

Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma

Miriam F. Moffatt^{1*}, Michael Kabesch^{2*}, Liming Liang^{3*}, Anna L. Dixon⁴, David Strachan⁵, Simon Heath⁶, Martin Depner², Andrea von Berg⁷, Albrecht Bufe⁸, Ernst Rietschel⁹, Andrea Heinzmann¹⁰, Burkard Simma¹¹, Thomas Frischer¹², Saffron A. G. Willis-Owen¹, Kenny C. C. Wong¹, Thomas Illig¹³, Christian Vogelberg¹⁴, Stephan K. Weiland¹⁵, Erika von Mutius², Gonçalo R. Abecasis³, Martin Farrall⁴, Ivo G. Gut⁶, G. Mark Lathrop⁶ & William O. C. Cookson¹

Asthma is caused by a combination of poorly understood genetic and environmental factors^{1,2}. We have systematically mapped the effects of single nucleotide polymorphisms (SNPs) on the presence of childhood onset asthma by genome-wide association. We characterized more than 317,000 SNPs in DNA from 994 patients with childhood onset asthma and 1,243 non-asthmatics, using family and case-referent panels. Here we show multiple markers on chromosome 17q21 to be strongly and reproducibly associated with childhood onset asthma in family and case-referent panels with a combined *P* value of $P < 10^{-12}$. In independent replication studies the 17q21 locus showed strong association with diagnosis of childhood asthma in 2,320 subjects from a cohort of German children ($P = 0.0003$) and in 3,301 subjects from the British 1958 Birth Cohort ($P = 0.0005$). We systematically evaluated the relationships between markers of the 17q21 locus and transcript levels of genes in Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lines from children in the asthma family panel used in our association study. The SNPs associated with childhood asthma were consistently and strongly associated ($P < 10^{-22}$) in *cis* with transcript levels of *ORMDL3*, a member of a gene family that encodes transmembrane proteins anchored in the endoplasmic reticulum³. The results indicate that genetic variants regulating *ORMDL3* expression are determinants of susceptibility to childhood asthma.

Our study design aimed to provide internal replication of potentially positive results with family and case-referent panels of subjects with childhood asthma (Fig. 1). This structure also allowed comparison for consistency between case-control and family-based tests of association. To systematically investigate the effects of polymorphisms on the transcription of positional candidate genes, we further measured global gene expression in B-cell-derived EBV-transformed lymphoblastoid cell lines (EBVL) in probands and siblings of the family panel.

The study subjects included a panel of 207 predominantly (99%) nuclear families (MRC-A). These were recruited through a proband with severe (Step 3) childhood onset asthma and contained 295 sib pairs, 11 half-sib pairs and 3 singletons (counting all possible sibs). Four-hundred and thirty-seven non-asthmatic Caucasian UK

controls (UK-C) were studied by the same protocols. We also genotyped 728 children of German origin recruited in the Multicentre Asthma Genetics in Childhood Study (MAGICS) study with physician-diagnosed asthma for comparison with 694 reference children recruited in the cross sectional International Study of Asthma and Allergies in Childhood (ISAAC) study⁴.

We genotyped all children in the primary association study with the Illumina Sentrix HumanHap300 BeadChip. After quality control and elimination of markers with low minor-allele frequencies ($< 0.5\%$), extreme Hardy–Weinberg equilibrium statistics ($\chi^2 > 25$) or low genotyping call rates ($< 95\%$), we retained 307,328 SNPs and 684 million genotypes (99.4% call rate) for analysis. Amongst the 2,236,212 common (minimum allele frequency > 0.05) SNPs in the HapMap panel, 79% were tagged with correlation $R^2 > 0.8$ and 90.9% with $R^2 > 0.5$ with our successfully typed markers.

We also typed the parents and children in the MRC-A panel with the Illumina Sentrix Human-1 Genotyping BeadChip (concentrated on genes and surrounding sequences) (Fig. 1), producing an

Subjects	Family MRC-A		Case-referent UK-C MAGICS		Total
	Parents	Children	Children	Children	
Non-asthmatic		112	437	694	1,243
Asthmatic		266	0	728	994
Total	405	378	437	1,422	2,642
Measurements					
300K Illumina	[Yellow bar]				
100K Illumina	[Green bar]				
Affy Hu133A 2.0	[Light blue bar]				
Analyses					
Transmit	[Blue bar: 100K]				
Regression model	[Green bar: 300K]				
Expression QTL	[Purple bar]				

Figure 1 | Study design. The subjects were recruited from family (MRC-A) and case-control panels (MAGICS and UK-C). All children were genotyped with the Illumina Sentrix HumanHap300 BeadChip. The children and parents in the MRC-A panel were in addition genotyped with the Illumina Sentrix Human-1 Genotyping BeadChip. Gene expression in lymphoblastoid cell lines (EBVL) was measured in the affected and unaffected children of the MRC-A panel. Replication of positive results was sought in two independent panels of subjects from the ISAAC Phase II and 1958 British Birth Cohort studies.

¹National Heart and Lung Institute, Imperial College, London SW3 6LY, UK. ²University Children's Hospital, Ludwig Maximilians University, D80337 Munich, Germany. ³Center for Statistical Genetics, Department of Biostatistics, SPH II, Ann Arbor, Michigan 48109-2029, USA. ⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁵Division of Community Health Science, St George's, University of London, London SW17 0RE, UK. ⁶Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique, 91057 Evry, France. ⁷Research Institute for the Prevention of Allergic Diseases, Children's Department, Marien-Hospital, D46483 Wesel, Germany. ⁸Department of Experimental Pneumology, Ruhr-University, D44789 Bochum, Germany. ⁹University Children's Hospital, University of Cologne, D50924 Cologne, Germany. ¹⁰University Children's Hospital, Albert Ludwigs University, D79106 Freiburg, Germany. ¹¹Children's Department, Feldkirch Hospital, A6800 Feldkirch, Austria. ¹²University Children's Hospital Vienna, A1090 Vienna, Austria. ¹³Institute of Epidemiology, GSF-Research Centre for Environment and Health, D85764 Neuherberg, Germany. ¹⁴University Children's Hospital, Technical University Dresden, D01307 Dresden, Germany. ¹⁵Institute of Epidemiology, Ulm University, D89081 Germany.

*These authors contributed equally to this work.

additional 91,293 SNPs with 36.0% average heterozygosity and 89,815,992 genotypes (99.0% call rate). We found only 0.412 mendelian errors per SNP: these were excluded from subsequent analyses.

We tested for association of the 300K panel to childhood onset asthma in the combined primary data set of 994 asthmatics and 1,243 non-asthmatics (Fig. 2). We calculated the 1% false-discovery rate (FDR) threshold⁵ to be $P \leq 6.8 \times 10^{-7}$ and the 5% FDR to be $P \leq 5.0 \times 10^{-6}$. We tested for population stratification in the combined data set, finding a genomic control parameter of 1.07. This minor degree of background stratification reflected small differences in allele frequencies within and between the two nationally defined European populations that contributed to the study.

We further tested the effects of stratification on our top results by analysing the 34 SNPs that surpassed the 5% FDR threshold with 102 randomly selected SNPs as covariates in a backward stepwise logistic regression procedure⁶. The results of the stepwise logistic regression analysis showed only modest reduction of the support for association for 31 of the SNPs. Of 34 SNPs, 16 continued to surpass the 5% FDR and 12 of 20 SNPs remained above the 1% FDR threshold (Supplementary Table 1).

Strikingly, 7 of the 12 markers still below the 1% FDR threshold mapped to a 112 kb interval on chromosome 17q21. Several other markers in this interval also showed strong evidence of association (Supplementary Table 2, Fig. 2 and Fig. 3a). The SNP with the strongest evidence of association within the interval was *rs7216389* (uncorrected $P = 9 \times 10^{-11}$). With the exception of the 17q21 locus, none of the markers below the 5% FDR, after controlling for stratification, were within 1 Mb of each other (Supplementary Table 1).

The patterns of association for the chromosome 17q21 markers were similar and significant in both the UK family panel and the German case-referent panel (Table 1 and Supplementary Table 2). There was no evidence of heterogeneity of the association, or of significant allele frequency differences in cases or controls from the UK and Germany.

We selected 27 markers from the NCBI database of genetic variation (dbSNP) that were within or adjacent to the strongly associated interval for genotyping. These exhibited similar patterns and strengths of association as the genome-wide association (GWA) markers (Supplementary Table 2).

SNPs from the region that we had typed in whole families in the MRC-A panel showed significant associations in a family-based association test (Supplementary Table 3). The most strongly associated marker was *rs8067378* (248 informative transmissions, $P = 3 \times 10^{-6}$; odds ratio, 1.84; 95% confidence interval, 1.43–2.42). The pattern of alleles transmitted in excess to affected offspring was consistent with the case-referent association results, further indicating that the association was robust to population stratification.

The trait-associated markers (with $P < 10^{-6}$ in the combined data set) from chromosome 17q21 fall within a 206.5 kb core containing

three consecutive haplotype blocks (linkage disequilibrium coefficient (D') > 0.94 for adjacent markers, $D' < 0.82$ between blocks), as well as one 'isolated' marker in moderate ($D' = 0.7$) linkage disequilibrium with an adjacent marker (Supplementary Fig. 1). Low linkage disequilibrium between markers that showed strong association with the disease trait indicated that multiple variants may independently confer disease susceptibility. We evaluated this hypothesis statistically in a forward stepwise regression potentially incorporating all the genotyped markers between 34.5 and 36.0 Mb on chromosome 17. This identified three SNPs (*rs7216389*, *rs11650680* and *rs3859192*), which jointly showed strong association to childhood asthma ($P < 10^{-12}$) and contributed statistically independent significant effects (Supplementary Table 4). This result is consistent with the possibility that more than one functional SNP underlies the locus or (less likely given the SNP density across this region) the presence of a single functional SNP in incomplete linkage disequilibrium with the typed markers. Further fine-mapping studies will be required to resolve these alternatives.

We further examined the impact of the chromosome 17q21 locus on childhood asthma with the aid of 200 asthmatic cases and 2,120 non-asthmatic controls from within the cross-sectional study population of ISAAC Phase II, recruited in Dresden and Munich. The corresponding DNAs were genotyped for a series of nine markers from across the locus that had shown evidence of association in the GWA (Table 1). These showed the same trends as observed in the GWA samples (odds ratios ranging from 1.52 to 1.11), and most were significantly associated with disease with $P < 0.001$ (4 out of 9 markers) or $P < 0.01$ (5 out of 9 markers).

One chromosome 17q21 SNP associated with childhood asthma in the GWA analysis (*rs3894194*, Table 1) had been previously examined in 3,301 subjects from the UK 1958 birth cohort. The genotype data for this marker were deposited for public use in the database maintained by the investigators responsible for the cohort study (<http://www.b58cgene.sgu.ac.uk/>). Although data were not available for the other disease-associated chromosome 17q21 markers, we were able to confirm association to childhood asthma with *rs3894194*, which is in linkage disequilibrium with the other markers at the locus (linkage disequilibrium block 3 in Supplementary Fig. 1). Restricting analyses to cohort members of Caucasian ethnicity, we found that the 398 cases recalling 'asthma ever' at age 42, showed a significant association (odds ratio, 1.21, 95% confidence interval, 1.04–1.40, $P = 0.012$). Ninety-three individuals were reported to have 'asthma attacks' in the first seven years of life (that is during 1958 to 1965), and these were strongly associated to *rs3894194* (odds ratio = 1.68, 95% confidence interval, 1.25–2.26, $P = 0.0005$).

The disease-associated chromosome 17q21 markers had consistent odds ratios and directions of association effects in the GWA panels and the two replication cohorts ($P = 0.19$ for heterogeneity)

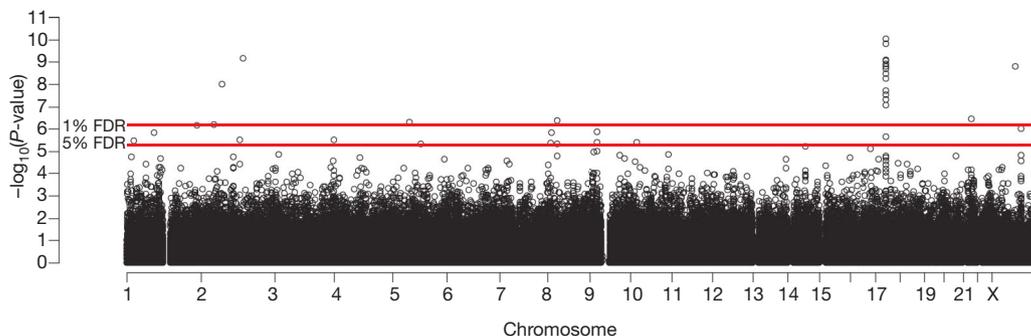


Figure 2 | Genome-wide association of 317,447 SNPs and asthma in 994 asthmatic children and 1,243 non-asthmatic children. Position in the genome, divided by chromosome, is depicted along the x axis. Strength of association is shown on the y axis. The result for each individual marker is

depicted as a black circle. The genome-wide thresholds for 1% and 5% false discovery rates (FDR) are shown as horizontal red lines. Numerous markers on chromosome 17q21 show association to asthma above the 1% FDR threshold in the region of maximum association.

Table 1 | Association of asthma and *ORMDL3* to chromosome 17q21 in combined and replication panels

Marker	Location (Mb)	Primary GWA study			eQTL -log ₁₀ (P) [<i>ORMDL3</i>] MRC-A	Ref. allele	Alt. allele	ISAAC II replication cohort			-log ₁₀ (P) asthma
		MRC-A	MAGICS	Combined				Odds ratio	Lower 95% CI	Upper 95% CI	
rs9303277	35.230	4.4	5.0	8.8	21.9	C	T	1.41	1.14	1.76	2.8
rs11557467	35.282	4.6	5.2	9.1	22.5	G	T	1.45	1.16	1.82	3.1
rs8067378	35.305	4.3	5.3	9.0	22.7						
rs2290400	35.320	4.0	6.4	9.8	22.4	A	G	1.47	1.18	1.82	3.2
rs7216389	35.323	4.3	6.4	10.0	22.4	T	C	1.45	1.17	1.81	3.1
rs4795405	35.342	4.1	5.3	8.7	14.6	C	T	1.52	1.20	1.89	3.6
rs8079416	35.346	2.4	6.5	8.2	10.9	C	T	1.30	1.05	1.61	1.9
rs4795408	35.361	2.1	6.1	7.5	10.3	A	G	1.25	1.01	1.54	1.4
rs3894194	35.376	1.9	6.6	7.7	11.0	T	C	1.22	0.98	1.51	1.2
rs3859192	35.382	2.0	6.0	7.3	3.5	T	C	1.11	0.89	1.37	0.4

SNPs are from the Illumina 300K panel. Results were calculated with logistic regression models. Association to *ORMDL3* transcript abundance is shown for comparison and is based on the family panel only. Replication results are shown for the ISAAC II population. CI, confidence interval; Ref., reference; Alt., alternative; eQTL, expression quantitative trait locus.

FDR threshold in either of the individual collections or the combined samples. Several other loci (Supplementary Table 1) passed a less stringent 5% FDR threshold after adjustment for population stratification, indicating that other susceptibility loci for childhood asthma could be present, perhaps having smaller effects than those found on chromosome 17q21. These results will be explored in large scale replication studies in multiple centres and population samples (the GABRIEL project: <http://www.gabriel-fp6.org/project>).

METHODS SUMMARY

Children and their parents from the UK panels were administered a standard questionnaire⁹ by a nurse or doctor. Asthma was defined as a positive response to the question "Has your doctor ever told you that you have asthma?" Proband had Step 3 asthma or worse according to British Thoracic Society guidelines¹⁰. Siblings were included regardless of asthma status¹¹. Asthma cases from the MAGICS were diagnosed by a paediatric pulmonologist or allergologist on the basis of clinical examination, history and objective tests of lung function. The transformation of peripheral blood lymphocytes in all children in the MRC-A panel was carried out by the ECACC (<http://www.ecacc.org.uk>). Microarray hybridization of EBVL RNA to the U133 Plus 2.0 GeneChips (Affymetrix) was under standard conditions. Whole-genome genotyping was carried out using Illumina Sentrix Human-1 Genotyping BeadChip¹² and Sentrix HumanHap300 Genotyping BeadChips¹³ (Illumina, San Diego). Genotyping of additional markers was performed on an ABI7900HT Sequence Detection System using TaqMan probes (Applied Biosystems, Foster City, California). German replication samples were genotyped using matrix-assisted laser desorption/ionization-time of flight (<http://www.sequenom.com>)¹⁴. Logistic regression models with robust sandwich estimation of the variance¹⁵ implemented in the Stata *logit* function were used to detect association to asthma in the combined panels. The TRANSMIT program¹⁶ was used to analyse nuclear family data (including parental genotypes). The false-discovery rate (FDR) method⁵ was used to assess the overall statistical significance of the genome-wide association results, and FDR thresholds were calculated by applying the QVALUE (<http://faculty.washington.edu/~jstorey/qvalue/>) package¹⁷. Gene expression data were normalized with the RMA package^{18,19} to remove any technical or spurious background variation. An inverse normalization transformation step was also applied to each trait to avoid any outliers. Association analysis was applied with Merlin (FASTASSOC option)²⁰, after probabilistically inferring missing genotypes²¹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April; accepted 14 June 2007.

Published online 4 July 2007.

1. Cookson, W. The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nature Rev. Immunol.* **4**, 978–988 (2004).
2. Ober, C. & Hoffjan, S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun.* **7**, 95–100 (2006).
3. Hjelmqvist, L. *et al.* *ORMDL* proteins are a conserved new family of endoplasmic reticulum membrane proteins. *Genome Biol.* **3**, RESEARCH0027 (2002).
4. Weiland, S. K. *et al.* Phase II of the International Study of Asthma and Allergies in Childhood (ISAAC II): rationale and methods. *Eur. Respir. J.* **24**, 406–412 (2004).
5. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* **57**, 289–300 (1995).

6. Setakis, E., Stirnadel, H. & Balding, D. J. Logistic regression protects against population structure in genetic association studies. *Genome Res.* **16**, 290–296 (2006).
7. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
8. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
9. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, November 1986. *Am. Rev. Respir. Dis.* **136**, 225–244 (1987).
10. British guideline on the management of asthma. *Thorax* **58** (Suppl 1), i1–i94 (2003).
11. Abecasis, G., Cardon, L. & Cookson, W. Selection strategies for disequilibrium mapping of quantitative traits in nuclear families. *Am. J. Hum. Genet.* **65**, A245 (1999).
12. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genet.* **37**, 549–554 (2005).
13. Steemers, F. J. *et al.* Whole-genome genotyping with the single-base extension assay. *Nature Methods* **3**, 31–33 (2006).
14. Buetow, K. H. *et al.* High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl Acad. Sci. USA* **98**, 581–584 (2001).
15. Williams, R. L. A note on robust variance estimation for cluster-correlated data. *Biometrics* **56**, 645–646 (2000).
16. Clayton, D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am. J. Hum. Genet.* **65**, 1170–1177 (1999).
17. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
18. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
19. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
20. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2002).
21. Burdick, J. T., Chen, W. M., Abecasis, G. R. & Cheung, V. G. *In silico* method for inferring genotypes in pedigrees. *Nature Genet.* **38**, 1002–1004 (2006).
22. Abecasis, G. R. & Cookson, W. O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The study was funded by the Wellcome Trust, the Medical Research Council, the French Ministry of Higher Education and Research, the German Ministry of education and research (BMBF), the national genome research network (NGFN), the National Institutes of Health (NHGRI and NHLBI; G.R.A.), and the European Commission as part of GABRIEL (a multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community). We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council and the Wellcome Trust. We thank J. Todd for genotyping rs3894194 in the 1958 British Birth cohort.

Author Information Microarray and chromosome 17 genotyping data have been deposited in the GEO database, with accession number GSE8052. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.O.C.C. (w.cookson@imperial.ac.uk).

METHODS

Subjects. Children and their parents from the UK panels were recruited as part of the MRC UK National family collection and were administered a standard questionnaire (based on the American Thoracic Society and International Study of Asthma and Allergies in Childhood (ISAAC) questionnaires⁹) by a nurse practitioner or a doctor. Asthma was defined as a positive response to the question “Has your doctor ever told you that you have asthma?” Proband had Step 3 asthma or worse according to the British Thoracic Society guidelines (high-dose inhaled steroids, or low-dose inhaled steroids and a long-acting β -agonist)¹⁰. Siblings were included regardless of asthma status¹¹. Asthma cases from the Multicentre Asthma Genetics in Childhood Study (MAGICS) were diagnosed by a paediatric pulmonologist or allergologist on the basis of clinical examination, case history and objective tests of lung function. Asthmatics (mean age 10.95 yr) were recruited from 7 centres located in Germany and Austria (Wesel, Bochum Cologne, Freiberg, Munich, Feldkirch and Vienna), and as a reference, 800 German children (mean age 9.62 yr) from Dresden ($n = 400$) and Munich ($n = 400$) were randomly drawn from all German children with DNA available in the cross sectional ISAAC Phase II study⁴. Further cases and controls for replication were drawn from the German ISAAC population in which asthma was diagnosed using standardized questionnaires and validated by lung function and bronchial hyper-responsiveness testing⁴. All study methods were approved by the appropriate ethics committees.

EBV. The transformation of the Peripheral Blood Lymphocytes (PBL) in all children in the MRC-A panel was carried out by the ECACC (<http://www.ecacc.org.uk>). Previously transformed cryo-preserved EBV cell lines were grown as 500 ml roller cultures. Once log phase had been obtained, cells were pelleted, media discarded and a mixture of RLT buffer and β -mercaptoethanol added. Pellets were vortexed to ensure thorough re-suspension, after which they were frozen at -70°C and stored at -80°C . RNA was extracted in batches after cell homogenization using RNeasy Maxi Kits (Quiagen), and quality and quantity assessed.

Microarray hybridization. RNA (10 μg) was used to synthesize double-stranded cDNA using the One-cycle cDNA synthesis kit (Affymetrix). Using the cDNA as a template, *in vitro* transcription of cRNA was carried out using the IVT kit (Affymetrix), following the manufacturer's protocol. A hybridization cocktail was made according to protocol, using 15 μg of labelled, fragmented cRNA, and hybridized to U133 Plus 2.0 GeneChips (Affymetrix) for 16 h at 45°C in a rotating oven. GeneChips were washed and stained according to manufacturer's protocols and scanned on a high-resolution scanner (Affymetrix).

Genotyping. Whole-genome genotyping (WGGT) was carried out using Illumina Sentrix Human-1 Genotyping BeadChip¹² and Sentrix HumanHap300 Genotyping BeadChips¹³ (Illumina, San Diego), according to the manufacturer's instructions in a BeadLab with full automation at the Centre National de Genotypage. All DNA samples were subjected to rigorous quality control to check for fragmentation and amplification. Twenty microlitres of DNA at a concentration of $50\text{ ng }\mu\text{l}^{-1}$ was used for each array. DNA samples were tracked using a Laboratory Information Management System. The HumanHap300 Genotyping BeadChip was used with an Illumina LIMS, whereas the Sentrix Human-1 Genotyping BeadChip was tracked through the Illumina process by hand. Groups of 24 samples were batched. Five percent of the samples were selected from different batches, re-genotyped and the results compared to the original data. No sample discrepancies were detected. Raw data were analysed using GTS Image and extracted for statistical analysis. Genotyping of additional markers on chromosome 17q21 was performed on an ABI7900HT Sequence Detection System using TaqMan probes (Applied Biosystems, Foster City, California). German replication samples were genotyped using matrix-assisted laser desorption/ionization time-of flight (MALDI-TOF) mass spectrometry (<http://www.sequenom.com>)¹⁴. Primer extension products were analysed by a MassARRAY mass spectrometer (<http://www.bdal.de>) and resulting mass spectra were analysed using the SpectroTYPER RT 2.0 software.

Association testing. Tests of Hardy–Weinberg equilibrium were performed in cases and controls using the genhw procedure (<http://www.biostat-resources.com/stata/>) and Stata version 9.2, and SNPs showing Hardy–Weinberg disequilibrium in controls ($\chi^2 > 25$) were excluded. As the data comprised a mixture of unrelated and related cases and controls, we used logistic regression models with robust sandwich estimation of the variance¹⁵ as implemented in the Stata logit function to model clustering of siblings' genotypes. Simulations using the MRC-A family structures (data available on request) confirmed that this method appropriately controls the Type I error. Heterogeneity of association between the two main strata (UK and Germany) was assessed by a weighted linear combination test using the results of an additive-effects-only regression analysis within each stratum. X-linked markers were analysed by fitting an additive-effects-only logit model that equates the risks of male hemizygotes with female homozygotes. The TRANSMIT program¹⁶ was used to analyse nuclear family

data (including parental genotypes), using the sandwich variance estimation option to robustly incorporate information from multiple affected siblings; confidence intervals for odds ratio estimates were computed as described. The false-discovery rate (FDR) method⁵ was used to assess the overall statistical significance of the genome-wide association results, taking into account the multiple hypothesis testing implications inherent in the analysis of more than 300K SNPs. The FDR thresholds were calculated by applying the QVALUE (<http://faculty.washington.edu/~jstorey/qvalue/>) software package¹⁷.

Association to transcript abundances. Data from the gene expression experiment were normalized together using the RMA package^{18,19} to remove any technical or spurious background variation. An inverse normalization transformation step was also applied to each trait to avoid any outliers. Association analysis was applied with Merlin (FASTASSOC option)²⁰. We estimated an additive effect for each SNP and tested its significance using a score test that adjusts for familiarity and takes into account uncertainty in the inference of missing genotypes. In the absence of a positive genomic control test, we did not adjust for stratification. We probabilistically inferred missing genotypes²¹ and adjusted for familiarity, but not for linkage signal.