LETTERS

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

Miriam F. Moffatt[1]*, Michael Kabesch[2]*, Liming Liang[3]*, Anna L. Dixon[4], David Strachan[5], Simon Heath[6], Martin Depner[2], Andrea von Berg[7], Albrecht Bufe[8], Ernst Rietschel[9], Andrea Heinzmann[10], Burkard Simma[11], Thomas Frischer[12], Saffron A. G. Willis-Owen[1], Kenny C. C. Wong[1], Thomas Illig[13], Christian Vogelberg[14], Stephan K. Weiland[15], Erika von Mutius[2], Gonçalo R. Abecasis[3], Martin Farrall[4], Ivo G. Gut[6], G. Mark Lathrop[6] & William O. C. Cookson[1]

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 reticulum[3]. 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) study[4].

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
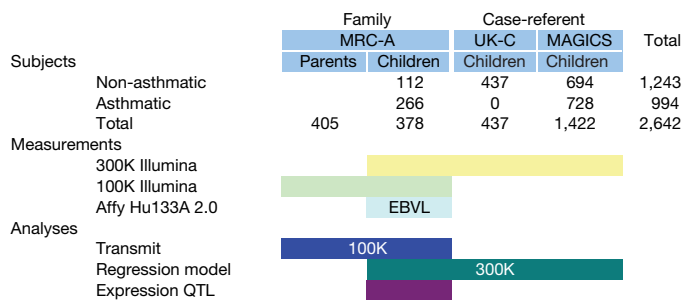

Subjects	Family		Case-referent		Total
	MRC-A		UK-C	MAGICS	
	Parents	Children	Children	Children	
Non-asthmatic		112	437	694	1,243
Asthmatic		266	0	728	994
Total	405	378	437	1,422	2,642
Measurements					
300K Illumina					
100K Illumina					
Affy Hu133A 2.0		EBVL			
Analyses					
Transmit		100K			
Regression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

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.

[1]National Heart and Lung Institute, Imperial College, London SW3 6LY, UK. [2]University Children's Hospital, Ludwig Maximilians University, D80337 Munich, Germany. [3]Center for Statistical Genetics, Department of Biostatistics, SPH II, Ann Arbor, Michigan 48109-2029, USA. [4]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [5]Division of Community Health Science, St George's, University of London, London SW17 0RE, UK. [6]Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique, 91057 Evry, France. [7]Research Institute for the Prevention of Allergic Diseases, Children's Department, Marien-Hospital, D46483 Wesel, Germany. [8]Department of Experimental Pneumology, Ruhr-University, D44789 Bochum, Germany. [9]University Children's Hospital, University of Cologne, D50924 Cologne, Germany. [10]University Children's Hospital, Albert Ludwigs University, D79106 Freiburg, Germany. [11]Children's Department, Feldkirch Hospital, A6800 Feldkirch, Austria. [12]University Children's Hospital Vienna, A1090 Vienna, Austria. [13]Institute of Epidemiology, GSF-Research Centre for Environment and Health, D85764 Neuherberg, Germany. [14]University Children's Hospital, Technical University Dresden, D01307 Dresden, Germany. [15]Institute 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) threshold[5] to be $P \le 6.8 \times 10^{-7}$ and the 5% FDR to be $P \le 5.0 \times 10^{-6}$. 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 procedure[6]. 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.sgul.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)
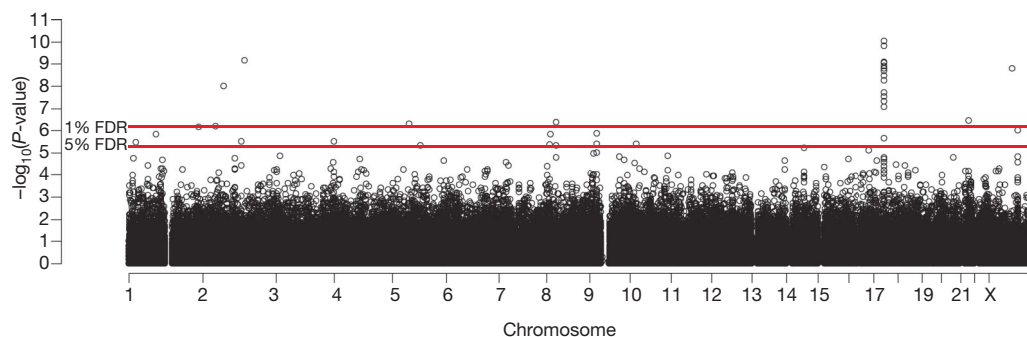

Figure 2 | Genome-wide association of 317,447 SNPs and asthma in 994 asthmatic children and 1,243 non-asthmatic children. Position in the genome, divided by chromosome, is depicted along the *x* axis. Strength of association is shown on the *y* axis. The result for each individual marker is depicted 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), suggesting that we have identified a robust risk factor for childhood asthma.

Variation in gene transcription is an important mechanism in mediating susceptibility to asthma and other diseases, and the transcript abundances of genes may be directly modified by poly-morphisms in regulatory elements[7,8]. We therefore measured global gene expression in EBVL from children in the MRC-A panel (Fig. 1). These same subjects had been typed with the Illumina Sentrix HumanHap300 and Sentrix Human-1 Genotyping BeadChips (Fig. 1). EBVLs represent the B-cell lineage, and are consequently of direct relevance to asthma. Cells were harvested at log-phase in the first growth after EBV transformation. Global transcript abundance was measured with Affymetrix HG-U133 Plus 2.0 chips. We used quantile normalization after Robust Multi-Array Average (RMA) to enforce normality and reduce outlier leverage. A complete description of the global results will be presented in another paper.

Expression data were available for 14 of 19 annotated genes in the region from 35.0 to 35.5 Mb on chromosome 17 (that is, within or near the 206 kb region of association on 17q21.1) (Fig. 3d). We found that transcripts in one gene, *ORMDL3,* were strongly ($P < 10^{-22}$ for *rs7216389*) and consistently positively associated to exactly the same SNPs from the Illumina Sentrix HumanHap300 BeadChip as childhood asthma (Fig. 3c). The disease-associated markers accounted for 29.5% of the variance of expression. No other markers were significantly correlated to *ORMDL3* expression after adjustment for genome-wide multiple testing. After accounting for the effects of the disease-associated markers, the residual heritability of the *ORMDL3* expression was not significant ($P = 0.29$ compared to $P = 0.0009$ before adjustment).

These findings suggested that the 17q21 disease-associated locus is the principal genetic determinant of *ORMDL3* expression. None of the other transcripts from the region or elsewhere in the genome showed a significant relationship to the disease-associated markers in our data set. Despite the strength of these effects, the presence of several SNPs independently associated to asthma nevertheless makes it possible that *ORMDL3* may not be the exclusive determinant of the disease susceptibility at this locus.

In the subset of individuals for whom expression data are available, the T nucleotide allele at *rs7216389* (the marker most strongly associated with disease in the combined GWA analysis) has a frequency of 62% amongst asthmatics compared to 52% in non-asthmatics ($P = 0.005$ in this sample). The additive effect of this allele corresponds to a change of 0.78 standard deviation units in *ORMDL3* expression ($P < 10^{-22}$). We saw the expected increase of ~0.064 standard deviation units in *ORMDL3* expression among asthmatics, but this was not significant given the sample size. Patterns of *ORMDL3* expression by genotype in asthmatics and non-asthmatics are shown in Supplementary Fig. 2.

ORMDL3 is the third member of a novel class of genes of unknown function that encode transmembrane proteins anchored in the endoplasmic reticulum (ER)[3]. We examined multiple tissue complementary (MTC) DNA panels by PCR with reverse transcription (RT–PCR), and found *ORMDL3* to be expressed in many tissues, particularly liver and peripheral blood lymphocytes (Fig. 3g). The SNPs showing the strongest association to asthma and *ORMDL3* transcript abundances are contained within an island of linkage disequilibrium between 35.2 and 35.4 Mb on chromosome 17q21 (Fig. 3b and 3c). The one-lod support unit for SNPs showing maximum association to *ORMDL3* levels lies within the first intron of the neighbouring *GSDML* gene. This non-coding sequence shows significant homology between species (Fig. 3e and 3f), and contains an element with high homology to the pro-inflammatory transcription factor C/EBPb (transcription factor score, 86.8; http://www.cbrc.jp/research/db/TFSEARCH.html).

Genomic regions other than the 17q21 locus did not contain multiple markers with significant evidence of association at the 1%
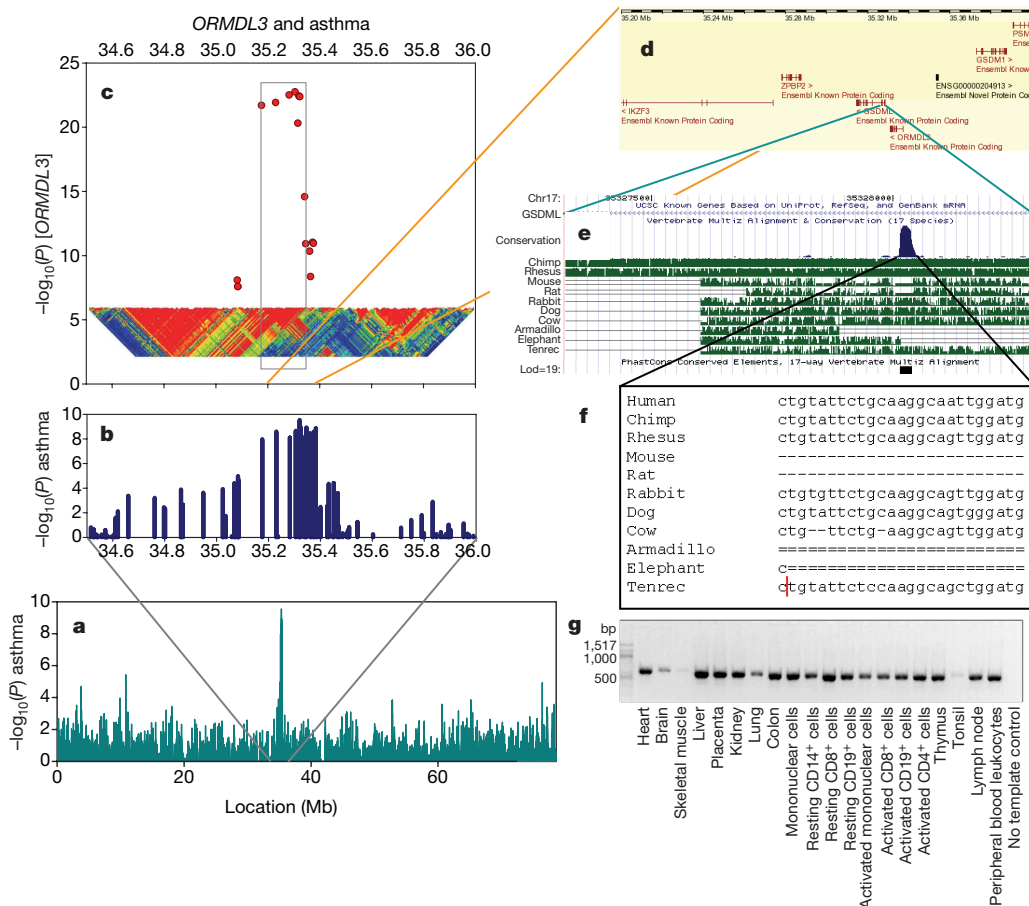

Figure 3 | Association to asthma and transcript abundances of *ORMDL3* on chromosome 17q21. a, Mapping of association to asthma on chromosome 17. **b**, Detail of association to SNPs on chromosome 17q21. **c**, Association to *ORMDL3* transcript abundance with the same markers. A GOLD plot[22] of linkage disequilibrium between markers is also shown, with red indicating high linkage disequilibrium and blue denoting low. The central island of linkage disequilibrium, which contains maximum association to *ORMDL3* and asthma, is contained within the grey rectangle. **d**, Genes contained within the associated interval. **e**, Homology plot from the region of maximum association. **f**, Sequence homology from intron I of *GSDML*. **g**, RT–PCR (34 cycles) of *ORMDL3* cDNA from representative tissues (Clontech).

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

Marker	Location (Mb)	Primary GWA study $-\log_{10}(P)$ asthma MRC-A	Primary GWA study $-\log_{10}(P)$ asthma MAGICS	Primary GWA study $-\log_{10}(P)$ asthma Combined	eQTL $-\log_{10}(P)$ [*ORMDL3*] MRC-A	Ref. allele	Alt. allele	ISAAC II replication cohort Odds ratio	ISAAC II replication cohort Lower 95% CI	ISAAC II replication cohort Upper 95% CI	$-\log_{10}(P)$ asthma
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 questionnaire[9] 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?" Probands had Step 3 asthma or worse according to British Thoracic Society guidelines[10]. Siblings were included regardless of asthma status[11]. 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 BeadChip[12] and Sentrix HumanHap300 Genotyping BeadChips[13] (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 variance[15] implemented in the Stata *logit* function were used to detect association to asthma in the combined panels. The TRANSMIT program[16] was used to analyse nuclear family data (including parental genotypes). The false-discovery rate (FDR) method[5] 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/) package[17]. 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 genotypes[21].

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).

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) questionnaires[9]) 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?" Probands 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 status[11]. 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 study[4]. 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 testing[4]. 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 BeadChip[12] and Sentrix HumanHap300 Genotyping BeadChips[13] (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 ng μl⁻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 variance[15] 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 program[16] 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) method[5] 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 package[17].

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 familiality 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 genotypes[21] and adjusted for familiality, but not for linkage signal.