nature genetics

Positional cloning of a novel gene influencing asthma from Chromosome 2q14

Maxine Allen[1,8], Andrea Heinzmann[2,8], Emiko Noguchi[2,8], Gonçalo Abecasis[2], John Broxholme[2], Chris P Ponting[3], Sumit Bhattacharyya[2], Jon Tinsley[1], Youming Zhang[2], Richard Holt[2], E Yvonne Jones[2], Nick Lench[1], Alisoun Carey[1], Helene Jones[1], Nicholas J Dickens[3], Claire Dimon[1], Rosie Nicholls[1], Crystal Baker[1], Luzheng Xue[1], Elizabeth Townsend[1], Michael Kabesch[4], Stephan K Weiland[5], David Carr[4], Erika von Mutius[4], Ian M Adcock[6], Peter J Barnes[6], G Mark Lathrop[7], Mark Edwards[1], Miriam F Moffatt[2] & William O C M Cookson[2]

Asthma is a common disease in children and young adults. Four separate reports have linked asthma and related phenotypes to an ill-defined interval between 2q14 and 2q32 (refs. 1–4), and two mouse genome screens have linked bronchial hyper-responsiveness to the region homologous to 2q14 (refs. 5,6). We found and replicated association between asthma and the *D2S308* microsatellite, 800 kb distal to the *IL1* cluster on 2q14. We sequenced the surrounding region and constructed a comprehensive, high-density, single-nucleotide polymorphism (SNP) linkage disequilibrium (LD) map. SNP association was limited to the initial exons of a solitary gene of 3.6 kb (*DPP10*), which extends over 1 Mb of genomic DNA. *DPP10* encodes a homolog of dipeptidyl peptidases (DPPs) that cleave terminal dipeptides from cytokines and chemokines, and it presents a potential new target for asthma therapy.

We studied 244 families including 239 children with asthma for association between asthma and polymorphisms in the *IL1* gene cluster[7–10]. We used a standard questionnaire that reproducibly identifies asthma[11] with a heritability of 60–70% (ref. 12). Childhood asthma is usually accompanied by atopic or immunoglobulin E (IgE)-mediated inflammation, and we used the total serum IgE concentration (LnIgE) as a quantitative measure of atopy[13]. We found no association with any *IL1* SNPs but found a strong association between asthma and the neighboring microsatellite *D2S308* ($P = 0.0001$ with a multiallelic extended transmission/disequilibrium test[14]). Allele 3 of *D2S308* (*D2S308*3*) was positively associated with asthma (130 transmissions and 68 nontransmissions, $P = 0.00001$; **Supplementary Table 1** online), and *D2S308*5* was negatively associated with asthma (31 transmissions and 64 nontransmissions, $P = 0.007$).

A D′ of 0.33 approximates the limit of detection of LD between a disease and a marker[15]. We observed a D′ < 0.33 for 80% of marker pairs >100 kb apart[15], suggesting that an asthma susceptibility gene was located within 100 kb of *D2S308*. We built a BAC/PAC contig covering 1.5 Mb and sequenced 462 kb from four contiguous clones surrounding *D2S308*. We detected SNPs systematically from DNA repeat–free sequences. We genotyped 82 of 105 identified SNPs, including all those with a frequency >0.1 within 100 kb of *D2S308*. LD was distributed into four islands (A, B, C and D; **Fig. 1**). Islands A and B were clearly separated, producing distinct boundaries for the localization of genetic associations. The border between islands A and

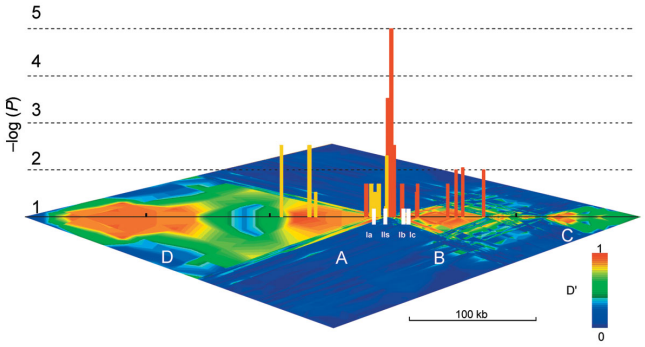

Figure 1 LD map of the locus associated with asthma and location of initial *DPP10* exons. The chromosomal region runs from left to right on the *x* axis at the bottom of the figure. The strength of association to asthma (red) and LnIgE (yellow) is plotted as –log(*P*) against position. The *D2S308* and *WTC122P* markers correspond to the highest peak. The scale bar indicates a distance of 100 kb. The graph is superimposed on the distribution of LD between markers: pairwise D′ values for LD are color-coded and plotted at the marker locations after completion by interpolation using the GOLD program[28]. Bright red and dark blue are opposite ends of the scale, with bright red indicating the most significant LD. The four initial exons of *DPP10* are shown as white bars. Later exons of *DPP10* are outside of the region of the LD map.

[1]Oxagen, Milton Park, Oxfordshire, UK. [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [3]MRC Functional Genetics Unit, Dept of Human Anatomy and Genetics, University of Oxford, UK. [4]University Children's Hospital, Munich, Germany. [5]Department of Epidemiology, University of Ulm, Ulm, Germany. [6]Department of Thoracic Medicine, National Heart and Lung Institute, London, UK. [7]Centre National De Genotypage, Evry, France. [8]These authors contributed equally to this work. Correspondence should be addressed to W.O.C.M.C. (william.cookson@ndm.ox.ac.uk).

Published online 19 October 2003; doi:10.1038/ng1256

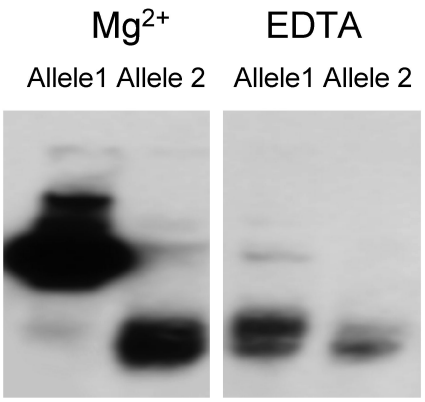

Figure 2 EMSA of *WTC122P*. Binding of *WTC122P* alleles to nuclear extracts from the T-cell line C8166, which expresses DPP10 as shown by western blotting. Protein bound to allele 1 sequence but not to allele 2 in the presence of 5 mM $MgCl_2$; binding was abolished in the presence of 10 mM EDTA.

B was flanked by the *DP1041* and *WTC121P* SNPs (**Supplementary Table 2** online), a distance of 913 bp.

Associations to asthma were confined to LD island B (**Fig. 1** and **Supplementary Table 2** online). We observed the strongest association with *WTC122P*, 1 kb proximal to *D2S308*. *WTC122P*1* was carried by 94.3% of asthmatic and 88.0% of nonasthmatic children. *WTC122P* alters the sequence of a known DPP family promoter element, CdxA[16], and *WTC122P* alleles showed differential protein binding with nuclear extracts from the T-cell line C8166 (**Fig. 2**). We saw weaker associations with asthma distally in island B, at *WTC124P* and around *WTC42P*. These associations could be attributed to a common haplotype, defined by $D' \geq 0.75$ between markers and containing the alleles *WTC122P*1*, *D2S308*3*, *WTC124P*1* and *WTC42P*2*. The haplotype had a frequency of 0.30 in founders. *WTC122P*1* and *D2S308*3* were not contained exclusively on the haplotype, and examination of all *WTC122P D2S308* haplotypes suggested susceptibility and protective effects additional to those of *WTC122P* (**Supplementary Table 3** online). We observed two peaks of association to LnIgE, which were confined to LD island A and centered around *DP1041* and *WTC100P* (**Fig. 1** and **Supplementary Table 2** online). These SNPs were not contained on a common haplotype, and inclusion of *DP1041* as a covariate in the analysis still left significant association ($P = 0.009$) to *WTC100P*. The results indicated a complex arrangement of susceptibility alleles affecting distinct but related phenotypes on either side of a hot spot of recombination.

We genotyped *WTC91P*, *WTC122P* and *D2S308* in 1,047 children aged 9–11 years from a cross-sectional study of Munich schoolchildren[17]. *D2S308*3* was positively associated with prick skin tests to common allergens ($P = 0.004$) and the presence of atopy ($P = 0.004$) and asthma ($P = 0.02$). Prick skin tests to common allergens ($P = 0.04$), the presence of atopy ($P = 0.02$) and asthma ($P = 0.02$) were associated with the haplotype *WTC91P*2 WTC122P*1 D2S308*3* but not with the haplotype *WTC91P*1 WTC122P*1 D2S308*3*, again indicating that *WTC122P*1* interacts with other unidentified susceptibility alleles.

We tested the relevance of the locus to clinical disease by measuring the frequency of the *WTC122P*1 D2S308*3* haplotype in 178 severe steroid-dependent asthmatics, 92 mild asthmatics and 304 unrelated controls. This haplotype was significantly more frequent among the severe asthmatics (odds ratio (OR) = 1.35, 95% confidence interval (c.i.) = 1.03–1.76, $P = 0.016$) but not among the mild asthmatics (OR = 0.91, 95% c.i. = 0.65–1.29).

We did not find any open reading frames (ORFs) from the region in public expressed sequence tag (EST) databases, so we amplified potential exons and used them in pools to screen a panel of cDNA libraries. We identified one clone (MEX4FB-1) from a fetal brain library. The clone contained an insert of 1,301 bp with an ORF of 1,137 bp. An exon of 60 bp (Ia), present at the 5′ end of the ORF, was recognized by three exon prediction programs and was 12.2 kb upstream of *D2S308*. Our extensive searching of the libraries, northern- and zoo-blot analyses and 5′ and 3′ rapid amplification of cDNA ends (RACE) experiments with all potential exons indicated that MEX4FB-1 represented the only gene expressed from the region. We found an overlap between the 3′ end of MEX4FB-1 and the 5′ end of a partial cDNA clone, KIAA1492. The sequences together encoded a full-length cDNA. Searches with the full sequence identified an additional clone (AK025075) that contained a 3′ poly(A)$^+$ tail upstream of that found in KIAA1492. We found this termination only in repeated 3′ RACE experiments in different tissues. The complete 3.6-kb cDNA of the gene results in a predicted ORF of 2,391 bp or 796 residues (**Supplementary Fig. 1** online). The gene spans approximately 1.4 Mb of genomic DNA and contains 26 exons. The protein, named DPP10, is the tenth member of the S9B family of DPPs and was previously called DRPR3 (ref. 18).

We isolated a mouse cDNA clone (BE862767) from a library and extended it to a full-length cDNA by three rounds of 5′ RACE. The primary transcript encodes a novel 2,370-bp ORF with a predicted peptide sequence of 789 residues. This sequence was 84% identical at the nucleotide level to the human gene (**Supplementary Fig. 2** online) and was within linkage peaks identified with mouse models of asthma[5,6].

We identified five different N termini by 5′ RACE from exon III of *DPP10* and designated the corresponding transcripts 1–5. These contain seven different 5′ exons, designated Ia–Ig (**Fig. 3**). We identified an alternate exon II (IIs) by 3′ RACE from exon Ia, which is 6,655 bp downstream of exon Ia (**Fig. 3**). The Ia–IIs transcript (transcript 6) contains a stop codon and encodes a 47-residue peptide. Three pre-

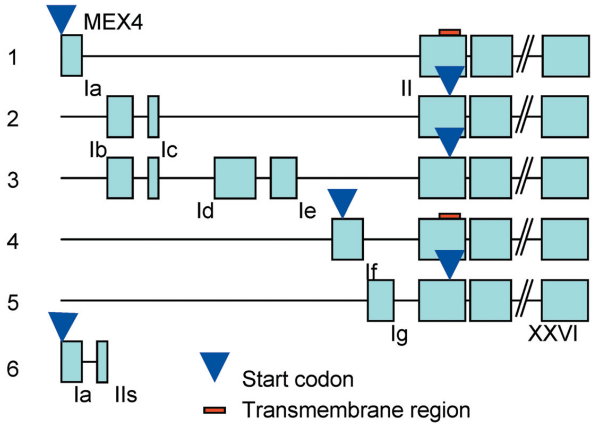

Figure 3 Exon structure of *DPP10*. Alternative transcripts are shown schematically and not to scale. Transcripts 1 and 4 (from exons Ia and If) provide an ORF from their first exon (20 and 23 residues upstream of the start of exon II) and encode a transmembrane domain. Transcripts 2, 3 and 5 begin their proteins within exon II. Transcripts 1–3 were isolated from brain and fetal brain cDNA and transcripts 4 and 5 were isolated from pancreas. Transcript 6 terminates within an alternative second exon (IIs) and was identified in cDNA from brain and testis.

dicted proteins are encoded by the other five transcripts (**Fig. 3**). Two proteins (from transcripts 1 and 4) contain a trans-membrane region. The other proteins are probably cytosolic. We confirmed this cellular localization by transfecting fusion protein vectors into HeLa cells. Transcript 1 was associated with plasma and internal membrane structures (**Fig. 4a**) and the cell surface (**Fig. 4b**). Transcript 2 showed a cytosolic distribution (**Fig. 4c**). Four alternate early exons (Ia, IIs, Ib and Ic) were contained within the LD map (**Fig. 1**) and were closely clustered at the junction between LD islands A and B. Exon IIs was close to *WTC122P* and *D2S308*, which showed the strongest association to asthma. We observed weaker associations near exons Ib and Ic.

Because there were no coding polymorphisms in *DPP10*, effects on asthma susceptibility may result in part from the presence or absence of the CdxA promoter element before exon Ib, leading to alternative splicing between membrane-bound and other forms of the protein. The alternate first exons of *DPP10* are spread over 500 kb of genomic DNA, and the distance between exon Ia and exon II is 866 kb. The position of exons Ia, IIs, Ib and Ic so distant from the body of the gene might reflect coregulation with the upstream IL-1 complex. Using northern blots, we found that *DPP10* was expressed strongly with multiple splice variants in brain, pancreas, spinal cord and adrenal glands (**Fig. 5a**). We saw less transcription in placenta, liver and airways (trachea). We found widespread expression of the initial exons of the gene in cDNA panels (**Fig. 5b**). High-molecular-weight complexes of DPP10 protein were free in serum, and the protein was abundant in T-cells (**Fig. 5c**,**d**).

The DPP proteins contain a β-propeller, which regulates substrate access to an α/β hydrolase catalytic domain. DPP4 and homologous enzymes remove N-terminal dipeptides from proteins, provided

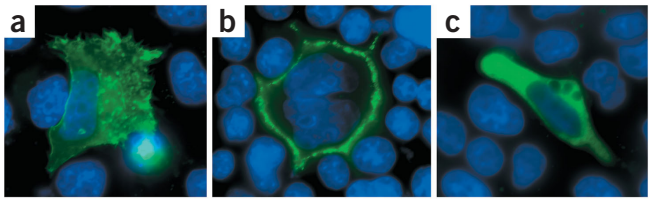

Figure 4 Cellular location of fusion proteins from transcripts 1 and 2. DPP10 from transcript 1 is a transmembrane protein whereas DPP10 from transcript 2 is cytosolic. (**a**) HeLa cells transfected with a fusion construct encoding green fluorescent protein (GFP) and DPP10 from transcript 1. The fusion protein (green) is observed on plasma membrane and intracellular membrane structures. (**b**) Prefixation staining using an antibody against V5 in a HeLa cell transfected with DPP10 from transcript 1 fused to V5. The fusion protein is detectable at the cell surface (green). (**c**) DPP10 from transcript 2 (green) shows a cytosolic pattern in a HeLa cell transfected with DPP10 fused to V5 with a histidine tag after immunostaining with antibody to V5. Nuclei are indicated by Hoechst staining (blue).

that the penultimate residue is proline[19]. Proteolytic specificity for a substrate containing proline is conferred by a proline-binding pocket adjacent to an active site triad. DPP10 lacks the serine of the catalytic triad (which is substituted by glycine) but conserves the aspartic acid and histidine (**Supplementary Fig. 3** online) as well as the hydrophobic nature of the residues lining the proline-binding pocket. A serine substitution within the catalytic triad is observed in the homologous proteins of human, mouse and bovine DPP6 and *Drosophila melanogaster* CG9059. The conservation of the catalytic histidine and aspartic acid residues in the absence of the catalytic serine in DPP10 and these homologs suggests that function is retained.

As DPP cleavage requires a penultimate proline in the substrate[19], we investigated whether catalytic serines may be provided by substrates that contain a serine after the proline at the cleavage point. We found that substrate-assisted catalysis was sterically feasible only for penultimate PxS motifs (proline, any amino acid, serine). We searched for cytokines with a signal peptide that was 20

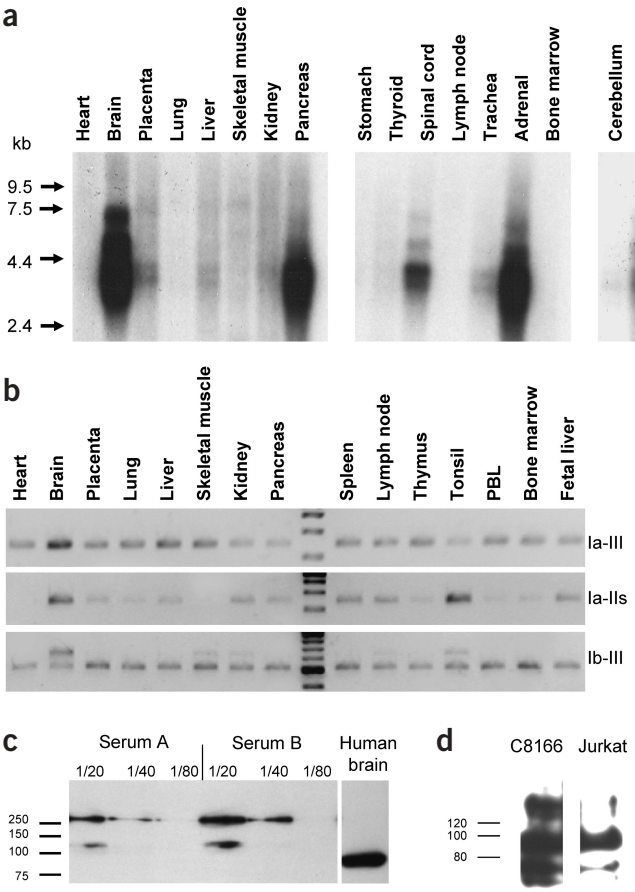

Figure 5 Tissue expression of *DPP10*. (**a**) Northern blots of *DPP10*. Gene expression is highest in brain, pancreas, spinal cord and adrenal gland, with lower levels of expression in placenta, liver and airways (trachea). Multiple splice variants are observed. (**b**) PCR amplification of initial exons of *DPP10* in multiple tissue cDNA panels. Amplification between exons Ia and III (top) and exons Ib and III (bottom) is observed in most tissues. Amplification between exons Ia and IIs (middle) is most marked in brain and tonsil. (**c**) Western blots of serum samples from two normal individuals with polyclonal antibodies to the C terminus of DPP10. (**d**) Western blots of lysates of C8166 and Jurkat T-cells. The expected molecular weight of the protein is ~75 kDa. The band observed in brain is consistent with a protein monomer, whereas serum and T-cell bands suggest protein oligomerization and differential glycosylation, such as that observed with other DPP family members[30].

capasa^SPYSSDTTPC	RANTES
lclsdg^KPVSLSYRCP	SDF-1
spggla^GPASVPTTCC	EOTAXIN
lsgigg^VPLSRTVRCT	IP10
spqgla^QPDSVSIPIT	MCP-2
iflglg^QPRSPKSKRK	IL17ß
alvtns^APTSSSTKKT	IL2
gplasa^GPVSAVLTEL	GCP-2
tllvra^TPVSQTTTAA	IL18 BP

Figure 6 Cytokines and chemokines containing penultimate PxS motifs. Human chemokines and cytokines that have a serine within 20 amino acids of a predicted signal peptide cleavage site and contain a PxS motif (where 'x' represents any amino acid). No other relative positions of proline and serine were compatible with substrate-assisted catalysis. The predicted signal peptide cleavage bond is indicated by a caret (^) and PxS motifs are highlighted in yellow. The molecules contain the PxS two residues after the cleavage site.

amino acids or fewer from the N terminus and PxS tripeptides starting at the +2 position among a redundant list of approximately 1,000 human cytokine sequences from the Entrez database. Nine cytokines fit these criteria (**Fig. 6**). No cytokines were identified by control searches. Several inflammatory cytokines contained the PxS motif at the +2 position, including SDF-1a, IP10, eotaxin and RANTES (**Fig. 6**). The nature of these putative substrates tentatively suggests a mechanism in which DPP10 may modulate asthmatic airway and other inflammation.

METHODS

Subjects. We used three panels of families to map the locus. The AUS1 panel consisted of 80 nuclear families subselected to be informative for atopy from a randomly ascertained Australian population sample of 230 families[1]. The panel contained a total of 203 offspring, 12% of whom were asthmatic. The UK1 panel consisted of 77 nuclear and extended families recruited from asthma and allergy clinics in the United Kingdom[1]. These families included 215 offspring, of whom 56% were asthmatic. The UK2 panel consisted of 87 nuclear families recruited through hospitals or general practitioner asthma clinics in the Oxford region. The families included 216 offspring, of whom 44% were asthmatic. As each panel showed evidence of association to *D2S308*, we combined them to maximize power. The combined panel contained 244 families and 1,122 subjects, with 239 asthmatic children and 103 asthmatic sibling pairs.

We selected 1,047 children representative of a general population from a cross-sectional study conducted in Munich, Germany, to assess the prevalence of asthma and allergies in schoolchildren aged 9–11 years[17]. The sample included 118 children with doctor-diagnosed asthma and 139 'super-normal' controls.

We recruited a replication panel of 129 severe adult asthmatics, 49 severe childhood asthmatics and 92 mild asthmatics at the Royal Brompton Hospital, London, as previously described[20]. The mean age of the adult severe asthmatics was 44.8 years, whereas that of the child asthmatics was 11.9 years and that of the mild asthmatics was 27.8 years. The FEV1 of the severe asthmatics was <65% predicted. Their genotypes were compared to those of 304 unrelated adults of European descent recruited in London as part of a study of dermatitis (mean age 37.1 years)[21]. The study was approved by local ethics committees. All subjects or their parents gave written informed consent to the study.

Phenotypes. We administered a standard respiratory questionnaire[22] to the subjects in the family panels. We defined "asthma" 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?" Answers to these questions corresponded closely (>95%) to the answer to the question "Has your doctor ever told you that you have asthma?"

Our German population sample was investigated as part of the International Study of Asthma and Allergies in Childhood according to its protocols[23]. Phenotyping procedures included parental questionnaires, skin prick testing, pulmonary function testing and bronchial challenge with hyperosmolar saline (4.5%) and measurements of total and specific serum IgE. We selected from the population all children with a doctor's diagnosis of asthma and compared their results to those of super-normals without any history or positive measurements of asthma or atopy (doctor diagnosis of asthma, negative; current wheeze, negative; skin prick test, <2 mm; bronchial hyperresponsiveness, negative; and total serum IgE concentration, ≤50 IU l^{-1}).

Asthma in the severe asthma panel was diagnosed according to American Thoracic Society criteria[24]. Individuals with severe asthma had daily symptoms requiring regular inhaled β-agonist therapy and high-dose inhaled (≥800 μg d^{-1} beclomethasone diproprionate or equivalent) or oral steroids and had impaired lung function. Individuals with mild asthma suffered intermittent symptoms treated with infrequent (<2 times per week) short-acting β-agonists, did not use maintenance inhaled steroids and had normal lung function.

Physical mapping. We extended a YAC contig around *D2S308* from a basic map published online by the Whitehead Institute. We used the central YACs as the basis for a BAC/PAC map and also mapped them using the restriction enzymes *Not*I and *Sal*I. We constructed a BAC/PAC contig extending for ~1 Mb by screening known sequence-tagged sites (STSs) and ESTs. We closed gaps by recovering clone ends using anchor-bubble PCR and rescreening libraries. The contig contained 22 clones with an average insert size of 110 kb and required the recovery of >40 new STSs to construct. We determined the sizes of all clones by pulsed field gel electrophoresis. We confirmed the localization of clones by fluorescence *in situ* hybridization.

Sequencing. We shotgun-sequenced four clones spanning 400 kb around the marker *D2S308* to 12× coverage. We assembled and analyzed the genomic sequence using a modification of HPREP; screened for repeat elements in RepBase using REPEATMASKER; screened for matches in human, rodent, EST, STS and other DNA databases; and predicted exons using GRAIL, GENSCAN, GENEPARSER and MZEF. We collated annotations using ACeDB.

SNP discovery and genotyping. We systematically detected SNPs in regions of DNA that were free of repeats by sequencing ten diploid genomes (five unrelated atopic subjects and five unrelated controls), a pool of DNA from 32 unrelated individuals and a pool of DNA from 150 asthmatic children. Dilution experiments with known alleles indicated that we could detect allele frequencies >0.15 with the pool. We carried out additional sequencing in and around putative exons in 22 unrelated individuals, 12 of whom were atopic and 7 of whom were asthmatic. This data would have 99.9% probability of identifying SNPs with >0.1 frequency and 95% probability of identifying SNPs with >0.01 frequency[25].

We genotyped all family members with respect to all markers. The microsatellite marker *D2S308* was genotyped by semiautomated fluorescence methods. We discovered SNPs through direct sequencing of nonrepetitive DNA and assembled traces 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. We genotyped SNPs without restriction sites and within 100 kb of *D2S308* by pyrosequencing. Outside this 200-kb region, choice was dictated by the availability of restriction sites (**Supplementary Table 2** online).

Statistical analysis of association. We dectected errors in SNP typing by testing for mendelian errors and by the MERLIN computer program[26], which identifies improbable recombination events from dense SNP maps. We examined association to asthma by the transmission/disequilibrium test using all affected children to maximize power and assessed association to LnIgE in all family members by variance components with the quantitative transmission/disequilibrium test[27]. We tested data in population and case-control panels using SPSS for OSF 6.1.4 and StatXact version 5.0. Haplotypes in families were generated by MERLIN[26]. We assessed LD between markers by estimating D′ from the

parental haplotypes[15] and portrayed LD using GOLD[28]. Haplotypes in unrelated samples were generated using SNPHAP.

Electrophoretic mobility shift assay (EMSA). We labeled oligonucleotides containing alleles *WTC122P*1* and *WTC122P*2* with biotin at the 3′ end and incubated them with complementary oligonucleotides to form double-stranded target DNA. We combined the annealed oligonucleotides with nuclear extracts from the C8166 T-cell line, which expressed DPP10 as shown by western blotting. After incubation at room temperature, we separated the samples on 5% polyacrylamide gels and electroblotted them onto Biodyne-charged membrane. We determined the positions of the biotin-labeled oligonucleotides using the LightShift chemiluminescent EMSA kit (Pierce) as described by the manufacturer and visualized them by exposure to film or CCD camera. We investigated protein binding in a range of concentrations of $MgCl_2$ and EDTA according to manufacturer's instructions.

Library screening. Potential exons were identified by at least two sequence analysis programs and were free of repeat sequences and at least 50 bp in length. We used PCR products in pools of 3–6 probes to screen commercial cDNA libraries prepared in the lambda-triplex phage vector (Clontech). The libraries included adult brain, fetal brain, lung, testis, trachea and skeletal muscle. We plated ~1 million phage clones. Positive phage plaques were 'cored', placed in dilution buffer and plated onto secondary plates. Single hybridizing plaques were picked and diluted and grown in a Luria broth culture of BM25.8 cells (Clontech) before replating. Single colonies were picked for plasmid DNA isolation and dynamic ET terminator cycle sequencing according to the manufacturer's protocol (Amersham). After using forward and reverse vector primers, we designed walking primers to complete the full-length sequence.

RACE. We carried out human 5′ and 3′ RACE experiments on a range of tissues using commercially prepared RACE-ready cDNA (Marathon Ready cDNA, Clontech) according to the manufacturer's instructions.

Transfection and immunocytochemistry. We transfected HeLa cells with pcDNA 3.1/NT-GFP-DPP10 (Transcript1), pcDNA3.1/V5-His-DPP10 (Transcript 1) or pcDNA3.1/V5-His-DPP10 (Transcript 2) using Lipofectamine (Invitrogen) following the manufacturer's protocol. After 2 d of expression, we directly immunostained the cells with mouse antibody against V5 (Invitrogen) at 4 °C (prefixation staining) or stained them after fixation with 3% paraformaldehyde and permeation with 0.1% saponin. We then stained cells with Alexa 488–conjugated antibody to mouse IgG (Molecular Probes) and washed them with buffer containing Hoechst 33342. The results were observed and recorded under a Leica microscope.

Northern blots. We created a probe by amplifying the MEX4FB-1 insert and digesting it with *Mbo*I and *Xba*I. This probe contained the last 8 bases of exon 3, all of exons 4–11 and the first 24 bases of exon 12. We hybridized 50 ng of α^{32}PdCTP-labeled probe to commercial northern blots (MTNI, MTNII and MTNIII and Human Brain II, Clontech), which represent mRNA from 31 different tissues, at 65 °C in EXPRESSHYBE buffer (Clontech), washed them twice with 500 ml of 2× saline sodium citrate with 0.01% SDS for 30 min at 50 °C and once with 500 ml 1× saline sodium citrate with 0.01% SDS for 30 min at 50 °C and then exposed them to autoradiographic film for 2 d with signal intensifying screens.

PCR screening of MTC panels. We used Human Multiple Tissue, Human Immune System and Human Blood Fractions cDNA Panels from Clontech for expression analysis by PCR amplification of target sequences. We amplified cDNA with 38 cycles of 1 min at 95 °C, 1 min at 54 °C and 1 min at 72 °C each. Primers are available on request.

Western blots. We diluted serum samples from volunteers (1:20, 1:40, 1:80) in sample buffer (7.5 mM Tris, pH 6.8, 3.8% SDS, 4 M urea, 20% glycerol, 5% mercaptoethanol) to a total volume of 20 µl and denatured them at 95 °C for 5 min. We loaded the samples onto a 12% polyacrylamide SDS denaturing gel and separated them by electrophoresis at 100 V for 60 min. After electrophoresis, we transferred the proteins to a 0.4-mm nitrocellulose membrane by blotting at 200 V for 2 h. We blocked the filters overnight in 5% milk solution at 4 °C before antibody detection. We generated the affinity-purified DPP10 C-terminal antibody against a DPP10 peptide (NH_2-CLK-EEI-SVL-PQE-PEE-DE) in rabbits. We incubated the filters with the DPP10 antibody (diluted 1:250) in 5% milk at room temperature for 60 min. After washing, we then incubated the filters with antibody to rabbit IgG conjugated to horseradish peroxidase (diluted 1:2,000) in 5% milk at room temperature for 60 min. After a final rigorous washing step, we detected bound antibody by chemiluminescence substrate (Roche) and autoradiography.

Identification of PxS motifs. We sought PxS motifs (where 'x' represents any amino acid) among a redundant list of ~1,000 human cytokine amino acid sequences from the Entrez database. We filtered sequences using a perl script and the sigcleave module from Bioperl. As controls, we used an identical protocol to detect cytokines with SxP, xSP and xPS tripeptides with serines at the –2, –1 and +1 positions, respectively, and found none. A search of a nonredundant list of all 112 human neuropeptides from the National Center for Biotechnology Information protein database under the same criteria identified no matches to the PxS or SxP motifs.

Examination of structure. We assessed the steric constraints of substrate-assisted catalysis using as a template the crystal structure of an inactive variant of PPOP complexed with an octapeptide (Protein Data Bank accession number 1E8N). Using the graphics program O[29], we substituted a serine residue into the substrate peptide at position +2 after the proline. From this position (without any adjustments of the mode), the substrate serine could adopt a side-chain conformation that placed its Oγ atom only 2 Å distant from the position required for the key Oγ atom of the catalytic serine in the PPOP active site. No relative positions of proline and serine other than the PxS motif were compatible with this putative mechanism.

URLs. The Whitehead Institute physical map is available at http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map. Linkage results for our genome screen are available at http://www.well.ox.ac.uk/asthma. The Entrez database is available at http://www.ncbi.nlm.nih.gov/entrez. ACeDB is available at http://www.acedb.org/. REPEATMASKER is available at http://repeatmasker.genome.washington.edu. QTDT is available at http://www.sph.umich.edu/csg/abecasis/QTDT/. MERLIN is available at http://www.sph.umich.edu/csg/abecasis/Merlin/. SNPHAP is available at http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt. Phage clone plating protocols PT3003-1, version PR09529 are available at http://www.clontech.com/libraries/#techinfo. The sigcleave module from Bioperl is available at http://bioperl.org.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank D. Gordon for support; L. Cardon for statistical and other advice; the people of Busselton and the many individuals who helped with their clinical testing; the subjects in other panels of individuals and families for their participation; and R. Hennion, S. Lees, J. Herbert, N. Robinson and J. Faux for their supporting contributions to experiments. Testing of unrelated severe asthmatics was supported by the Asmarley Trust. The study was funded by the Wellcome Trust and the National Asthma Campaign. E.Y.J. is a Cancer Research UK Principal Fellow.

COMPETING INTERESTS STATEMENT

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

Received 23 July; accepted 29 September 2003
Published online at http://www.nature.com/naturegenetics/

1. Daniels, S.E. *et al.* A genome-wide search for quantitative trait loci underlying asthma. *Nature* **383**, 247–250 (1996).
2. 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).
3. Koppelman, G.H. *et al.* Genome-wide search for atopy susceptibility genes in Dutch families with asthma. *J. Allergy Clin. Immunol.* **109**, 498–506 (2002).
4. Wjst, M. *et al.* A genome-wide search for linkage to asthma. German Asthma Genetics Group. *Genomics* **58**, 1–8 (1999).

5. Ewart, S.L. *et al.* Quantitative trait loci controlling allergen-induced airway hyperresponsiveness in inbred mice. *Am. J. Respir. Cell. Mol. Biol.* **23**, 537–545 (2000).
6. De Sanctis, G.T. *et al.* Quantitative locus analysis of airway hyperresponsiveness in A/J and C57BL/6J mice. *Nat. Genet.* **11**, 150–154 (1995).
7. Pociot, F., Molvig, J., Wogensen, L., Worsaae, H. & Nerup, J. A TaqI polymorphism in the human interleukin-1 beta (IL-1 beta) gene correlates with IL-1 beta secretion *in vitro*. *Eur. J. Clin. Invest.* **22**, 396–402 (1992).
8. di Giovine, F.S., Takhsh, E., Blakemore, A.I. & Duff, G.W. Single base polymorphism at –511 in the human interleukin-1 beta gene (IL1 beta). *Hum. Mol. Genet.* **1**, 450 (1992).
9. Bailly, S., di Giovine, F.S., Blakemore, A.I. & Duff, G.W. Genetic polymorphism of human interleukin-1 alpha. *Eur. J. Immunol.* **23**, 1240–1245 (1993).
10. Epplen, C. *et al.* Dinucleotide repeat polymorphism in the human IL1A gene. *Hum. Mol. Genet.* **3**, 1710 (1994).
11. O'Connor, G.T. & Weiss, S.T. Clinical and symptom measures. *Am. J. Respir. Crit. Care Med.* **149**, S21–S28 (1994).
12. Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. Genetics of asthma and hay fever in Australian twins. *Am. Rev. Respir. Dis.* **142**, 1351–1358 (1990).
13. Zhang, Y. *et al.* Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat. Genet.* **34**, 181–186 (2003).
14. Sham, P.C. & Curtis, D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* **59** (Pt 3), 323–336 (1995).
15. Abecasis, G.R. *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**, 191–197 (2001).
16. Rao, N.V., Rao, G.V. & Hoidal, J.R. Human dipeptidyl-peptidase I. Gene characterization, localization, and expression. *J. Biol. Chem.* **272**, 10260–10265 (1997).
17. Kabesch, M. *et al.* Association between polymorphisms in caspase recruitment domain containing protein 15 and allergy in two German populations. *J. Allergy Clin. Immunol.* **111**, 813–817 (2003).
18. Qi, S.Y., Riviere, P.J., Trojnar, J., Junien, J.L. & Akinsanya, K.O. Cloning and characterization of dipeptidyl peptidase 10, a new member of an emerging subgroup of serine proteases. *Biochem. J.* **373**, 179–189 (2003).
19. Lambeir, A.M. *et al.* Kinetic investigation of chemokine truncation by CD26/dipeptidyl peptidase IV reveals a striking selectivity within the chemokine family. *J. Biol. Chem.* **276**, 29839–29845 (2001).
20. Pulleyn, L.J., Newton, R., Adcock, I.M. & Barnes, P.J. TGFbeta1 allele association with asthma severity. *Hum. Genet.* **109**, 623–627 (2001).
21. Cox, H.E. *et al.* Association of atopic dermatitis to the beta subunit of the high affinity immunoglobulin E receptor. *Br. J. Dermatol.* **138**, 182–187 (1998).
22. Hill, M.R. *et al.* Fc epsilon RI-beta polymorphism and risk of atopy in a general population sample. *Brit. Med. J.* **311**, 776–779 (1995).
23. ISAAC. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. *Lancet* **351**, 1225–1232 (1998).
24. 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).
25. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat. Genet.* **27**, 234–236 (2001).
26. 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).
27. 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).
28. Abecasis, G.R. & Cookson, W.O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
29. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47** (Pt 2), 110–119 (1991).
30. Duke-Cohan, J.S., Morimoto, C., Rocker, J.A. & Schlossman, S.F. A novel form of dipeptidylpeptidase IV found in human serum. Isolation, characterization, and comparison with T lymphocyte membrane dipeptidylpeptidase IV (CD26). *J. Biol. Chem.* **270**, 14107–14114 (1995).