. Different types of pathway analyses have been used [for review, see Holmans, 2010]. The Global test is one example, based on a regression model that includes all SNPs of the pathway. It tests whether on average these are associated with the phenotype [Goeman et al., 2004]. Interpretation is intuitive as the analysis of a pathway is a direct extension of the analysis of single genes or single SNPs. Further, the Global test performed consistently well in the analysis of multiple SNPs in candidate genes [Chapman and Whittaker, 2008; Pan, 2009] and candidate pathways [Deelen et al., 2011].

In order to unravel the heterogeneity of symptoms among patients with MDD, we aimed to identify pathways associated with specific dimensions of depression and anxiety. Among individuals with a lifetime MDD, we used the Global test to identify gene sets associated with the symptom dimensions of the tripartite model.

**METHODS**

**Subjects**

Data from 1,598 MDD cases from NESDA that passed quality control in the GAIN-MDD study were used [Sullivan et al., 2009]. NESDA is a cohort study that follows patients with depressive and/or anxiety disorders recruited between September 2004 and February 2007 from mental health care organizations, primary care, and community samples [Penninx et al., 2008]. Inclusion criteria for genotyping were a lifetime diagnosis of MDD according to DSM-IV, assessed with the composite international diagnostic interview (World Health Organisation version 2.1), age 18–65 years, and self-reported western European ancestry. Individuals not fluent in Dutch or with a primary diagnosis of a psychotic disorder, obsessive compulsive disorder, bipolar disorder, alcohol or substance use disorder were excluded. For details see Boomsma et al. [2008]. The Ethical Review Boards of all participating universities approved the research protocol. All subjects provided written informed consent. For 200 of the 1,598 cases, no dimensional phenotype data were available (questionnaires not returned). These respondents were younger, more often men, and more often with current MDD. The remaining 1,398 individuals were the basis of the present study.

**Genotype Data**

Genotype data were collected as part of the GAIN-MDD study. Genotyping was carried out by Perlegen Sciences according to strict standard operating procedures. High-density oligonucleotide arrays were used yielding 599,164 SNPs. Eight SNPs with duplicate numbers were deleted and 73 mitochondrial SNPs were removed for later analysis. From the remaining 599,083 SNPs on the Perlegen chip 435,291 passed quality control. For details see Boomsma et al. [2008], Sullivan et al. [2009], Bosker et al. [2011].

**Annotation**

SNPs were assigned to genes using RefSeq alignments to the genome (UCSC, May 2010) and gene to RefSeq annotation (NCBI, build 35/hg17) [Wheeler et al., 2007]. We included SNPs between 5 kb from 5' end and 2 kb from the 3' end of genes, as did others [Luciano et al., 2011]. SNPs mapped to multiple genes were assigned to each of these genes. The resulting gene map contained 18,601 well-oriented Entrez Gene RefSeq clusters that defined the gene boundaries, and covered 170,795 SNPs. With inclusion of 5 kb at 5' and 2 kb at 3' end, 188,017 SNPs were included. For men, SNPs on the X-chromosome were encoded 0 or 2 [Clayton, 2008].

**Gene Sets**

Manually curated pathway databases like Biocarta and Kegg [Ogata et al., 1999] contain high-quality information, whereas the Gene
Ontology (GO) database has extensive coverage by electronic annotation [Ashburner et al., 2000]. Genes assigned the same GO term have some aspect of their biology in common. In a comparison of several pathway databases, results were complementary [Elbers et al., 2009]. The combined use was advised to ensure both comprehensive coverage and use of the available high quality information [Wang et al., 2010]. In the present study, five groups of gene sets were analyzed: (1) 12 candidate pathways manually curated from published data (see Supplementary Online), (2) 117 GO cellular components, (3) 211 GO molecular functions, (4) 609 GO biological processes, (5) 880 canonical pathways from the Molecular Signatures Database v3.0 [Subramanian et al., 2007], which consists of 217 pathways from BioCarta, 186 from KEGG [Ogata et al., 1999], and 430 from Reactome [Croft et al., 2011].

Selection of GO Terms
GO terms were associated with genes using all evidence codes and with the genes of its offspring. For this, the Entrez Gene to GO association map org.Hs.eG02ALLEGS from the R package org.Hs.eG.db (v2.4.1) was used. This database contained 12,327 GO terms; to restrict multiple testing, only those with 20–100 genes were included. From this subset, only the roots of sub-trees were selected to avoid overlap between gene sets. A similar approach has been used by Wang et al. [2007], who selected level 4 GO terms among those with 20–200 genes.

Global Test
To test for association of gene sets with General Distress, Anhedonic Depression and Anxious Arousal, the R package GlobalTest (v5.1.5) was used [Goeman et al., 2004]. The genotypes for all SNPs in a pathway were included in a linear regression model, and the tested null hypothesis was that no SNP in the pathway was associated with the symptom dimensions. Because the Global test is based on a multiple regression model, it automatically takes correlation between SNPs, such as caused by linkage disequilibrium, into account. Furthermore, because the Global test is constructed as a test for a single parameter, regardless of the number of SNPs in the pathway, the test is not biased towards large or small pathways [Goeman et al., 2006]. For pathways associated with phenotypes and for candidate pathways, additional Global tests were done for the genes in the pathway to investigate which genes contributed to the pathway-signal.

Control of False Discoveries
To control the risk of false-positive discoveries, Benjamini and Hochberg false discovery rates (FDR-BH) were calculated for each group of gene sets and for each of the outcomes. The use of FDR is preferable to more traditional multiple testing controls because they provide a better balance between the competing goals of finding true positives versus controlling false discoveries, and are much less dependent on the number of tests conducted, and are relatively robust against the effects of correlated tests [Benjamini and Hochberg, 1995]. Five percent of the significant findings were allowed to be false discoveries.

Calculation of R2
An R2 measure of explained variance was calculated by fitting a linear ridge regression model to the MASQ-D30 phenotype using the SNPs in the pathway as predictor covariates. The parameter of the ridge penalty was determined by 10-fold cross-validation using the penalized package (v0.9-32) in R [Goeman and Oosting, 2011], and the cross-validated results were used for the R2 calculation. Because the R2 measure is obtained from a penalized model and calculated using cross-validation, it can be interpreted as an adjusted R2 measure.

Weighted Analysis
In our main analysis, all SNPs contribute equally to the pathway score. As larger genes contain more SNPs, these get more weight than small genes. Thus, this analysis favors pathways in which multiple (independent) polymorphisms in large genes contribute to the phenotype. If the phenotype is influenced by a few SNPs in small genes, the effect is diluted by the many non-associated SNPs in the large genes. In this case, effects can more readily be detected by an analysis in which genes rather than SNPs have equal contribution to the pathway score. Therefore, all pathways were also analyzed with the Global test using weights assigned to each of the SNPs. The weight of a SNP was 1 divided by the number of SNPs in the gene.

Analyses were carried out with R statistical software (v2.11.1).

RESULTS
Sample Description
Table I shows that 548 (39%) respondents had current MDD and 850 (61%) remitted MDD. For General Distress, the median score was 21 (interquartile range (IQR) 15–29), for Anhedonic Depression 23 (IQR 16–31), for Anxious Arousal 15 (IQR 12–20). Thus, there was considerable variability on the dimensions.

Gene Sets
Surprisingly, pathways around important molecules like serotonin, noradrenalin, or BDNF were not associated with symptom dimen-

**TABLE I. Characteristics of the Study Population (N = 1,398)**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)a</td>
<td>43.0 (12.4)</td>
<td>27.4 (12.4)</td>
</tr>
<tr>
<td>Gender (n, % women)</td>
<td>965 (69%)</td>
<td>548 (39)</td>
</tr>
<tr>
<td>Onset age (years)a</td>
<td>27.4 (12.4)</td>
<td>850 (61)</td>
</tr>
<tr>
<td>Current MDD (past month)</td>
<td>21 (15–29)</td>
<td>23 (16–31)</td>
</tr>
<tr>
<td>Remitted MDD</td>
<td>21 (15–29)</td>
<td>15 (12–20)</td>
</tr>
<tr>
<td>MASQ-D30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General distress</td>
<td>21 (15–29)</td>
<td></td>
</tr>
<tr>
<td>Anhedonic depression</td>
<td>23 (16–31)</td>
<td></td>
</tr>
<tr>
<td>Anxious arousal</td>
<td>15 (12–20)</td>
<td></td>
</tr>
</tbody>
</table>

aMean (SD).

bMedian (interquartile range, IQR).
sions. Only Epigenetic Changes after Fear Conditioning was associated with Anhedonic Depression.

After correction for multiple testing, no pathway remained associated with General Distress (see Supplementary Online). However, Anhedonic Depression was associated with six GO-terms, and three gene sets were associated with Anxious Arousal. The variance explained by these pathways ranged from 0.67% to 1.69%.

Figure 1 demonstrates that pathways associated with one of the symptom dimensions show little signal for the other dimensions. Thus, distinct gene sets contribute to different symptom dimensions.

SNPs
The candidate pathway Epigenetic Changes after Fear Conditioning was associated with Anhedonic Depression (FDR 0.0267). Figure 2A shows that the signal for this pathway was not driven by a few important SNPs, as few SNPs showed small \( P \)-values. Further, the signal was driven by SNPs in different genes (see Supplementary Online for the other pathways). Importantly, different SNPs contributed to different dimensions.

Genes
For all significant gene sets, several genes contributed to the pathway signal. Even without adjustment for multiple testing few genes had \( P \)-values < 0.05 (see Supplementary Online). Figure 2B shows associations of the genes in pathway Epigenetic Changes after Fear Conditioning.

However, all GO terms associated with Anhedonic Depression contained the gene NRG1 (unadjusted \( P \)-value \( 5.36 \times 10^{-5} \)), and all GO-terms associated with Anxious Arousal contained FMN2 (unadjusted \( P \)-value \( 1.08 \times 10^{-3} \)). These genes contained 190 and 110 SNPs, respectively, and thus contributed importantly to the pathway scores. The influence of these genes was much diminished in the weighted analysis. Only the KEGG pathway Dorso-Ventral Axis Formation and the GO-term Meiosis showed weak evidence for association in the weighted analysis (both with adj. \( P = 0.0726 \) (see Supplementary Online).

Weighted Analysis
In the weighted analysis, we again tested the 12 candidate pathways, 937 GO-terms and 880 canonical pathways for association with the symptom dimensions. After adjustment for multiple testing, only GO-term Endoplasmatic Reticulum Lumen was associated with General Distress (see Table II).

DISCUSSION
We sought a genetic basis for the clinical heterogeneity among patients with MDD. Our study revealed that distinct genetic factors contribute to different dimensions of depression and anxiety. We
identified seven pathways for Anhedonic Depression. If verified, this implies that Anhedonic Depression is highly complex, and a "simple" treatment might be out of reach. However, in six of these pathways the gene encoding neuregulin (NRG1) appeared. This gene is an important candidate gene for schizophrenia [Mei and Xiong, 2008]. If neuregulin proves to be a key feature of the pathogenesis of Anhedonic Depression, it may provide an important drug target. Three pathways were associated with Anxious Arousal, and only one pathway was associated with General Distress, possibly due to the heterogeneity of this dimension [Den Hollander-Gijsman et al., 2010]. Together, these results indicate that the symptom dimensions underlying anxiety and depression have different etiologies.

General distress has been identified as the core of internalizing disorders [Hettema et al., 2006; Goldberg et al., 2009; Griffith et al., 2010]. Factor analysis of general distress items has clearly established that the items cannot be reduced to a single dimension. Separate subdimensions of depression, anxiety, and anger could be distinguished. Although these factors were strongly correlated, this indicates that General Distress is not unidimensional [Watson et al., 2008]. The structure of General Distress needs to be unraveled to identify more homogeneous phenotypes of this important aspect of emotional disorders.

The gene set associated with General Distress concerned molecules located in the endoplasmatic reticulum (ER). In the ER, secretory and membrane proteins are synthesized and folded with help from ER-chaperones. Malfolded proteins form aggregates which are toxic and cause an ER stress response. A dysregulated ER stress response was shown to be involved in neurodegenerative and in mood disorders [Yoshida, 2007]. Furthermore, mood stabilizers valproate and lithium increased expression of ER chaperones. In our study, unfavorable genotypes at these genes were associated with General Distress.

Seven gene sets were associated with Anhedonic Depression. Several of these had no clear function, but were collections of genes with a common theme. They contained genes that were also included in the candidate pathways, including BCL2, HDAC5, ADRB2, and ACVR1. Also, the themes were highly relevant: muscle cell differentiation contained genes implicated in cell death, mesenchymal stem cells increased hippocampal neurogenesis and ameliorated depression in rodents [Tfilin et al., 2010], cardiac cell differentiation contained motility-genes that are highly expressed in the brain. However, it is unclear how these gene

### TABLE II. Gene Sets Associated With Symptom Dimensions

<table>
<thead>
<tr>
<th>Description</th>
<th>Nr SNPs</th>
<th>P-value</th>
<th>FDR-BH</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Endoplasmatic reticulum lumen G0:0005788</td>
<td>470</td>
<td>4.10 × 10⁻⁴</td>
<td>0.0410</td>
<td>0.93</td>
</tr>
<tr>
<td>Anhedonic depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Regulation muscle cell differentiation G0:0051147</td>
<td>634</td>
<td>3.99 × 10⁻⁵</td>
<td>0.0243</td>
<td>1.69</td>
</tr>
<tr>
<td>2 Regulation of protein binding G0:0043393</td>
<td>596</td>
<td>1.30 × 10⁻⁴</td>
<td>0.0260</td>
<td>0.86</td>
</tr>
<tr>
<td>3 Macromolecular complex disassembly G0:0032984</td>
<td>680</td>
<td>1.76 × 10⁻⁴</td>
<td>0.0260</td>
<td>0.84</td>
</tr>
<tr>
<td>4 Notch signalling pathway G0:0007219</td>
<td>1,041</td>
<td>1.78 × 10⁻⁴</td>
<td>0.0260</td>
<td>0.79</td>
</tr>
<tr>
<td>5 Cardiac cell differentiation G0:0035051</td>
<td>503</td>
<td>2.14 × 10⁻⁴</td>
<td>0.0260</td>
<td>0.67</td>
</tr>
<tr>
<td>6 Epigenetic changes after fear conditioning</td>
<td>367</td>
<td>1.27 × 10⁻³</td>
<td>0.0267</td>
<td>1.22</td>
</tr>
<tr>
<td>7 Mesenchyme development G0:0060485</td>
<td>690</td>
<td>3.38 × 10⁻⁴</td>
<td>0.0343</td>
<td>0.78</td>
</tr>
<tr>
<td>Anxious arousal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Dorso ventral axis formation Kegg hsa04320</td>
<td>478</td>
<td>1.85 × 10⁻⁵</td>
<td>0.0163</td>
<td>1.19</td>
</tr>
<tr>
<td>2 Meiosis I G0:0007127</td>
<td>371</td>
<td>3.60 × 10⁻⁵</td>
<td>0.0221</td>
<td>0.73</td>
</tr>
<tr>
<td>3 Cytokinesis G0:0000910</td>
<td>390</td>
<td>9.62 × 10⁻⁵</td>
<td>0.0293</td>
<td>0.74</td>
</tr>
</tbody>
</table>
sets influence Anhedonic Depression. Below, we discuss GO terms that represent biological processes.

Macromolecular complex disassembly (GO:0032984) contributed to Anhedonic Depression and consists of genes that regulate the dynamic structure of the cytoskeleton. These genes regulate the shape and function of dendritic spines, structures critically involved in learning and memory [Hotulainen and Hoogenraad, 2010]. Many of these genes have been associated with autism, mental retardation, Alzheimer’s disease, and schizophrenia, which like depression are neurodevelopmental or neurodegenerative conditions. In rodents, both early and adult stresses lead to impaired memory accompanied by morphological changes in the hippocampus and frontal cortex. There, cytoskeletal modifications were associated with reduced density of dendritic spines and altered structure of synaptic terminals [Yan et al., 2010]. Thus, a genetic makeup associated with less synaptic structures necessary for memory and learning may contribute to Anhedonic Depression.

Notch signaling (GO:0007219) also contributed to Anhedonic Depression. The adult hippocampus generates new neurons, a process in which Notch1 signaling is involved [Androutsellis-Theotokis et al., 2006]. This neurogenesis has been associated with sad mood [Doetsch and Hen, 2005] and was necessary for behavioral effects of antidepressants [Santarelli et al., 2003; Guo et al., 2009].

Candidate pathway Epigenetic Changes after Fear Conditioning contributed to Anhedonic Depression. Epigenetic changes are modifications of chromatin that have lasting influence on gene expression. For instance, acetylation of histones by histone acetyltransferases reduces their affinity for DNA and facilitates gene expression. In contrast, deacetylation by histone deacetylases (HDACs), or DNA methylation by DNA methyltransferases (DNMTs) is associated with gene silencing. Classical fear conditioning is based on Pavlov’s observations that neutral stimuli can acquire affective properties due to association with aversive stimuli. Animal models of fear conditioning have provided insight into the biology of fear, much of which can be extended to human fear and anxiety [Delgado et al., 2006]. Fear conditioning induced hippocampal expression of DNMT3A and -3B, and DNMT blockers inhibited fear conditioning [Miller and Sweatt, 2007]. Protein phosphatase-1 (PP1) suppresses memory formation and was methylated (silenced) in fear conditioning [Miller and Sweatt, 2007]. RELN encodes reelin, a memory-promoting gene, which was demethylated (activated) in fear conditioning [Miller and Sweatt, 2007]. These epigenetic changes were also involved in depression. In mice, anhedonia was accompanied by histone methylation in hippocampus. Antidepressant imipramine reversed the anhedonia, not by demethylation, but by increased acetylation of the histones [Tsankova et al., 2006]. Thus, aversive environmental signals cause epigenetic changes with lasting effects on mood and anxiety. The genetic profile for this pathway may render individuals more or less sensitive to adverse events. In our study, it was associated with Anhedonic Depression.

Three pathways contributed to Anxious Arousal. The pathway Dorso-Ventral Axis Formation represents a developmental pathway, although the genes are also highly expressed in adult brain. The pathway Cytokinesis regards the division of cytoplasm into two daughter cells, but the generation of dendritic spines is a similar process that shares involvement of many genes. Remarkable was the association of Anxious Arousal with GO-term Meiosis 1. This pathway consists of genes encoding elements of the synaptonemal complex (SC) and regulators of DNA recombination. The SC plays a major role in chromosome pairing and genetic recombination [Page and Hawley, 2003]. Intriguingly, components of the SC [Dietrich and Been, 2001], and pairing of homologous chromosomes [Arnoldus et al., 1989] have been demonstrated in neurons in brain tissue of rats. Moreover, DNA recombination and/or DNA repair in the brain was necessary for long-term memory in fear conditioning [Wang et al., 2003; Colon-Cesario et al., 2006; Saavedra-Rodriguez et al., 2009]. Many proteins involved in homologous recombination are also active in somatic DNA recombination/repair. DNA recombination has been suggested to contribute to neuronal diversity in the developing brain, like V(D)J rearrangement contributes to antibody diversity in the immune system [Chun and Schatz, 1999; Muotri and Gage, 2006]. Interestingly, in a GWAS for bipolar disorder, GO-term V(D)J recombination was high-ranking [Holmans et al., 2009]. Thus, genetically determined variation in DNA recombination in the brain may contribute to variation in Anxious Arousal. However, the nature of this process remains to be elucidated.

Surprisingly, candidate pathways had little effect on the symptom dimensions. Explorative network analyses may be more likely to identify gene sets around seemingly important genes, as they visualize interactions or shared functions of selected SNPs [Baranzini et al., 2009; Vink et al., 2009]. Subsequent verification with Global test in a replication cohort would provide convincing evidence for association.

In the pathways that showed significant association, only few genes and SNPs had small P-values. This is in line with the idea that many SNPs with small effects are involved. Recent GWASs in thousands of individuals have identified 50 SNPs associated with height, which together explained 5% of phenotypic variance [Yang et al., 2010]. Another 40% of variation was explained by the rest of the SNPs. In this light, our observations of 0.67–1.69% variance explained by the gene sets are promising.

In most pathway analyses, the pathway score is calculated from gene scores. Often, the genes are summarized by the most significant SNP [Wang et al., 2007; Holmans et al., 2009; Yu et al., 2009]. As larger genes may contain more SNPs, these are more likely to be selected [Wang et al., 2007]. To prevent this gene size bias, such analyses need adjustment for gene size. However, adjustment for gene size is not necessary but optional in the Global test. The Global test uses all SNPs in a pathway, adjusts for correlation between markers (i.e., linkage disequilibrium), and its results are valid. Importantly, gene-based pathway tests have different properties than SNP-based pathway tests. In the SNP-based analysis, NRG1 strongly influenced the pathway score for six GO-terms in association with Anhedonic Depression. This gene contained 190 SNPs, and the signal was not due to a few highly significant SNPs, but rather to a larger number of less impressive signals. In the weighted analysis, the contribution of these 190 SNPs was down-weighted to let every gene have an equal contribution to the pathway. This diminished the effect of NRG1 to a large extent. SNP-based analyses may favor pathways in which multiple independent signals are present in large genes. In gene-based analyses,
single signals in smaller genes have as much weight as larger genes. Therefore, it is not surprising that we observed different results in the SNP-based and gene-based analyses. For only three pathways, both the SNP-based and the gene-based results were in the top-10 associations.

Two other (regression) methods also favor pathways containing genes with multiple, independent disease-associated SNPs. GRASS jointly analyses principal components within genes to compute a pathway score [Chen et al., 2010], and the Plink pathway score is the average statistic of all the (selected) SNPs in the pathway [Purcell et al., 2007]. In the brain, many very large genes are expressed, so for psychiatric disorders, the power of the Global test, GRASS and Plink set-test to incorporate multiple independent effects in these genes may be favorable over other pathway approaches. These methods were shown to yield similar results [Deelen et al., 2011].

Symptoms of depression and anxiety are often used to measure state effects, like response to treatment. However, as these symptoms were shown to be quite stable during a 7- to 8-year follow-up period, they also possess trait-like characteristics [Ormel and Wohlfarth, 1991; Kendler and Gardner, 2011]. The stable part of symptoms was demonstrated to reflect the level of neuroticism, which is predetermined by genes and childhood adversities. The variable part of symptoms of depression and anxiety was shown to be due to episodes of MDD and GAD, caused by genetic risk and adversities during childhood and recent adulthood [Kendler and Gardner, 2011]. In line with this, among patients with a diagnosis of depression and/or anxiety at baseline, 2-year course trajectories were predicted by baseline scores for dimensions of depression and anxiety [Wardenaar et al., 2012]. Thus, symptoms of depression and anxiety seem to have a strong genetic basis. Indeed, twin studies have demonstrated large effects of genetic factors on symptoms of depression and anxiety [Gillespie et al., 2004; Boomsma et al., 2005; Kendler et al., 2008]. Likewise, a substantial genetic component has been demonstrated for General Distress and Anhedonic Depression [Clark and Watson, 2008].

Previously, our group has shown distinct associations of General Distress, Anhedonic Depression, and Anxious Arousal with depression and anxiety [Bielcikova et al., 2007; Kendler et al., 2008]. Likewise, a substantial genetic component has been demonstrated for General Distress and Anhedonic Depression [Clark and Watson, 2008].

To conclude, we identified pathways associated with Anhedonic Depression and Anxious Arousal. Many of these processes were known from animal studies to be involved in depression or anxiety, and now we demonstrated that common variability in these processes contributes to psychopathology in humans. Our observations help to understand the heterogeneity of symptoms among patients with depression. These pathways provide leads for further study to understand the neurobiology of depression and anxiety.


