nature genetics

Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma

Youming Zhang[1], Nicholas I Leaves[1,6], Gavin G Anderson[1,6], Chris P Ponting[2], John Broxholme[1], Richard Holt[1], Pauline Edser[1], Sumit Bhattacharyya[1], Andy Dunham[3], Ian M Adcock[4], Louise Pulleyn[4], Peter J Barnes[4], John I Harper[5], Gonçalo Abecasis[1], Lon Cardon[1], Melanie White[1], John Burton[3], Lucy Matthews[3], Richard Mott[1], Mark Ross[3], Roger Cox[1], Miriam F Moffatt[1] & William O C M Cookson[1]

Atopic or immunoglobulin E (IgE)-mediated diseases include the common disorders of asthma, atopic dermatitis and allergic rhinitis[1]. Chromosome 13q14 shows consistent linkage to atopy and the total serum IgE concentration[2–6]. We previously identified association between total serum IgE levels and a novel 13q14 microsatellite (*USAT24G1*; ref. 7) and have now localized the underlying quantitative-trait locus (QTL) in a comprehensive single-nucleotide polymorphism (SNP) map. We found replicated association to IgE levels that was attributed to several alleles in a single gene, *PHF11*. We also found association with these variants to severe clinical asthma. The gene product (PHF11) contains two PHD zinc fingers and probably regulates transcription. Distinctive splice variants were expressed in immune tissues and cells.

Asthma may be recognized by standard questionnaires or by physician diagnosis[3,8]. Atopy is detected by skin prick tests or by measurement of specific serum IgE titers against allergens or elevation of the total serum IgE concentration. The genetic analysis of quantitative rather than categorical traits has greater power to detect linkage and association[9]. The total serum IgE has a heritability of 40–50% (refs. 8,10). It is measured by standardized protocols and is log-normally distributed with well defined effects of age and sex. When elevated, it has a close

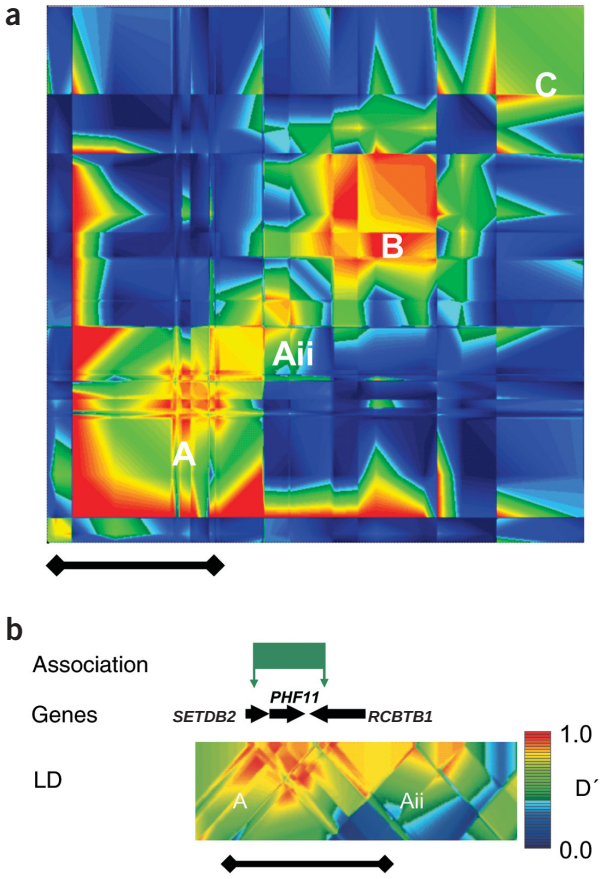


Figure 1 LD map of the locus associated with atopy. (**a**) A GOLD plot[16] of color-coded pair-wise disequilibrium statistics (D′) between markers. The locus extends from the bottom left of the figure to the top right. Red and yellow indicate areas of strong LD. LD is approximately divided into three major islands (A, B and C) and one minor island (Aii).The scale bar at the bottom indicates a distance of 200 kb. (**b**) Detail of LD around the *PHF11* gene complex. Association to IgE levels is shown above the figure. Genes are shown as black arrows, pointing in the direction of transcription. The scale bar at the bottom indicates a distance of 100 kb.

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [2]Medical Research Council Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, UK. [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [4]Department of Thoracic Medicine, National Heart and Lung Institute, London, UK. [5]Great Ormond Street Hospital and the Institute for Child Health, London, UK. [6]These authors contributed equally to this work. Correspondence should be addressed to W.O.C.M.C. (e-mail: william.cookson@ndm.ox.ac.uk).

Table 1 Associations between LnIgE and DNA polymorphism

Marker	Position	Allele frequency	*P* (QTDT) LnIgE
b13_1	–5527	0.28	
b15_1	3706	0.53	
b10_1	22872	0.72	
b11_2	71508	0.14	
d8ex7	134465	0.30	0.02
b2_2	138153	0.28	
b2_1	138191	0.72	
d8ex10	140942	0.74	
b1_2	141131	0.54	0.008
b1_1	141478	0.42	0.002
d8in12	146172	0.72	
d8in13	148684	0.27	
GGGCn	154089	0.18	0.06
b7_3	157408	0.26	0.03
b7_2	158597	0.66	0.009
b7_1	158601	0.66	0.004
ren_in1	159774	0.32	0.018
154016_2R	164692	0.33	0.016
ren_in2	170987	0.30	0.048
ren_4bp	175973	0.61	0.015
b4_1	178455	0.74	
b4_2	178972	0.43	0.002
154016_in5r	179288	0.45	0.004
b4_3	179798	0.40	0.003
154016_5S	179972	0.30	0.04
ren_in7	181824	0.18	0.004
b5_1	185499	0.56	0.003
b5_2	186064	0.39	0.0005
b5_3	187062	0.55	0.001
b6_1	190800	0.45	0.004
44593_15	191057	0.45	0.002
b6_3	191749	0.55	0.002
b6_4	192097	0.45	0.005
b33_2	239484	0.16	0.008
b33_1	240010	0.10	
b32_2	241987	0.95	
b32_1	242784	0.34	
5133822_2	278252	0.56	
b43_1	282003	0.71	0.02
b43_2	282243	0.36	
1895799_1	327886	0.22	
143317_1	332711	0.26	
b38_3	359360	0.30	
b38_1	360186	0.69	
626789_3	379960	0.91	
b35_1	434774	0.32	
b37_1	449093	0.72	
b37_2	449664	0.04	
b17_1	478470	0.65	
b28_1	516429	0.39	
b28_2	516829	0.90	
b26_2	615612	0.22	
b26_1	615975	0.06	

SETDB2 extends from marker d8ex7 to marker d8in13; *PHF11* extends from marker GGGCn to marker b5_3 and *RCBTB1* extends from marker b6_1 to marker b32_1. Markers used to test for confirmation are shown in bold. The position is given in bp, with position 1 being nucleotide chr13:49306000 from the UCSC golden path, 5 April 2002. Allele frequencies are for allele *1.

relationship to asthma[11]. We have consequently used it as a quantitative trait to map susceptibility genes for atopy and asthma.

After confirming linkage of atopy and related phenotypes to chromosome 13q14 (refs. 2–6), we made a saturation map that indicated the locus associated with atopy was within a 7.5-cM region centering on *D13S161* (ref. 12). We found association between the total serum IgE and the microsatellite *USAT24GI* (ref. 7). The limit of detection of linkage disequilibrium (LD) between a disease and a marker approximates a D′ of 0.33 (ref. 13). Approximately 80% of marker pairs at this locus showed a D′ < 0.33 by 100 kb[13], suggesting that the gene associated with atopy was within 100 kb of *USAT24G1*.

We constructed a 1.5-Mb BAC and PAC contig centering on *USAT24GI* (ref. 7). We then used scaffold sequence-tag sites (STSs) to prioritize genomic sequencing. We systematically identified expressed sequences by examination of expressed-sequence tag (EST) databases, by cDNA selection[7] and by 3′ and 5′ RACE.

We then constructed a dense SNP map around *USAT24G1*. We aimed to identify alleles with a minimum frequency of 0.15, because the variation underlying complex traits probably consists of common alleles that are evolutionarily old[14], and rare alleles may require prohibitively large sample sizes to detect associations.

We identified 49 SNPs, 4 deletion-insertion polymorphisms and a (GGGC)n repeat with minor allele frequencies ≥15%. We genotyped these in a primary panel of 364 individuals in 80 nuclear families (AUS1). We carried out error checking and haplotype generation by MERLIN[15]. We measured D′ between markers from parental haplotypes[13] and depicted by GOLD[16] (**Fig. 1a**).

We examined the evidence for association to the $\mathrm{Log_e}$(IgE concentration) (LnIgE) by variance components analyses[17] (**Table 1**). Positive associations extended for approximately 100 kb in the A and Aii islands of LD (**Table 1 and Fig. 1b**).

We tested three-, four-, and five-marker haplotypes for association in a sliding window across the locus. The results were not essentially different from the single-point associations (data not shown). We found no evidence of genetic admixture with any marker[17]. We obtained similar results using transmission tests of association, except that the expected reduction in power gave weaker significance levels.

The region of association to LnIgE centered on one gene (*PHF11*) and extended to two flanking genes (*SETDB2* and *RCBTB1*; **Table 1** and **Fig. 1b**). We carried out a stepwise procedure in which the most significantly associated SNP was included as a covariate in the analysis using the QTDT program. We then repeated the analysis serially with the next most significantly associated remaining SNP included as an additional covariate until no significant associations remained. This analysis identified three SNPs as having independent effects (b5_2: $\chi^2 = 12.6$, $P = 0.0004$; b4_2: remaining $\chi^2 = 7.13$, $P = 0.008$; and b5_3: remaining $\chi^2 = 3.96$, $P = 0.05$). These SNPs are within 8.1 kb of each other (**Table 1**) and are found, respectively, in intron IX, intron V and the 3′ untranslated region of *PHF11*. Inclusion of SNPs in flanking genes as covariates did not abolish these associations.

We estimated the variation in the IgE levels from the locus attributable to genetic linkage as 11.5% (± 5.2% s.d.) by MERLIN. The variation attributable to association with the three SNPs measured by QTDT was 6.1%, suggesting the possible presence of other alleles contributing to linkage. Sequencing of cDNA from multiple subjects, however, identified no other SNPs. Our full coverage of the *PHF11* complex suggests that if other alleles contribute to linkage, they must be located elsewhere within the 7.5-Mb linkage peak.

Table 2 Association of common b4_2 b5_3 b43_1 haplotypes to total IgE

	AUS1			AUS2			UK2			ECZ*		
b4_2 b5_3 b43_1 haplotye	frequency	*P* (mul)	*P* (ind)	frequency	*P* (mul)	*P* (ind)	frequency	*P* (mul)	*P* (ind)	frequency	*P* (mul)	*P* (ind)
A: *2 *1 *1	0.449	0.011	0.014	0.396	0.12	0.078	0.378	0.0039	0.0153	0.294	0.0012	0.0001
B: *1 *2 *1	0.22		–	0.229		–	0.198		–	0.177		–
C: *1 *2 *2	0.193		–	0.238		–	0.192		–	0.234		0.0019
D: *2 *1 *2	0.084		0.005	0.077		0.036	0.09		0.0012	0.113		–
	n = 296			*n* = 533			*n* = 323			*n* = 531		

*RAST index included as a covariate. *P* (mul), multiallelic test of significance; *P* (ind), individual allele (haplotype) test of significance.

We typed six markers (b11_2, b4_2, b5_3, b43_1, b38_1 and b28_2) in other panels of subjects (AUS2, UK2 and ECZ). We assembled genotypes into three-marker haplotypes and carried out multi-allelic tests of association before examining individual haplotypes. The b4_2, b5_3, b43_1 haplotype showed consistent association to the LnIgE in each of the panels tested (**Table 2**). Two haplotypes containing the b4_2*2 and the b5_3*1 alleles (A and D; **Table 2**) showed negative association with the LnIgE, although they differed at b43_1. We observed positive association with the C haplotype (containing b4_2*1 and b5_3*2) in families with atopic dermatitis (ECZ). These results further localized the polymorphisms influencing IgE levels to b4_2 and b5_3.

The combined panel of Busselton families (AUS1 and AUS2) were recruited at random from the general Australian population[18]. We observed association to asthma in these subjects with the b4_2 and b5_3 markers ($P = 0.024$ and $P = 0.017$, respectively) using a transmission disequilibrium test[19]. In unrelated British adults, genotypes containing the b4_2*1 allele were significantly more common in severe asthmatics compared to non-atopic controls and were progressively less common in children with severe asthma and adults with mild asthma ($P = 0.001$ for trend; **Table 3**). We observed a similar pattern with b5_3. These results show relevance of the locus to severe clinical asthma.

LD mapping, therefore, consistently identified polymorphism in *PHF11* as being responsible for variation in LnIgE. *PHF11* encodes NY-REN-34, which was originally identified from patients with renal cell carcinoma[20], and its transcripts are highly represented in stomach, tonsil and B cells.

The gene product contains two PHD (plant homeodomain) zinc fingers, which suggest involvement in chromatin-mediated transcriptional regulation[21] (**Fig. 2**). PHD fingers normally posses two Zn^{2+} coordinating groups that contain cysteine and histidine residues. The N-terminal (5′) finger of the PHF11 finger pair, however, lacks one of the two coordinating groups.

The arrangement of PHD fingers in PHF11 is characteristic of human proteins, such as ALL-1 and AF10, whose genes are fused in some cases of acute lymphoblastic leukemia[22] (**Fig. 2**). Analogy to AF10 and ALL PHD fingers suggests that the PHF11 PHD finger pair probably has a role in homodimerization, protein binding or both.

The gene *PHF11* contains 10 exons. Public databases identified an alternative first exon with multiple overlapping variants that produce alternative start methionines for protein translation. We were able to confirm these variants by sequencing specific PCR products from cDNA panels (data not shown).

PCR analysis of a segment of the gene from the orthodox first exon (exon I) to exon III showed a uniform distribution in cDNA from all tissues studied (**Fig. 3a**). PCR products from the alternative first exon (exon Ia) were distributed in immune-related cells and tissues (**Fig. 3a**). We observed multiple splice variants. Cloning and sequencing of these showed that exon Ia and II contain multiple splice-donor and -acceptor sequences. We commonly found cDNA variants that showed skipping of exon II in these PCR products and in the cDNA databases. Loss of exon II results in an incomplete first PHD domain.

We also found variants that contained additional exons between exons V and VI (named exon Va and Vb). These variants differed by 54 bp and were only present in lung and peripheral blood leucocytes (PBL; **Fig. 3a**). Examination of PBL fractions showed that the variants were present in unactivated $CD4^+$, $CD8^+$ and $CD19^+$ cells but were absent in activated cells. Exons Va and Vb both result in premature stop codons. Alternative splicing with a premature stop codon has been identified as a mechanism for negative control of transcription[23], and a negative role for these variants is consistent with their expression in inactive T and B cells.

PHF11 is closely flanked by *SETDB2* and *RCBTB1*. *SETDB2* is 4 kb proximal to *PHF11* and is transcribed in the same direction. It contains methyl-CpG-binding (MBD) and SET domains (**Fig. 2**). The MBD seems to lack amino acids required to bind methylated CpG but retains the ability to bind unmethylated DNA. SET domains modulate gene expression epigenetically through histone H3 methylation[24,25]. *SETDB2* is probably a histone H3 methyltransferase, as it contains both active site and flanking cysteine residues that are important for catalytic activity[25].

We identified two coding variants in *SETDB2*, an A→G conservative substitution in exon 7 leading to the amino acid change Q117G (d8ex7) and a G→A variant in exon 10 leading to the amino acid change V473M (d8ex10), which showed weak associations with LnIgE ($P = 0.02$ and $P = 0.10$, respectively; **Table 1**). These effects disappeared when polymorphism in *PHF11* was taken into account.

The close genomic proximity of *SETDB2* and *PHF11* suggested that the two genes might be coordinately expressed or that a combined product, similar to IL-5 promoter REII region–binding protein, might be expressed (**Fig. 2**).

Table 3 Association of the b4_2 SNP with various degrees of clinical asthma

	Genotype		Odds ratio 95% confidence interval		
	*1/*1 and*1/*2	*2/*2	*P* compared to controls		
Adults with severe asthma	104 (80%)	26 (20%)	4.0	1.70–9.4	0.0008
Children with severe asthma	36 (75%)	12 (25%)	3.0	1.12–8.1	0.015
Adults with mild asthma	58 (69%)	26 (31%)	2.2	0.93–5.3	0.06
Non-atopic controls	14 (50%)	14 (50%)			

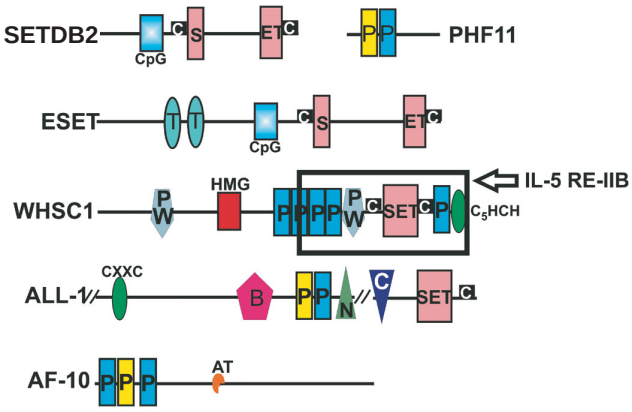
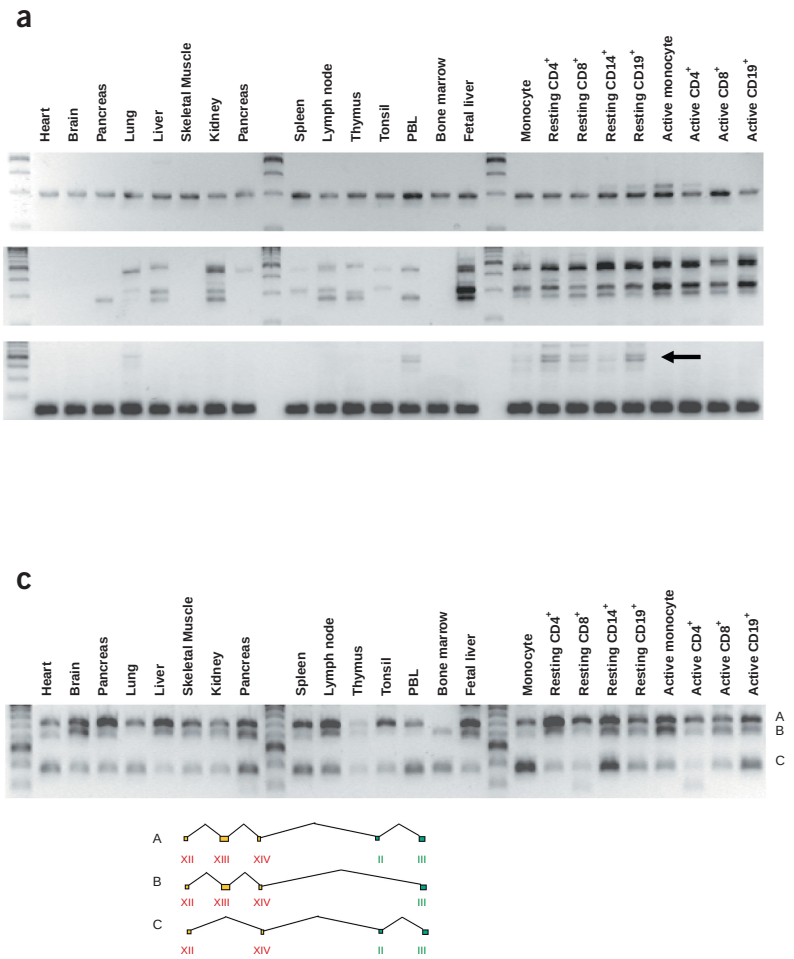


Figure 2 Domain architectures of SETDB2, PHF11 and related proteins, drawn approximately to scale. The SET domains of SETDB2 and ESET are bifurcated due to the presence of large insertions. The IL-5 promoter REII region–binding protein (RE-IIBP), which arises from an mRNA initiated in a middle intron of the gene *WHSC1* (ref. 32), is shown. Domain symbols: AT, AT-hook DNA-binding motif; B, bromodomain; C (white-on-black), cysteine-rich regions flanking SET domains; C, FYRC domain; C_5HCH, zinc finger domain; CpG, methyl-CpG-binding domains; CXXC, CXXC-type zinc finger; HMG, high mobility group domains; N, FYRN domain; P, PHD domains with two (blue rectangle) and one (yellow rectangle) Zn^{2+}-coordinating groups; PW, PWWP domains; SET, Su(var)3-9, Enhancer-of-zeste, trithorax methyltransferase domains; T, tudor domains.

A northern blot for *PHF11* showed high molecular weight bands that were most prominent in immune-related tissues (**Fig. 3b**). We observed bands of similar size with *SETDB2* (**Fig. 3b**). PCR between exon XII of *SETDB2* and exon III of *PHF11* identified three variably expressed products (A, B and C) that contained different exons (**Fig. 3c**). Sequence of a limited number of clones did not identify any extended open reading frames, but the presence of small variation of splice boundaries in exon II of *PHF11* meant that we were not able to rule out an ORF in some transcripts.

RCBTB1 is only 3 kb 3′ of *PHF11* but is transcribed in the opposite direction. It shows strong protein sequence similarity to RCC1 (regulator of chromatin condensation 1) and RLG (RCC1-like G exchanging factor). RCC1 binds to DNA and to histones H2A and H2B[26]. The close proximity of *RCBTB1* to *PHF11* and *SETDB2* might suggest coregulation of all three genes.

QTLs in other organisms are characterized by variation in a number of clustered genes[27]. Although our genetic analysis indicates that variation in IgE levels at this locus is attributable to variation in *PHF11*, the coordinate regulation of *PHF11* and *SETDB2* and the possible presence of more distant alleles influencing linkage suggest that investigation of the function of the locus should include the three genes of this complex.



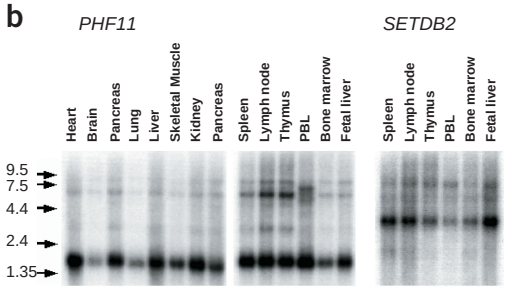


Figure 3 Expression of *PHF11*. (**a**) Splice variation in *PHF11* in multiple tissue cDNA panels Top, PCR amplification of orthodox exons I–III. A relatively consistent level of expression was seen in all tissues studied. Middle, PCR amplification of alternative exon Ia–III. The presence of multiple splice variants was observed. Transcripts originating from exon Ia were strongly expressed in tissues and cells associated with immune function. Bottom, PCR amplification of exons IV–VI. The presence of additional bands indicating retention of additional exons (Va and Vb) was observed in lung and immune tissues. The bands were present in cDNAs from unactivated lymphocytes and absent in activated lymphocytes. (**b**) Northern blots of *PHF11* and *SETDB2*. The figures show alternative probings of the same northern blot. The expected transcript size of approximately 1.6 kb for *PHF11* was seen in all tissues. A band of 3.0 kb was prominent in lymph node and thymus, and polymorphic higher molecular weight bands between 6.0 and 8.0 kb were visible in immune tissues. The expected band of 4.0 kb was seen for *SETDB2* in all tissues. Additional higher molecular weight bands were also seen. (**c**) Co-expression of the *SETDB2* and *PHF11* gene complex. PCR amplification between exon XII of *SETDB2* and exon III of *PHF11*. Three bands were observed (A, B and C). Their splice structure is depicted below.

We have previously shown linkage between this locus and the total serum concentration of immunoglobulins A, G and M. Testing the SNPs for association to these traits showed no evidence for association to IgA or IgG levels, but the total serum IgM concentration was associated with the b5_2 SNP ($P = 0.004$; variance = 3.6%). In contrast to serum IgE, high IgM levels accompanied the b5_2*2 rather than the b5_2*1 allele.

This narrow segment of chromosome 13q14 is within the minimum region commonly deleted in B-cell chronic lymphocytic leukemia[28]. Both *PHF11* and *SETDB2* show strong homologies with genes activated in multiple myeloma and leukemia, suggesting that B-cell clonal expansion and regulation of immunoglobulin expression may operate through shared mechanisms at this locus.

METHODS

Subjects. We carried out primary mapping on 364 subjects in 80 nuclear families selected from a population sample of 230 families of European descent from the rural town of Busselton in Western Australia[3,18] (the AUS1 panel). These families were selected to be informative for atopy. They all included both atopic and non-atopic members, and sibships of three or more were not exclusively atopic or non-atopic. The AUS2 panel consisted of the remaining 150 nuclear families from the population sample. The UK2 panel consisted of 87 nuclear families of British descent recruited when a child of the family attended an asthma clinic in the Oxford region. The families contained 216 offspring (148 sibling pairs). The ECZ panel consisted of 150 nuclear families predominately of European descent recruited from the dermatology clinics at the Great Ormond Street Hospital for Children, when one or more children of the family had active atopic dermatitis, as previously described[29].

A case-control panel of British descent including 130 adults with severe asthma, 49 children with severe asthma and 92 individuals with mild asthma was recruited at the Royal Brompton Hospital, London, as previously described[30]. The mean age of the adults with severe asthma was 44.8 years, that of the children with severe asthma was 11.9 years and that of the individuals with mild asthma was 27.8 years. The ratio of the forced expiratory volume in 1 s to the forced vital capacity (FEV1/FVC) of the individuals with severe asthma was predicted to be <65%. We compared their genotypes to those of adults recruited by the blood transfusion service and designated them non-atopic with a total serum IgE concentration <100 IU l^{-1}.

All studies of subjects were approved by the Central Oxford Regional Ethics Committee and by the Human Rights Committee of the University of Western Australia. All subjects or their parents gave written informed consent to the study.

Phenotypes. The subjects were administered a standard respiratory questionnaire[18]. 'Asthma' was defined as a positive answer to the questions "Have you ever had an attack of asthma?" and "If yes, has this happened on more than one occasion?". We measured the total serum IgE by solid-phase immuno-assay (Pharmacia CAP system) and $\log_e$ normalized the values before statistical analyses. Asthma in the case-control panel was diagnosed according to American Thoracic Society (ATS) criteria. Individuals with mild asthma suffered intermittent symptoms treated with infrequent (less than 2 times per week) short-acting β-agonists, did not use maintenance inhaled steroids and had normal lung function. Individuals with severe asthma had daily symptoms requiring regular inhaled β-agonist therapy and high-dose inhaled (at least 800 μg d^{-1} beclomethasone diproprionate or equivalent) or oral steroids and had impaired lung function.

SNP discovery and typing. We sequenced 10 diploid genomes (5 unrelated individuals with atopic disease and 5 unrelated control individuals) together with a pool of DNA from 32 unrelated individuals. This gave us 99.9% probability of detecting alleles with a minimal frequency of 0.2 and 99% probability of detecting alleles with a minimum frequency of 0.1 (ref. 31). Dilution experiments with known alleles indicated that we were able to detect allele frequencies >0.15 with the pool. In addition, we sequenced all *PHF11* exons in cDNA from 22 unrelated individuals, 12 of whom were atopic and 7 of whom were asthmatic. We extracted RNA using PAXgene blood RNA kit and subjected it to RT–PCR (QIAGEN OneStep RT-PCR) using two pairs of different primers for the alternate first exons that generated two major bands with or without exon II (primer sequences available on request). We sequenced 5–6 clones from the major bands using big dye labels. To control for *Taq* error, we screened for possible mutations in genomic DNA. This data, taken with the 9 complete cDNAs in the public databases, would have 99.9% probability of identifying SNPs with >0.1 frequency and 95% probability of identifying SNPs with >0.01 frequency[31]. Traces were assembled by the POLYPHRED/PHRAP programs. We genotyped SNPs by PCR and restriction digestion. In the absence of a natural restriction sequence, we modified a primer to generate a site. SNP primers are available on request.

Statistical analysis of association. We detected errors in SNP typing by testing for mendelian errors and by the MERLIN computer program[15], which identifies improbable recombination events from dense SNP maps. We generated SNP haplotypes by MERLIN and recoded them as individual alleles.

We carried out tests of association to quantitative traits using the QTDT program, which allowed use of markers and phenotypes as covariates in analyses[17]. We examined association to asthma and categorical traits by the Monks test routine of QTDT in family samples[17,19] and by exact tests of contingency tables in unrelated individuals with the StatXact program, version 4.0, reporting one-sided *P* values.

Sequence analyses. We mapped STSs onto BAC contigs built by a combination of *Hin*dIII digest fingerprinting and STS content. We sequenced an overlapping set of clones from the RPCI-11 BAC library using a hierarchical shotgun sequencing strategy. We analyzed genomic sequence using a modification of HPREP (G. Micklem, unpublished data) and screened for repeat elements in RepBase using REPEATMASKER and for matches in human, rodent, EST, STS and other DNA databases. We screened SWISSPROT, TREMBL and TREMBLNEW peptide databases using BLASTX for homologs to transcripts of unknown function. We screened for CpG islands using CPG, transcription factor elements and putative promoter regions using PROMOTERSCAN and exon predictions using GRAIL, GENSCAN, GENEPARSER and MZEF. We collated annotations using ACeDB. We identified known genes using BLASTN against the EMBL DNA database. Putative roles for remaining transcripts were established using PSI-BLAST and SMART.

IMAGE clone sequence and extension. We obtained IMAGE clones mapping to the region from Research Genetics and sequenced them on a 377 DNA sequencer using ABI Prism Big Dye Terminator (PE Applied Biosystems). We aligned consensus sequences for each IMAGE clone by the GCG program. We obtained Marathon-Ready cDNA RACE libraries from CLONTECH to extend 5′ and 3′ cDNA ends of the IMAGE clones. We designed two gene-specific primers for each direction for each consensus. We cut distinct bands from RACE PCR from gels and purified them. We cloned the bands with ZERO Blunt PCR Cloning kit (Invitrogen). We sequenced the inserts using Big Dye Terminator and integrated them into consensus sequences with GCG.

Tissue expression and splice variation. We obtained Human Multiple Tissue Northern (MTN) Blots and Human Immune System MTN blots from CLONTECH. We used Human Multiple Tissue, Human Immune System and Human Blood Fractions Multiple Tissues cDNA Panels from CLONTECH for expression analysis by PCR amplification of target sequences. We systematically investigated the exonic and intronic structure of the splice variants of *PHF11* by selective PCR, gel separation of products, cloning with ZERO Blunt PCR Cloning kit and Big Dye Terminator sequencing.

Accession numbers. *PHF11* and related transcripts are contained in UniGene cluster 279799. Alternative first exons for *PHF11* are observed in EMBL sequences AF155105, AL552215, BI463029 and BG759124. cDNA variants that show skipping of *PHF11* exon II include ESTHUM accession numbers BF662927 and BE787177 and EMBL accession number AL552215.

ACKNOWLEDGMENTS

We thank D. Gordon for support and advice, the people of Busselton and the many individuals who helped with their clinical testing and the subjects in the other panels of families for their participation. The study was funded by the Wellcome Trust and the National Asthma Campaign. Individuals with severe asthma were recruited with support from the Asmarley Trust.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 21 December 2002; accepted 18 April 2003
Published online 18 May 2003; doi:10.1038/ng1166

1. Jarvis, D. & Burney, P. ABC of allergies. The epidemiology of allergic disease. *Br. Med. J.* **316**, 607–610 (1998).
2. Eiberg, H., Lind, P., Mohr, J. & Nielsen, L.S. Linkage relationship between the human immunoglobulin E polymorphism and marker systems. *Cytogenet. Cell Genet.* **40**, 622 (1985).
3. Daniels, S.E. *et al.* A genome-wide search for quantitative trait loci underlying asthma. *Nature* **383**, 247–250 (1996).
4. Kimura, K. *et al.* Linkage and association of atopic asthma to markers on chromosome 13 in the Japanese population. *Hum. Mol. Genet.* **8**, 1487–1490 (1999).
5. Hizawa, N. *et al.* Genetic regulation of *Dermatophagoides pteronyssinus*-specific IgE responsiveness: a genome-wide multipoint linkage analysis in families recruited through 2 asthmatic sibs. Collaborative Study on the Genetics of Asthma (CSGA). *J. Allergy Clin. Immunol.* **102**, 436–442 (1998).
6. Beyer, K. *et al.* Evidence for linkage of chromosome 5q31–q33 and 13q12–q14 markers to atopic dermatitis. *J. Allergy Clin. Immunol.* **101**, 152 (1998).
7. Anderson, G.G. *et al.* Positive association to IgE levels and a physical map of the 13q14 atopy locus. *Eur. J. Hum. Genet.* **10**, 266–270 (2002).
8. Palmer, L.J. *et al.* Independent inheritance of serum immunoglobulin E concentrations and airway responsiveness. *Am. J. Respir. Crit. Care Med.* **161**, 1836–1843 (2000).
9. Risch, N.J. & Zhang, H. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. *Am. J. Hum. Genet.* **58**, 836–843 (1996).
10. Gerrard, J., Rao, D. & Morton, N. A genetic study of immunoglobulin E. *Am. J. Hum. Genet.* **30**, 46–58 (1978).
11. Burrows, B., Martinez, F., Halonen, M., Barbee, R. & Cline, M. Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N. Engl. J. Med.* **320**, 271–277 (1989).
12. Bhattacharyya, S., Leaves, N.I., Wiltshire, S., Cox, R. & Cookson, W.O. A high-density genetic map of the chromosome 13q14 atopy locus. *Genomics* **70**, 286–291 (2000).
13. Abecasis, G.R. *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**, 191–197 (2001).
14. Chakravarti, A. Population genetics—making sense out of sequence. *Nat. Genet.* **21**, 56–60 (1999).
15. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
16. Abecasis, G.R. & Cookson, W.O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
17. Abecasis, G.R., Cardon, L.R. & Cookson, W.O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292 (2000).
18. Hill, M.R. *et al.* Fc-ε RI-β polymorphism and risk of atopy in a general population sample. *Br. Med. J.* **311**, 776–779 (1995).
19. Monks, S.A., Kaplan, N.L. & Weir, B.S. A comparative study of sibship tests of linkage and/or association. *Am. J. Hum. Genet.* **63**, 1507–1516 (1998).
20. Scanlan, M.J. *et al.* Antigens recognized by autologous antibody in patients with renal-cell carcinoma. *Int. J. Cancer* **83**, 456–464 (1999).
21. Aasland, R., Gibson, T.J. & Stewart, A.F. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20**, 56–59 (1995).
22. Angioni, A. *et al.* Interstitial insertion of AF10 into the ALL1 gene in a case of infant acute lymphoblastic leukemia. *Cancer Genet. Cytogenet.* **107**, 107–110 (1998).
23. Walker, W., Girardet, C. & Habener, J. Alternative exon splicing controls a translational switch from activator to repressor isoforms of transcription factor CREB during spermatogenesis. *J. Biol. Chem.* **271**, 20145–20150 (1996).
24. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
25. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
26. Nemergut, M.E., Mizzen, C.A., Stukenberg, T., Allis, C.D. & Macara, I.G. Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. *Science* **292**, 1540–1543 (2001).
27. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
28. Oscier, D.G. Cytogenetic and molecular abnormalities in chronic lymphocytic leukaemia. *Blood Rev.* **8**, 88–97 (1994).
29. Cox, H.E. *et al.* Association of atopic dermatitis to the β subunit of the high affinity immunoglobulin E receptor. *Br. J. Dermatol.* **138**, 182–187 (1998).
30. Pulleyn, L.J., Newton, R., Adcock, I.M. & Barnes, P.J. TGFβ1 allele association with asthma severity. *Hum. Genet.* **109**, 623–627 (2001).
31. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat. Genet.* **27**, 234–236 (2001).
32. Garlisi, C.G. *et al.* A unique mRNA initiated within a middle intron of WHSC1/MMSET encodes a DNA binding protein that suppresses human IL-5 transcription. *Am. J. Respir. Cell Mol. Biol.* **24**, 90–98 (2001).