

ORIGINAL ARTICLE

Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers

Kazuhiko Nakabayashi^{1,2,7}, Gen Komaki³, Atsushi Tajima⁴, Tetsuya Ando³, Mayuko Ishikawa¹, Junko Nomoto¹, Kenichiro Hata⁵, Akira Oka⁴, Hidetoshi Inoko⁴, Takehiko Sasazuki⁶, Japanese Genetic Research Group for Eating Disorders (JGRED) and Senji Shirasawa²

The Japanese Genetic Research Group for Eating Disorders (JGRED) is a multisite collaborative study group that was organized for the systematic recruitment of patients with an eating disorder for the purpose of genetic study in Japan. We conducted a genome-wide case-control association study using 23 465 highly polymorphic microsatellite (MS) markers to identify genomic loci related to anorexia nervosa (AN). Pooled DNA typing in two screening stages, followed by individual typing of 320 AN cases and 341 controls, allowed us to identify 10 MS markers to be associated with AN. To narrow down genomic regions responsible for the association of these MS markers, we further conducted a single-nucleotide polymorphism (SNP) association analysis for 7 of the 10 loci in 331 AN cases and 872 controls, which include the 320 AN cases and the 341 controls genotyped in the MS screening, respectively. Two loci, namely 1q41 and 11q22, remained significantly associated with AN in the SNP-based fine mapping, indicating the success in narrowing down susceptibility regions for AN. Neither of these loci showed a positive evidence of association with bulimia nervosa. The most significant association was observed at SNP rs2048332 (allelic P -value=0.00023) located at 3'-downstream of the *SPATA17* gene on the 1q41 locus. The association analysis for MS-SNP haplotypes detected a statistically significant association (permutation P -value=0.00003) of the A-4-G-T haplotype that comprised four SNP/MS markers (rs6590474-D11S0268i-rs737582-rs7947224) on the 11q22 locus with AN. This linkage disequilibrium block spanning a 20.2-kb interval contains exon 9 of the *CNTN5* gene encoding contactin 5.

Journal of Human Genetics (2009) 54, 531–537; doi:10.1038/jhg.2009.74; published online 14 August 2009

Keywords: anorexia nervosa; bulimia nervosa; eating disorders; genome-wide association study; microsatellite markers; SNPs

INTRODUCTION

Eating disorders (EDs) are psychiatric conditions characterized by severe disturbances in eating behavior, and can be classified into three major types, namely anorexia nervosa (AN), bulimia nervosa (BN) and ED not otherwise specified. AN and BN are not mutually exclusive conditions as some individuals cross over between both conditions.^{1,2} It is generally believed that ED has become more frequent over recent decades. Fairburn and Harrison³ summarized the prevalence of AN and BN to be 0.7% in teenage girls and 1–2% in 16–35-year-old females, and the incidence (per 100 000 per year) of AN and BN to be 19 in females and 2 in males, and 29 in females and 1 in males, respectively.³ AN has the highest mortality rate (5–6%) of any psychiatric disease,⁴ whereas the mortality rate of BN is reported to be 0.3%.⁵

The cause of ED is complex and poorly understood. However, the involvement of genetic factors in the etiology of ED has been demonstrated in family and twin studies.^{6–9} Twin studies have estimated the contribution of genetic factors in AN to be between 33 and 84%⁹ and that in BN to be between 28 and 83%.⁸ To search for the genetic etiology of ED, two types of molecular genetic approaches, namely linkage studies and association studies, have been carried out (reviewed in Bulik *et al.*¹⁰ and Pinheiro *et al.*¹¹). Linkage studies for AN have detected significant linkage at two regions on chromosome 1^{12,13} and an additional suggestive linkage at a number of loci.^{13,14} From the 1p33-p36 region, one of the regions showing significant linkage to AN,¹² serotonin 1D receptor (*HTR1D*) and opioid delta receptor (*OPRD1*) genes were further evaluated and found to exhibit significant association with AN.¹⁵ Linkage analysis of a BN cohort detected

¹Department of Pathology, Research Institute, International Medical Center of Japan, Shinjuku, Tokyo, Japan; ²Department of Cell Biology, Faculty of Medicine, Fukuoka University, Jonan, Fukuoka, Japan; ³Department of Psychosomatic Research, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawahigashicho, Kodaira, Japan; ⁴Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan; ⁵Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Setagaya, Tokyo, Japan and ⁶International Medical Center of Japan, Shinjuku, Tokyo, Japan

⁷Current address: Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo 157-8535, Japan
Correspondence: Dr K Nakabayashi, Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo 157-8535, Japan.

E-mail: knakabayashi@nch.go.jp

Received 15 April 2009; revised 2 July 2009; accepted 15 July 2009; published online 14 August 2009

significant linkage at 10p13 and a suggestive linkage at 14q22.2-23.1.¹⁶ Fine mapping of these regions with significant/suggestive linkage signals will clarify whether the regions contain gene(s) relevant to AN or BN. Association studies for ED have so far been limited to candidate gene approaches that focused on the genes, such as those involved in the regulation of feeding and body composition and those implicated in neurotransmitter pathways regulating behavior.^{10,11,17} Although many association studies performed for ED are considered to be statistically underpowered because of their small sample sizes and/or suffer from multiple testing,¹⁰ positive findings on *HTR1D*, *OPRD1* and *BDNF* genes seem promising, as their association with AN/BN has been replicated in more than one study in which relatively large numbers of samples were enrolled (reviewed in Bulik *et al.*¹⁰).

Completion of the Human Genome Project¹⁸ and the rapid progress of the International HapMap Project¹⁹ dramatically increased the amount of information on genetic markers, such as microsatellite (MS) markers and single-nucleotide polymorphisms (SNPs). Consequently, statistical strategies and genotyping platforms for genome-wide association studies (GWASs) have been established and prevailed as a means of identifying disease susceptibility genes. Disease association studies using MS markers distributed across the human genome have advantages over linkage analysis and the candidate gene approach. MS markers are highly polymorphic and show a high degree of heterozygosity (~70% on average), and their linkage disequilibrium (LD) lengths are in the 100 kb range.^{20–22}

In this study, we adopted a practical and efficient GWAS strategy for AN using a set of 23 465 MS markers^{22,23} and the DNA pooling method, which has been adopted to identify novel susceptibility genes of rheumatoid arthritis²² and candidate loci for hypertension²⁴ and adult height.²³ We identified 10 novel loci related to AN by the MS marker-based GWAS strategy, and subsequently conducted an SNP-based association analysis for 7 of the 10 loci to further narrow down candidate genomic intervals responsible for AN susceptibility.

MATERIALS AND METHODS

Subjects

The patients enrolled in this study were recruited through the efforts of the Japanese Genetic Research Group for Eating Disorders (JGRED), which comprises 67 nation-wide hospitals/institutions (the full list is available in Supplementary Table 1). A total of 456 unrelated Japanese female patients with ED (331 cases with AN and 125 cases with BN) participated in this study. According to the Diagnostic and Statistical Manual of Mental Disorders,²⁵ 218 and 113 cases were diagnosed as AN-restricting type (AN-R) and as AN with binge eating/purging type (AN-BP), respectively. Among the 125 BN cases, 46 had histories of AN and 79 cases did not. The average age at assessment was 23.2 ± 7.91 (s.d.) years for AN-R, 25.7 ± 7.22 years for AN-BP and 26.3 ± 6.77 years for BN. The lifetime minimum body mass index was 12.4 ± 1.90 (s.d.) kg m^{-2} for AN-R, 12.8 ± 2.67 kg m^{-2} for AN-BP and 16.2 ± 3.47 kg m^{-2} for BN. A total of 872 Japanese healthy individuals participated in this study: 180 female individuals whose average age was 34.5 years (control group 1) and 692 male and female volunteers recruited among university students (control group 2). The average age and gender ratio of control group 2 were unavailable. The ethics committees of all facilities approved the investigation. All subjects gave their written informed consent before participation in the study.

Pooled DNA construction and MS Genotyping

Among the 27 037 MS markers developed by Tamiya *et al.*,²² 23 465 (with average spacing 118.0-kb) were used in this study. Detailed information on the 27 037 MS markers is also available at the Japan Biological Information Research Center website (<http://jbirc.jbic.or.jp/gdbs/database/viewer/download/list.jsp>). The pooled DNA method for MS typing was performed according to the protocol of Collins *et al.*²⁶ with a slight modification.²¹ Genomic

DNA was extracted from peripheral blood using the Qiagen blood mini kit (Qiagen, Hilden, Germany). DNA concentration was determined using Pico-Green fluorescence assay (Molecular Probes, Eugene, OR, USA) as described previously.^{22,26} The detailed conditions for PCR amplification and peak detection for pooled DNA and individual genotyping were described previously.^{22,24} A total of 320 AN cases and 341 controls were subjected to MS genotyping. The number of subjects for pooled DNA typing was 90 AN-R cases and 90 female controls from control group 1 in the first-stage screening, and another 90 AN-R cases and another 90 female controls from control group 1 in the second-stage screening. The number of subjects for individual genotyping in the third screening was 140 AN cases (composed of 32 AN-R and 108 AN-BP cases) and 161 controls from control group 2. As the final step of MS screening, positive markers were subjected to individual genotyping of all 320 AN cases and 341 controls used in the three screening stages. MS markers showing statistical significance ($P < 0.05$) in Fisher's exact test for the 2×2 contingency table in the first screening were subjected to the second screening. In the second and third screening stages, besides statistical significance ($P < 0.1$), consistency in the directions of effect of the associated MS allele on AN susceptibility was considered; when the allele holding statistical significance showed the opposite direction of effect compared with that in the previous screening stage, such a marker was excluded from further analysis.

SNP genotyping

Single-nucleotide polymorphisms in candidate regions were selected from the SNP database (<http://www2.appliedbiosystems.com/>) using SNPbrowser software 3.5 (Applied Biosystems, Foster City, CA, USA). SNPs were genotyped using TaqMan assays, which were carried out using standard protocols for ABI PRISM 7900HT Sequence Detection Systems (Applied Biosystems). A total of 331 AN cases, 125 BN cases and 872 control individuals were subjected to SNP genotyping. The 331 AN cases consisted of the 320 cases genotyped in the MS screening and 11 additional cases. The 872 controls consisted of 180 female controls from control group 1 and 692 controls from control group 2.

Statistical analyses

Allelic frequencies in pooled DNA genotyping were estimated from the height of peaks measured using the PickPeak and MultiPeaks programs (Applied Biosystems). To calculate allelic P -values from MS genotyping data, we used Fisher's exact test for the 2×2 contingency tables for each individual allele and for the $2 \times m$ contingency tables for each locus, where m refers to the number of marker alleles observed in the genotyped population.

The power of each of the three-stage MS screenings was calculated using a 'Genetic Power Calculator'²⁷ (<http://pngu.mgh.harvard.edu/~purcell/gpc/>) in the AN prevalence rate of 1%. In the three successive screening, the overall power of the study was 72 and ~50% for detecting an AN-susceptible allele with a genotype relative risk of 2.0 and 1.8, respectively, under an additive model in log-odds scale when the susceptible allele (with a frequency of 0.2) and an AN-associated MS marker were in complete LD.

For SNP genotyping, disease associations were assessed by the χ^2 test mainly using Haploview 4.0 software (<http://www.broad.mit.edu/mpg/haploview/>).²⁸ As a multistep analysis was used, nominal P -values were corrected with 10 000 iterated permutations for a series of SNPs selected for each candidate genomic interval. The significance level for SNP association was set at 0.05 throughout the study. Haplotype association analysis (100 000 iterated permutations) was also performed using Haploview 4.0. LD blocks in each candidate locus were defined using the default algorithm, namely confidence intervals,²⁹ of Haploview 4.0. SNPAnalyze v7.0 software (Dynacom, Mobarra, Japan) was used to perform LD calculation, haplotype inference, identification of haplotype-tagging SNP/MS and case-control haplotype analysis (100 000 iterated permutations) for combined data of MS and SNP genotypes.

RESULTS

Phased genomic screens using DNA pools

In the first-stage screening, among the 23 465 MS markers subjected to genotyping, 1414 (6.0%) were considered to be statistically significant ($P < 0.05$) and were subjected to the next screening stage. In the second-stage screening, among the 1414 markers tested, 158 satisfied

our selection criteria: statistical significance ($P < 0.1$) and the same direction of allelic effect on AN susceptibility between the first and the second screening results. In the third screening, among the 158 markers tested, 16 satisfied our selection criteria: statistical significance ($P < 0.1$) and the same direction of allelic effect throughout the MS screening stages. We determined a significance threshold to control false-positive rates (nominal $\alpha = 0.05$) in the first stage of MS screening. In the second and third stages, considering that our sample sizes of cases and controls are not large, we set the significance thresholds (nominal $\alpha = 0.1$) to maintain the overall statistical power of the screening.

To determine the definite allele frequencies of the selected 16 MS markers, we performed individual genotyping on all the AN cases ($n = 320$) and controls ($n = 341$) used in the first to third screenings. Of the 16 markers, 10 showed a statistically significant difference by Fisher's exact test in the comparison between controls and the AN cohort (Table 1). After correction of multiple tests with the number of alleles, 7 of 10 markers remained statistically significant ($P_c < 0.05$).

SNP association analysis to narrow down the regions responsible for AN susceptibility

From the 10 MS markers that were found to be associated with AN, we selected 7 (shown in bold in Table 1), on the basis of the gene content around each marker, as targets for SNP association analysis to narrow down disease susceptibility loci. We primarily selected a collection of evenly spaced SNPs (11.1-kb interval on average) within a several 100-kb region surrounding each candidate MS marker, although it should be noted that intragenic SNPs were preferentially selected from the loci of 1q41 (spermatogenesis associated 17 (*SPATA17*)), 5q15 (*CDH18*) and 18q22 (*NETO1*). The number of SNPs subjected to association analysis and the size and nucleotide positions of the corresponding genomic interval are listed for each locus in Table 2.

In total, we performed genotyping for 333 SNPs on 331 AN cases and 872 controls. Among the 251 SNPs that satisfied thresholds for the Hardy-Weinberg equilibrium (exact test $P > 0.01$) and minor allele frequency ($> 5\%$), 24 showed a statistically significant association (nominal $P < 0.05$) with the AN cohort (Table 2). For each of the seven loci analyzed, nominal P -values were corrected with 10 000 iterated permutations using Haploview 4.0. In all, 3 SNPs, all of which

are located on 1q41, out of the 24 SNPs remained statistically significant ($P_c < 0.05$) (underlined in Table 2).

Subsequently, using Haploview 4.0, we inferred LD block structures for each candidate chromosomal region, and performed a haplotype association analysis (100 000 iterated permutations) for the constructed LD blocks. Significant association ($P_c < 0.05$) with the AN cohort was detected in three of the six SNP haplotype blocks defined in the 1q41 locus, and in one of the eight blocks defined in the 11q22 locus (Figure 1 and Table 3).

1q41

A total of 38 SNPs were selected for genotyping within a 337.3-kb interval, including the AN-associated MS marker D1S0562i. Among 30 SNPs subjected to association analysis, 7 showed a statistically significant association ($P < 0.05$) with the AN cohort (Table 2). All the seven SNPs were located at 3'-downstream of the *SPATA17* gene (Figure 1, left). SNP rs2048332 showed the most significant association (allelic $P = 0.00023$) and was further analyzed under different genetic models. Association analysis under a recessive model for rs2048332 showed the lowest P -value of 0.00015 with the CC genotype, indicating that the CC genotype of rs2048332 has a susceptible effect on the AN phenotype in the Japanese (odds ratio = 1.73, confidence interval, 1.30–2.31). Among the three AN-associated haplotype blocks (1q41-#4, #5 and #6 in Figure 1, left and Table 3), 1q41-#5 that comprised two SNPs, namely rs1397178 and rs2048332, spanning a 10.2-kb interval, was found to be most significantly associated ($P_c = 0.0039$).

The AN-associated MS marker D1S0562i was located in block 1q41-#6, which comprised five SNPs spanning a 38.1-kb interval, and was also associated with AN ($P_c = 0.038$). Four of the five SNPs binned to haplotype block 1q41-#6, rs17691163, rs34418611, rs1934216 and rs1538555, were in a relatively strong pairwise LD ($D' = 0.72$ – 0.75) with D1S0562i, whereas the most significant SNP, rs2048332, in 1q41-#5 block was in modest LD ($D' = 0.46$) with it. The four SNPs and D1S0562i (rs17691163–D1S0562i–rs34418611–rs1934216–rs1538555) were selected as tags captured through LD in block 1q41-#6. These haplotype tags were subjected to an MS-SNP haplotype-based association analysis: one haplotype (G-2-A-T-G), tagged by an AN-associated risk allele of D1S0562i (Supplementary Table 2), was significantly associated with AN ($P_c = 0.0065$) (Table 4).

Table 1 Ten microsatellite markers showing statistically significant differences in the individual genotyping

MS Marker	Cytoband	No. of alleles	Positive alleles ^a	Allele frequencies		2×2		2×m	Odds ratio	95% CI
				Control (N=341)	AN (N=320)	P	P _c	P		
D1S0016i	1p36	12	2	13.0%	20.3%	<u>0.00047</u>	<u>0.0056</u>	<u>0.014</u>	1.70	1.26–2.28
D1S0562i	1q41	9	2	12.4%	18.1%	<u>0.0043</u>	<u>0.039</u>	0.054	1.57	1.16–2.12
D5S0853i	5q15	17	1	12.5%	7.7%	<u>0.0047</u>	0.080	0.29	0.59	0.41–0.85
D11S0389i	11q13	16	1	6.5%	10.8%	<u>0.038</u>	0.61	0.85	1.75	1.18–2.60
D11S0268i	11q22	16	2	17.8%	26.0%	<u>0.00039</u>	<u>0.0062</u>	<u>0.0057</u>	1.62	1.24–2.11
D12S0245i	12q14.1	3	2	50.0%	57.5%	<u>0.0067</u>	<u>0.020</u>	<u>0.024</u>	1.36	1.09–1.69
D12S0848i	12q23.2	11	1	14.6%	21.1%	<u>0.0029</u>	<u>0.032</u>	<u>0.040</u>	1.56	1.17–2.08
G09961	16q12	11	1	17.6%	11.1%	<u>0.00075</u>	<u>0.008</u>	<u>0.0043</u>	0.58	0.43–0.80
D18S0019i	18q22	13	3	36.8%	26.5%	<u>0.000070</u>	<u>0.00091</u>	<u>0.0071</u>	0.61	0.48–0.77
D19S0081i	19p13.3	13	1	14.2%	10.4%	<u>0.044</u>	0.57	<u>0.017</u>	0.70	0.50–0.98

Abbreviations: P, Fisher's exact test P-value; P_c, P-value corrected for the number of alleles.

P- and P_c-values smaller than 0.05 are underlined.

^aNumber of the alleles showing significant differences ($P < 0.05$). When more than one positive allele were detected, the allele frequencies and the P and P_c-values of the positive allele showing the most significant difference are listed.

Table 2 SNP allelic association with AN

Cytoband, # of SNPs subjected to association analysis, interval size, and nucleotide positions (NCBI Build 36.1)	SNP ID	Location (NCBI Build 36.1)	Gene symbol	SNP Type	Allele	Frequency		HWE P-value	P-value	Pc	Odds ratio	95% CI
						Control (n=872)	Case (n=331)					
1p36	rs3753499	chr.1 7834024	UTS2	Intron	T	0.868	0.903	0.59	0.019	0.375	1.42	1.06~1.90
54 SNPs, 332.8 kb chr1: 7695967-8028788	rs106968	chr.1 7897131		Intergenic	G	0.397	0.447	0.017	0.027	0.480	1.23	1.02~1.47
1q41	rs11117962	chr.1 216116886		Intergenic	T	0.424	0.500	0.41	0.00088	0.106	1.36	1.13~1.62
30 SNPs, 337.3 kb chr1: 215887318-216224643	rs907057	chr.1 216125221		Intergenic	C	0.354	0.400	0.29	0.039	0.710	1.21	1.01~1.46
	rs7538903	chr.1 216135516		Intergenic	C	0.453	0.530	0.60	0.00074	0.022	1.36	1.14~1.63
	rs1397178	chr.1 216166200		Intergenic	T	0.283	0.326	0.16	0.039	0.643	1.23	1.01~1.49
	rs2048332	chr.1 216176389		Intergenic	C	0.446	0.530	0.61	0.00023	0.009	1.40	1.17~1.68
	rs12125437	chr.1 216186564		Intergenic	T	0.311	0.382	0.75	0.00092	0.062	1.37	1.14~1.65
	rs1538555	chr.1 216224643		Intergenic	G	0.311	0.388	0.71	0.00036	0.039	1.40	1.16~1.69
11q13	rs11227668	chr.11 66568914	SYT12	Intron	A	0.214	0.258	0.39	0.019	0.243	1.28	1.04~1.58
20 SNPs, 441.3 kb chr11: 66365567-66806868	rs3741190	chr.11 66573083	SYT12	UTR 3	T	0.207	0.248	0.82	0.029	0.330	1.26	1.02~1.56
11q22	rs12574821	chr.11 99333169	CNTN5	Intron	G	0.207	0.264	0.74	0.0027	0.100	1.38	1.12~1.69
56 SNPs, 699.8 kb chr11: 98931758-99631604	rs1349782	chr.11 99335431	CNTN5	Intron	G	0.395	0.443	0.49	0.034	0.678	1.22	1.01~1.46
	rs6590474	chr.11 993358383	CNTN5	Intron	A	0.164	0.214	0.68	0.004	0.137	1.39	1.11~1.75
16q12	rs2270355	chr.16 50958219		Intergenic	G	0.086	0.115	0.29	0.031	0.512	1.38	1.03~1.84
34 SNPs, 324.9 kb chr16: 50819155-51144082	rs1111482	chr.16 51021032		Intergenic	T	0.858	0.902	0.23	0.0040	0.094	1.52	1.14~2.04
	rs11647880	chr.16 51022612		Intergenic	A	0.858	0.902	0.49	0.0045	0.102	1.52	1.14~2.03
	rs8062936	chr.16 51033470	TOX3	Intron	A	0.863	0.903	0.30	0.0077	0.171	1.48	1.11~1.99
	rs2052287	chr.16 51059669	TOX3	Intron	G	0.755	0.801	0.73	0.019	0.347	1.30	1.04~1.62
	rs2287144	chr.16 51059819	TOX3	Intron	G	0.756	0.800	0.72	0.022	0.397	1.29	1.04~1.61
	rs3095620	chr.16 51068809	TOX3	Intron	T	0.880	0.913	0.016	0.020	0.376	1.44	1.06~1.96
18q22	rs2000728	chr.18 68647991	NETO1	Intron	G	0.554	0.599	0.04	0.047	0.667	1.20	1.00~1.44
27 SNPs, 1026.4 kb chr18: 68441726-69468090	rs1477490	chr.18 68910075		Intergenic	T	0.385	0.447	0.92	0.0061	0.137	1.29	1.07~1.55
	rs1477491	chr.18 68910515		Intergenic	A	0.383	0.447	0.84	0.0044	0.103	1.30	1.09~1.56

Abbreviations: P, chi-square test; P-value; Pc, 10000 iterated permutation P-value.

Pc-values smaller than 0.05 are underlined. For the 5q15 locus, 30 SNPs within a 521.8 kb interval (chr5: 19501188-20341733) were subjected to association analysis, and none of them showed a statistically significant association (P<0.05) with AN.

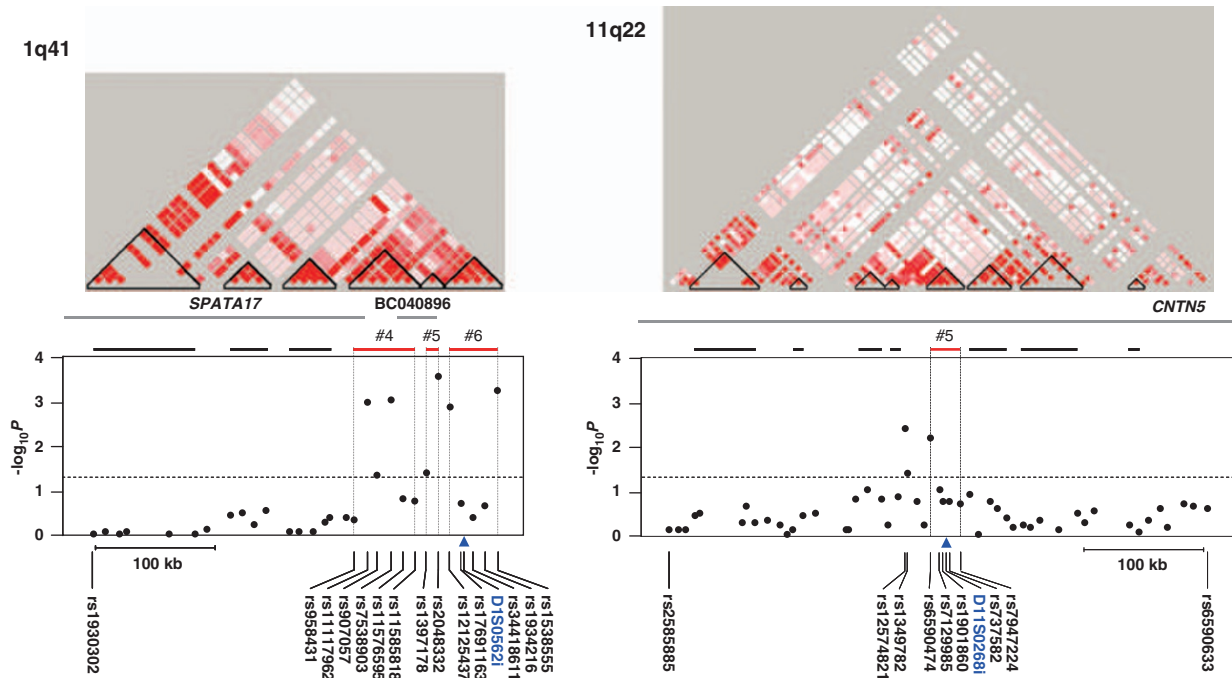


Figure 1 Single-nucleotide polymorphism (SNP) and haplotype association analyses for the 1q41 (left) and the 11q22 (right) loci. For each locus, the linkage disequilibrium (LD) plot (top), resided gene(s) (middle) and the P -value plot (bottom) are shown. In LD plots, the extent of LD between two SNPs is shown by the standard color scheme (D' /LOD) of Haploview 4.0. In P -value plots, closed dots show the minus log P -value (y axis) and the physical location (x axis) of SNPs. Minus log P -values were calculated by χ^2 tests for the genotyping data of anorexia nervosa (AN) cases ($n=331$) and controls ($n=872$). The horizontal dashed line corresponds to the P -value of 0.05. Black and red horizontal bars above the P -value plots correspond to the LD blocks defined by the confidence intervals method (Haploview 4.0). LD blocks showing statistical significance in the haplotype association analysis ($P_c < 0.05$ in Table 3) are shown by red bars. The rs numbers of the SNPs showing statistical significance ($P < 0.05$ in Table 1), the SNPs binned to the AN-associated LD blocks and the SNPs at the ends of the genomic interval are shown underneath the P -value plot. The positions of AN-associated MS markers (D1S0562i and D11S0268i) are shown by blue rectangles. The *SPATA17* gene and an uncharacterized mRNA sequence, BC040896, which are transcribed from left to right orientation, are mapped in the 337.3-kb interval between SNPs rs1930302 and rs1538555 on 1q41. For the 11q22 locus, the 474.6-kb region between SNPs rs2585885 and rs6590633, which are located in intron 2 and intron 16 of the *CNTN5* gene (NM_014361), respectively, is shown.

Table 3 LD blocks in 1q41 and 11q22 loci and haplotype association analysis

Locus & block#	Size (kb)	Start	End	Gene	# of SNPs	# of positive		Frequency (controls)	Frequency (AN)	P	P_c (permutation P-value)
						haplotypes	haplotypes ($P < 0.05$)				
1q41											
#1	85.1	rs1930302	rs11578064	<i>SPATA17</i>	6	3	0	0.676	0.676	0.99	1
#2	29.8	rs6604558	rs10495075	<i>SPATA17</i>	4	3	0	0.535	0.559	0.29	1
#3	32.9	rs11578620	rs1510262	<i>SPATA17</i>	5	4	0	0.586	0.570	0.48	1
#4	49.9	rs958431	rs11585818	<i>SPATA17/BC040896</i>	6	7	2	0.520	0.447	<u>0.0014</u>	<u>0.031</u>
#5	10.2	rs1397178	rs2048332	BC040896	2	3	2	0.548	0.461	<u>0.00010</u>	<u>0.0039</u>
#6	38.1	rs12125437	rs1538555		5	4	2	0.139	0.191	<u>0.0019</u>	<u>0.038</u>
11q22											
#1	48.0	rs2585885	rs1145408	<i>CNTN5</i>	5	5	0	0.487	0.508	0.37	1
#2	8.6	rs4754649	rs11221713	<i>CNTN5</i>	2	3	1	0.037	0.020	<u>0.032</u>	0.56
#3	22.7	rs3824932	rs10894179	<i>CNTN5</i>	3	3	0	0.654	0.623	0.16	0.99
#4	6.5	rs11221996	rs1530997	<i>CNTN5</i>	2	3	0	0.445	0.431	0.55	1
#5	20.2	rs6590474	rs7947224	<i>CNTN5</i>	5	3	2	0.730	0.655	<u>0.00030</u>	<u>0.0078</u>
#6	32.0	rs770569	rs4754665	<i>CNTN5</i>	5	5	0	0.482	0.447	0.13	1
#7	45.9	rs10750469	rs12806530	<i>CNTN5</i>	5	6	0	0.409	0.397	0.61	1
#8	7.3	rs7115626	rs6590633	<i>CNTN5</i>	2	3	0	0.798	0.793	0.81	1

Abbreviations: P , chi-square test P -value; P_c , 100 000 iterated permutation P -value. P - and P_c -values smaller than 0.05 are underlined.

When no haplotype shows statistically significant association among multiple haplotypes inferred, the frequencies, P value, and P_c value of the major haplotype are shown. When more than one haplotype show statistically significant association, the frequencies, P value, and P_c value of the most significantly associated haplotype are shown.

Table 4 Case-control association analysis for MS-SNP haplotypes

Haplotype	Frequencies		P _c
	control	case	
<i>rs17691163-D1S0562i-rs34418611-rs1934216-rs1538555 (1q41-#6)</i>			
G-6-A-T-A	0.279	0.218	<u>0.025</u>
G-7-A-T-A	0.141	0.127	0.51
G-2-A-T-G	0.098	0.153	<u>0.0065</u>
G-4-G-A-G	0.111	0.105	0.76
G-5-A-T-A	0.102	0.080	0.20
G-4-A-T-A	0.066	0.073	0.68
A-6-A-T-A	0.051	0.052	0.89
<i>rs6590474-D11S0268i-rs737582-rs7947224 (11q22-#5)</i>			
C-2-G-T	0.479	0.408	<u>0.023</u>
C-3-G-T	0.288	0.242	0.105
A-4-G-T	0.124	0.212	<u>0.00003</u>
C-5-A-C	0.050	0.062	0.377

Abbreviation: P_c (permutation P-value, n=100000).

Haplotypes whose control frequency is >=0.05 are shown.

P_c-values smaller than 0.05 and the corresponding haplotypes are underlined.

This association was comparable with those observed in the SNP-haplotype analysis in two haplotype blocks 1q41-#5 (P_c=0.0039) and 1q41-#6 (P_c=0.038) in terms of statistical significance.

11q22

A total of 66 SNPs were selected from a 699.8-kb interval surrounding the AN-associated MS marker D11S0268i. Among the 56 SNPs subjected to association analysis, 3 (rs12574821, rs1349782 and rs6590474) showed a statistically significant association (P<0.05) with the AN cohort (Table 2). These associated SNPs were found to be located in the eighth intron of the *CNTN5* gene (GenBank accession no. NM_014361). Although these three SNPs did not hold statistical significance after multiple-testing correction by permutation tests, a haplotype composed of five SNPs (rs6590474, rs7129985, rs1901860, rs737582 and rs7947224) spanning a 20.2-kb interval showed a statistically significant association with AN (P_c=0.0082) in the haplotype association analysis (11q22-#5 in Figure 1, right and Table 3). Exon 9 of *CNTN5* was included in the 20.2-kb interval.

The AN-associated MS marker, D11S0268i, was located in the AN-associated 11q22-#5 block. In this block, D11S0268i was in a strong pairwise LD (D' = 0.81–0.90) with each of the five SNPs binned to this block. As three SNPs (rs6590474, rs737582 and rs7947224) and D11S0268i were selected as tags captured through LD in the 11q22-#5 block, we further conducted an MS-SNP haplotype-based association study within the block using these four markers (rs6590474, D11S0268i, rs737582 and rs7947224). As shown in Table 4, the A-4-G-T haplotype was overrepresented in AN cases with the greatest statistical significance (P_c=0.00003). This MS-SNP haplotype contained both of the significantly associated risk alleles (A allele and 4) at SNP rs6590474 (Table 2) and D11S0268i (Supplementary Table 3).

Assessment of possible gender effects in the detected association of 1q41 and 11q22 with AN

Owing to the limited number of female individuals whose age matches with the average age of the AN cases in our control samples, we adopted a population-based control group to search for AN susceptibility loci. Therefore, although 180 female controls (control group 1, average age: 34.5 years) were genotyped in the first and second stages of MS

screening, an additional 692 control individuals enrolled in the later stages (161 individuals in the third stage of MS screening and all of 692 individuals in the SNP association analysis) consisted of male and female individuals. To assess whether the detected association of the 1q41 and 11q22 loci to AN was inflated by this partial mismatch in gender between case and control populations, we conducted a stratified analysis for 7 SNPs on 1q41 and for 3 SNPs on 11q22 (Table 2) that showed a significant association with AN. When control group 1 (180 females) and the AN cohort were subjected to association analysis, all 10 SNPs were detected to be associated (P<0.05) with AN (Supplementary Table 4). These results assure that the association of the 1q41 and 11q22 loci with AN detected in this study is not because of inflation caused by the inclusion of male individuals in the control population.

Association analysis for a BN cohort

To assess whether the genomic intervals identified to be associated with AN in this study are also involved in the genetic etiology of BN, we conducted an SNP association analysis for BN cases and controls. The 7 SNPs from the 1q41 locus and the 3 SNPs from the 11q22 locus, which showed a statistically significant association (P<0.05) with AN before multiple-testing correction, were subjected to SNP genotyping on the cohort of 125 BN cases. None of the 10 SNPs showed a statistically significant association with BN (data not shown).

DISCUSSION

We have completed a genome-wide association analysis for AN using 23 465 MS markers. To our knowledge, this is the first GWAS performed for EDs. Among the 10 candidate loci we identified, 9 are reported to be associated with AN for the first time in this study. Only one locus, D1S0016i on 1p36, overlaps with the chromosomal region of 1p33-p36 that has already been reported to show significant linkage to AN.¹²

Through an SNP association analysis for the seven selected candidate regions to narrow down genomic intervals involved with AN susceptibility, we tentatively identified a 10.2-kb genomic interval (the haplotype block 1q41-#5) located at 3'-downstream of the *SPATA17* gene as a region associated with AN. The MS-SNP haplotype-based association analysis also indicated the association of haplotype block 1q41-#6 with AN with a similar statistical significance. *SPATA17* encodes a 361 amino-acid protein that contains three highly conserved IQ motifs and is strongly expressed in the testis.³⁰ It is unknown whether the *SPATA17* protein has any physiological roles in neuronal tissues. It should be noted that the 10.2-kb critical interval coincides with the exon-intron structure of the uncharacterized mRNA sequence BC040896, which is derived from a cDNA library made from brain (adult medulla) RNA.

Another genomic region identified to be associated with AN in this study is a 20.2-kb interval (the haplotype block 11q22-#5), spanning from the eighth to the ninth intron of the *CNTN5* gene on 11q22. Furthermore, we found that one MS-SNP haplotype (A-4-G-T), which includes two AN-associated risk alleles at SNP rs6590474 and MS D11S0268i, was significantly overrepresented in AN cases. *CNTN5* encodes a member of the contactin family known to function during the formation of neuronal interactions. It is reported that the mouse line deficient of *Cntn1*, another member of the contactin family, exhibits an ataxic and anorectic phenotype.^{31,32} In human adult tissues examined using northern blot analysis, *CNTN5* has been shown to be predominantly expressed in the brain and thyroid.³³ In various regions of an adult brain examined, the gene was found to be expressed with highest levels in the occipital lobe and amygdala,

followed by the cerebral cortex, frontal lobe, thalamus and temporal lobe.³³ Although neuronal activity in the auditory system is reported to be impaired in the mouse line deficient for *Cntn5*, no anorexic phenotype has been described.³⁴

Although causative SNPs are not yet determined, we have successfully mapped genetic association with AN to at least two genomic regions on 1q41 and 11q22 and narrowed down an AN-associated genomic interval for each locus by haplotype association analysis. Further replication analysis using independent patient/control populations for AN-associated SNPs and functional analyses for the genes or for particular genomic regions in these loci will better clarify the impact of these SNPs/genes in the genetic etiology of AN. It should also be noted that additional common variants are likely to have roles in the development of AN because this study was not well powered to detect susceptible loci with relatively small genetic risks. Additional gender/age-matched cohorts consisting of much larger numbers of cases and controls need to be used to improve statistical power in MS-based genome-wide association analysis.

ACKNOWLEDGEMENTS

This study was performed under the management of the Japan Biological Informatics Consortium (JBIC) and was supported by grants from the New Energy and Industrial Technology Development Organization (NEDO). This study was also supported by Grant-in-Aid for Scientific Research on Priority Areas from MEXT, and by Grant-in-Aid for Young Scientists (B) from JSPS. We thank Karin Ohki for her technical assistance.

- Tozzi, F., Thornton, L. M., Klump, K. L., Fichter, M. M., Halmi, K. A., Kaplan, A. S. *et al*. Symptom fluctuation in eating disorders: correlates of diagnostic crossover. *Am. J. Psychiatry* **162**, 732–740 (2005).
- Nishimura, H., Komaki, G., Ando, T., Nakahara, T., Oka, T., Kawai, K. *et al*. Japanese Genetic Research Group for Eating Disorders. Psychological and weight-related characteristics of patients with anorexia nervosa-restricting type who later develop bulimia nervosa. *Biopsychosoc. Med.* **2**, 5 (2008).
- Fairburn, C. G., Harrison, P. J. Eating disorders. *Lancet* **361**, 407–416 (2003).
- Birmingham, C. L., Su, J., Hlynsky, J. A., Goldner, E. M., Gao, M. The mortality rate from anorexia nervosa. *Int. J. Eat. Disord.* **38**, 143–146 (2005).
- Keel, P. K., Mitchell, J. E. Outcome in bulimia nervosa. *Am. J. Psychiatry* **154**, 313–321 (1997).
- Lilenfeld, L. R., Kaye, W. H., Greeno, C. G., Merikangas, K. R., Plotnicov, K., Pollice, C. *et al*. A controlled family study of anorexia nervosa and bulimia nervosa: psychiatric disorders in first-degree relatives and effects of proband comorbidity. *Arch. Gen. Psychiatry* **55**, 603–610 (1998).
- Strober, M., Freeman, R., Lampert, C., Diamond, J., Kaye, W. Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. *Am. J. Psychiatry* **157**, 393–401 (2000).
- Bulik, C. M., Sullivan, P. F., Wade, T. D., Kessler, K. S. Twin studies of eating disorders: a review. *Int. J. Eat. Disord.* **27**, 1–20 (2000).
- Wade, T. D., Bulik, C. M., Neale, M., Kessler, K. S. Anorexia nervosa and major depression: shared genetic and environmental risk factors. *Am. J. Psychiatry* **157**, 469–471 (2000).
- Bulik, C. M., Slof-Op't Landt, M. C., van Furth, E. F., Sullivan, P. F. The genetics of anorexia nervosa. *Annu. Rev. Nutr.* **27**, 263–275 (2007).
- Pinheiro, A. P., Sullivan, P. F., Bacaltchuck, J., Prado-Lima, P. A., Bulik, C. M. Genetics in eating disorders: extending the boundaries of research. *Rev. Bras. Psiquiatr.* **28**, 218–225 (2006).
- Grice, D. E., Halmi, K. A., Fichter, M. M., Strober, M., Woodside, D. B., Treasure, J. T. *et al*. Evidence for a susceptibility gene for anorexia nervosa on chromosome 1. *Am. J. Hum. Genet.* **70**, 787–792 (2002).
- Devlin, B., Bacanu, S. A., Klump, K. L., Bulik, C. M., Fichter, M. M., Halmi, K. A. *et al*. Linkage analysis of anorexia nervosa incorporating behavioral covariates. *Hum. Mol. Genet.* **11**, 689–696 (2002).
- Bacanu, S. A., Bulik, C. M., Klump, K. L., Fichter, M. M., Halmi, K. A., Keel, P. *et al*. Linkage analysis of anorexia and bulimia nervosa cohorts using selected behavioral phenotypes as quantitative traits or covariates. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **139**, 61–68 (2005).
- Bergen, A. W., van den Bree, M. B., Yeager, M., Welch, R., Ganjei, J. K., Haque, K. *et al*. Candidate genes for anorexia nervosa in the 1p33-36 linkage region: serotonin 1D and delta opioid receptor loci exhibit significant association to anorexia nervosa. *Mol. Psychiatry* **8**, 397–406 (2003).
- Bulik, C. M., Devlin, B., Bacanu, S. A., Thornton, L., Klump, K. L., Fichter, M. M. *et al*. Significant linkage on chromosome 10p in families with bulimia nervosa. *Am. J. Hum. Genet.* **72**, 200–207 (2003).
- Klump, K. L., Gobrogge, K. L. A review and primer of molecular genetic studies of anorexia nervosa. *Int. J. Eat. Disord.* **37**, S43–S48 (2005).
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- Oka, A., Tamiya, G., Tomizawa, M., Ota, M., Katsuyama, Y., Makino, S. *et al*. Association analysis using refined microsatellite markers localizes a susceptibility locus for psoriasis vulgaris within a 111 kb segment telomeric to the HLA-C gene. *Hum. Mol. Genet.* **8**, 2165–2170 (1999).
- Oka, A., Hayashi, H., Tomizawa, M., Okamoto, K., Suyun, L., Hui, J. *et al*. Localization of a non-melanoma skin cancer susceptibility region within the major histocompatibility complex by association analysis using microsatellite markers. *Tissue Antigens* **61**, 203–210 (2003).
- Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K. *et al*. Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Hum. Mol. Genet.* **14**, 2305–2321 (2005).
- Kimura, T., Kobayashi, T., Munkhbat, B., Oyungerel, G., Bilegtsaikhan, T., Anar, D. *et al*. Genome-wide association analysis with selective genotyping identifies candidate loci for adult height at 8q21.13 and 15q22.33-q23 in Mongolians. *Hum. Genet.* **123**, 655–660 (2008).
- Yatsu, K., Mizuki, N., Hirawa, N., Oka, A., Itoh, N., Yamane, T. *et al*. High-resolution mapping for essential hypertension using microsatellite markers. *Hypertension* **49**, 446–452 (2007).
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. (American Psychiatric Association: Washington, DC, 1994).
- Collins, H. E., Li, H., Inda, S. E., Anderson, J., Laiho, K., Tuomilehto, J. *et al*. A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. *Hum. Genet.* **106**, 218–226 (2000).
- Purcell, S., Cherny, S. S., Sham, P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
- Barrett, J. C., Fry, B., Maller, J., Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B. *et al*. The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- Deng, Y., Hu, L. S., Lu, G. X. Expression and identification of a novel apoptosis gene Spata17 (MSRG-11) in mouse spermatogenic cells. *Acta. Biochim. Biophys. Sin. (Shanghai)* **38**, 37–45 (2006).
- Johansen, J. E., Fetissov, S. O., Bergström, U., Nilsson, I., Faÿ, C., Ranscht, B. *et al*. Evidence for hypothalamic dysregulation in mouse models of anorexia as well as in humans. *Physiol. Behav.* **92**, 278–282 (2007).
- Fetissov, S. O., Bergström, U., Johansen, J. E., Höfkelt, T., Schalling, M., Ranscht, B. Alterations of arcuate nucleus neuropeptidergic development in contactin-deficient mice: comparison with anorexia and food-deprived mice. *Eur. J. Neurosci.* **22**, 3217–3228 (2005).
- Kamei, Y., Takeda, Y., Teramoto, K., Tsutsumi, O., Taketani, Y., Watanabe, K. Human NB-2 of the contactin subgroup molecules: chromosomal localization of the gene (CNTN5) and distinct expression pattern from other subgroup members. *Genomics* **69**, 113–119 (2000).
- Li, H., Takeda, Y., Niki, H., Ogawa, J., Kobayashi, S., Kai, N. *et al*. Aberrant responses to acoustic stimuli in mice deficient for neural recognition molecule NB-2. *Eur. J. Neurosci.* **17**, 929–936 (2003).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)