Psychiatry Research 206 (2013) 117–121

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Brief report

The study of candidate genes related to the neurodevelopmental hypothesis of anorexia nervosa: Classical association study versus decision tree

Monika Dmitrzak-Weglarz [a,*], Jerzy Moczko [c], Maria Skibinska [a], Agnieszka Slopien [a,b], Marta Tyszkiewicz [b], Joanna Pawlak [a], Dorota Zaremba [a], Aleksandra Szczepankiewicz [a], Andrzej Rajewski [b], Joanna Hauser [a]

[a] Psychiatric Genetics Unit, Department of Psychiatry, Poznan University of Medical Sciences, Poznan, Poland
[b] Department of Child and Adolescent Psychiatry, Poznan University of Medical Sciences, Poznan, Poland
[c] Department of Computer Sciences and Statistics, Poznan University of Medical Sciences, Poznan, Poland

ARTICLE INFO

Article history:
Received 16 May 2012
Received in revised form
4 September 2012
Accepted 7 September 2012

Keywords:
Anorexia nervosa
Multifactor dimensionality reduction
Decision trees

ABSTRACT

In this research, we conducted a study of genes connected with the neurodevelopmental hypothesis of anorexia nervosa, using classical statistical and data-mining methods to establish a relationship with disease risk and algorithms to identify the best genetic predictors of anorexia nervosa.

1. Introduction

Anorexia nervosa (AN) is a serious and ever-growing social problem. Although family and twin studies have suggested the existence of a strong genetic component in AN (heritability 33–84%) (Klump et al., 2001; Kortegaard et al., 2001; Wade et al., 2000), the search for genetic susceptibility factors is still underway. However, the results of classical candidate gene association studies of AN have been inconclusive (Helder and Collier, 2011; Hinney et al., 2010).

In association studies of polygenic disorders, complex analysis (e.g., genome-wide association studies (GWASs)) is used more often than the analysis of individual polymorphisms. The use of haplotype analysis in association studies increases both the power and the sensitivity of the test (Liu et al., 2008). It has also been proposed that, in the case of multifactorial disorders that arise due to the cumulative impact of small effects of many genetic variants, gene–gene interactions should be analysed. Several candidate gene approaches disclosed an intriguing association between AN and several genetic variants, whereas a GWAS identified only two loci (Wang et al., 2010). One potential explanation for this discrepancy is that GWASs provide a huge quantity of information, which is a challenge for bioinformatics. With the use of algorithms during multiple testing, many individual single-nucleotide polymorphisms (SNPs) (especially with relatively small impacts on the disease risk) can be missed. Several authors have suggested that GWASs are more informative about novel biological pathways than for clinical diagnosis (Hirschhorn and Gajdos, 2011; Maher et al., 2008).

This report proposes the use of data-mining methods to develop predictive models with results that could be compared with the results of classic association analyses. We decided to use two techniques to identify the best genetic predictors of AN status: logistic regression and decision trees (DTs) (Hastie et al., 2001).

To test classical and novel statistical methods, we selected genes based on the neurodevelopmental model of AN. Connan et al. (2003) proposed that the development of a neurodevelopmental model for AN is a promising line of research for the explanation of the primary predisposition to AN. In this hypothesis, genetic factors involved in the development and maturation of the central nervous system (CNS) combined with environmental stress factors acting at an early stage in CNS development can cause a dysfunction of the hypothalamic–pituitary–adrenal (HPA) axis and create a predisposition to disease later in life (Connan et al., 2003). According to neurodevelopmental concepts, neurotrophic factors and their receptors might play a significant role in the pathophysiology and aetiology of AN. In particular, we focussed on genes that are related to neurotrophic signalling cascade encoding: the neurotrophin brain-derived neurotrophic

* Correspondence to: Psychiatric Genetics Unit, Department of Psychiatry; University of Medical Sciences, Poznan, Poland ul. Szpitalna 27/33, 60-572, Poznan, Poland. Tel.: +48 61 8491311; fax: +48 61 8480392.
E-mail address: mweglarz@ump.edu.pl (M. Dmitrzak-Weglarz).

0165-1781/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psychres.2012.09.005

factor (BDNF) and its receptor, neurotrophic tyrosine receptor kinase, type 2 (NTRK2) (Kernie et al., 2000; King, 2006; Xu et al., 2003); the downstream intercellular kinases *FYN* (Yamada and Nabeshima, 2004) and *GSK3β* (Foulstone et al., 1999; Mai et al., 2002); genes linked with cognition, learning and memory, such as *GRIN2A* and *GRIN2B*, which encode subunits 2A and 2B of the *N*-methyl-D-aspartic acid (NMDA) receptor (de Quervain and Papassotiropoulos, 2006; Endele et al., 2010; King et al., 2010); and the synaptosomal associated protein of 25 kD (SNAP-25), which is one of the core proteins involved in the regulation of neurotransmitter release (Frassoni et al., 2005).

2. Methods

2.1. Patients and controls

All patients and healthy controls involved in the study were of Polish origin. Patients ($n=256$; mean age 17.5 ± 3.3 years) were recruited from inpatients treated in the Department of Child and Adolescent Psychiatry of Poznan University of Medical Sciences. AN was diagnosed according to International Classification of Diseases Tenth Revision (ICD-10) and Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM-IV) criteria.

Control subjects ($n=167$; mean age 19.6 ± 3.2) were recruited from among carefully chosen volunteers without eating disorders or other mental disorders as well as without a family history of these disorders among their first-degree relatives. A detailed description of the patient and control groups is presented in Table 1.

2.2. SNP selection and genotyping method

We genotyped 23 SNPs from seven genes selected based on the neurodevelopmental hypothesis of AN (*BDNF, NTRK2, FYN, GSK3β, GRIN2B, GRIN2A* and *SNAP-25*). The included polymorphisms fulfilled the following criteria: functionality (in experimental functional studies), high frequency (minor allele frequency (MAF) above 0.10) or previously reported associations for mental disorders (both positive and negative findings).

The genotypes were determined by a previously described method (Dmitrzak-Weglarz, 2012).

2.3. Data analysis

The statistical analyses were performed with statistical packages and online packages such as STATISTICA v. 9.0 (Pearson's chi-squared (χ^2) test; Fisher's exact test; logistic regression and DTs), QUANTO v. 1.2.4 (power calculations), Haploview v. 4.2 (linkage disequilibrium (LD) analysis, Hardy–Weinberg equilibrium (HWE)), MDR 2.0_beta_8.3 (gene × gene interaction) and GraphPad InStat 3 (odds ratios (ORs) with a 95% confidence interval (CI)).

3. Results

3.1. Association analysis

In the molecular analysis of 23 polymorphisms in seven candidate genes, associations with AN risk were found for the *BDNF* gene (rs6265 and rs2030324), *NTRK2* gene (rs2289656), *FYN* gene (rs6916861), *GSK3β* gene (rs334558) and *GRIN2B* gene (rs890, rs1806201) with nominal *P* values in the range of 0.002–0.021 for genotypes and 0.000–0.039 for alleles (Table 2).

Power testing was determined for all analysed SNPs based on the OR arising from the observed allele frequencies in the population of patients and healthy controls. The power calculations for empirically estimated ORs ranging from 0.12 to 1.83 were evaluated for all polymorphisms as follows:

BDNF: rs2203877 –5%, rs6265 – 65%, rs988748 –30%, rs2030324 – 83%;
NTRK2: rs1187326 – 26%, rs993315 – 23%, rs1187327 – 12%, rs2289656 – 16%;
FYN: rs6916861 – 64%, rs3730353 – 20%, rs706895 – 28%;
GSK3β: rs334558 – 66%;
GRIN2B: rs890 – 98%, rs1806201 – 70%, rs7301328 – 9%, rs1019385 – 5%, rs3764028 – 5%;
GRIN2A: rs1014531 – 40%, rs727605 – 5%, rs11859727 – 10%;
SNAP-25: rs363050 – 16%, rs8636 – 5%, rs362552 – 5%

3.1.1. HWE analysis

Genotype distributions for all studied polymorphisms were in concordance with HWE ($P > 0.05$) except for the rs2203877 polymorphism of the *BDNF* gene.

3.2. LD disequilibrium and haplotype analysis

The LD analysis showed the existence of haplotype variants associated with either increased or decreased risk of AN. The haplotype rs890G/rs1806201A of the *GRIN2B* gene was associated with the risk of AN ($P^* < 0.000$). We also found three haplotypes: the haplotype rs6265G/rs988748C/rs2030324C of the *BDNF* gene ($P^*=0.038$); haplotype rs993315C/rs1187327G of the *NTRK2* gene ($P^*=0.001$); and haplotype rs890T/rs1806201G of the *GRIN2B* gene ($P^*=0.032$), that may be protective against AN. The associations in all of the above-mentioned haplotypes were sufficiently strong to persist after adjustment (P^*) for multiple comparisons using 10 000 permutations. For analysed polymorphisms of *FYN*, *GRIN2A* and *SNAP-25* genes, no LD was observed.

3.3. Gene–gene interactions

Multifactor dimensionality reduction (MDR) was used to analyze gene–gene interaction models in AN; however, we did not observe any significant interaction between analysed genes ($P=0.096$).

3.4. Predictive models for AN

3.4.1. Regression analysis

All analysed polymorphisms were included in the logistic regression analysis. A significant model with a *P* value of 0.00004 was created. In this model, a significant predictive value was observed for polymorphisms, in *NTRK2* (rs993315 and rs1187327), *FYN* (rs6916861), *GRIN2B* (rs890 and rs3764028) and *SNAP-25* (rs362552). The quality of classification of analysed objects using this model was estimated to be 73.7% (with an 88.5% positive classification of the patients and 45.28% classification of the control group). The reduction of analysed polymorphisms to polymorphisms with significant predictive value showed a decreased prediction accuracy (64.45%).

3.4.2. DT

In the first stage, all of the analysed polymorphisms in the DT analysis were included. In this analysis, a minimum incorrect amount equal to 30 was accepted. With this number of incorrect classifications, the growth of the tree was completed. In the direct

Table 1
Description of analyzed population.

	Anorexia nervosa patients	Control group
Number of subjects (*n*)	256	167
Age (years)	17.482 ± 3.341	19.617 ± 3.186
BMI	14.389 ± 1.945	21.469 ± 9.578
BECK	17.500 ± 11.585	6.708 ± 6.830
Education (years)	11.306 ± 2.852	13.509 ± 3.008
Age of onset (years)	14.983 ± 2.842	–
Family history (*n*) (%)	71 (27.73)	–