nature genetics

A genome-wide association study identifies *CDHR3* as a susceptibility locus for early childhood asthma with severe exacerbations

Klaus Bønnelykke[1,24,25], Patrick Sleiman[2,24], Kasper Nielsen[3,24], Eskil Kreiner-Møller[1], Josep M Mercader[4], Danielle Belgrave[5,6], Herman T den Dekker[7–9], Anders Husby[1,10], Astrid Sevelsted[1], Grissel Faura-Tellez[11,12], Li Juel Mortensen[1], Lavinia Paternoster[13], Richard Flaaten[1], Anne Mølgaard[1], David E Smart[10], Philip F Thomsen[14], Morten A Rasmussen[15], Silvia Bonàs-Guarch[4], Claus Holst[16], Ellen A Nohr[17,18], Rachita Yadav[3], Michael E March[2], Thomas Blicher[19], Peter M Lackie[11], Vincent W V Jaddoe[7,9,20], Angela Simpson[5], John W Holloway[11], Liesbeth Duijts[8,9,21], Adnan Custovic[5], Donna E Davies[10], David Torrents[4,22], Ramneek Gupta[3], Mads V Hollegaard[23], David M Hougaard[23], Hakon Hakonarson[2,25] & Hans Bisgaard[1,25]

© 2013 Nature America, Inc. All rights reserved.

Asthma exacerbations are among the most frequent causes of hospitalization during childhood, but the underlying mechanisms are poorly understood. We performed a genome-wide association study of a specific asthma phenotype characterized by recurrent, severe exacerbations occurring between 2 and 6 years of age in a total of 1,173 cases and 2,522 controls. Cases were identified from national health registries of hospitalization, and DNA was obtained from the Danish Neonatal Screening Biobank. We identified five loci with genome-wide significant association. Four of these, *GSDMB*, *IL33*, *RAD50* and *IL1RL1*, were previously reported as asthma susceptibility loci, but the effect sizes for these loci in our cohort were considerably larger than in the previous genome-wide association studies of asthma. We also obtained strong evidence for a new susceptibility gene, *CDHR3* (encoding cadherin-related family member 3), which is highly expressed in airway epithelium. These results demonstrate the strength of applying specific phenotyping in the search for asthma susceptibility genes.

Acute asthma exacerbations are among the most frequent causes of hospitalization during childhood and are responsible for large healthcare expenditures[1–4]. Available treatment options for prevention and treatment of asthma exacerbations are inadequate[5], suggesting that asthma with severe exacerbations may represent a distinct subtype of disease and demonstrating a need for improved understanding of its pathogenesis.

Asthma heritability is estimated to be 70–90% (refs. 6,7), but only a limited number of susceptibility loci have been verified in genome-wide association studies (GWAS)[8–13]. Larger GWAS may identify new susceptibility loci with smaller effects, but, owing to the large heterogeneity in asthma[14], an alternative strategy is to increase phenotype specificity in genome-wide analyses. A specific phenotype is likely to be more closely related to a specific pathogenetic mechanism, and focusing on a particular phenotype may increase the power of genetic studies.

We aimed to increase understanding of the genetic background of early childhood asthma with severe exacerbations by conducting a

[1]Copenhagen Prospective Studies on Asthma in Childhood, Health Sciences, University of Copenhagen & Danish Pediatric Asthma Center, Copenhagen University Hospital, Gentofte, Denmark. [2]Center for Applied Genomics, Children's Hospital of Philadelphia (CHOP), Philadelphia, Pennsylvania, USA. [3]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark. [4]Joint Institute for Research in Biomedicine and Barcelona Supercomputing Center (IRB-BSC) Program on Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain. [5]Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester and University Hospital of South Manchester, Manchester, UK. [6]Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester, UK. [7]Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands. [8]Department of Pediatrics, Division of Respiratory Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [9]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [10]Brooke Laboratory, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, University Hospital Southampton, Southampton, UK. [11]Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton, UK. [12]Pediatric Pulmonology and Pediatric Allergology, University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Groningen Research Institute for Asthma and COPD, Groningen, The Netherlands. [13]Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, UK. [14]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [15]Department of Food Science, University of Copenhagen, Copenhagen, Denmark. [16]Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark. [17]Institute of Clinical Research, University of Southern Denmark, Aarhus, Denmark. [18]Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark. [19]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [20]Department of Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands. [21]Department of Pediatrics, Division of Neonatology, Erasmus Medical Center, Rotterdam, The Netherlands. [22]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [23]Danish Centre for Neonatal Screening, Department of Clinical Biochemistry and Immunology, Statens Serum Institut (SSI), Copenhagen, Denmark. [24]These authors contributed equally to this work. [25]These authors jointly directed this work. Correspondence should be addressed to K.B. (kb@copsac.com).

Received 27 May; accepted 28 October; published online 17 November 2013; doi:10.1038/ng.2830

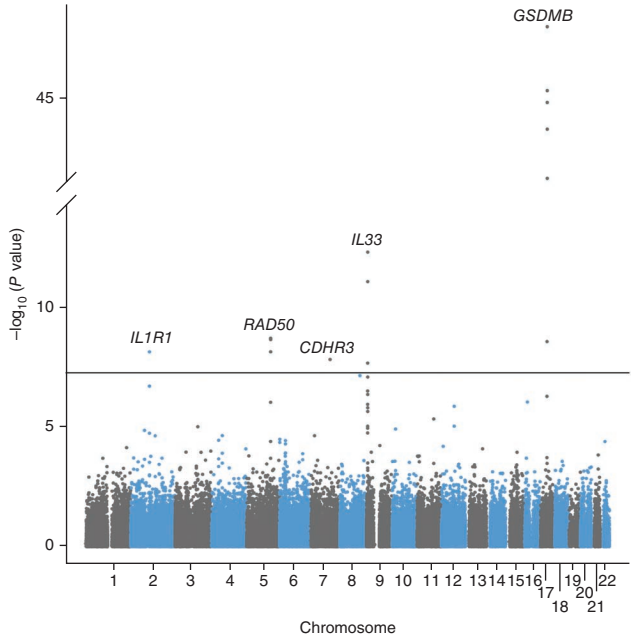

Figure 1 Manhattan plot for the discovery genome-wide association analysis. The horizontal line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$).

GWAS of this particular asthma phenotype. We identified children with recurrent acute hospitalizations for asthma occurring between 2 and 6 years of age (cases) from the Danish National Patient Register. We then extracted and amplified DNA from dried blood spot samples isolated from the Danish Neonatal Screening Biobank, as previously described[15,16], before genome-wide array genotyping (Affymetrix Axiom CEU array).

Case criteria were fulfilled for 2,029 of 1.7 million children born in Denmark between 1982 and 1995 (1.1/1,000 children). The final case cohort (Copenhagen Prospective Studies on Asthma in Childhood exacerbation cohort, $COPSAC_{exacerbation}$) after genotyping and quality control comprised 1,173 children (**Supplementary Fig. 1**). Compared to the general population, cases were more often boys (67 versus 51%) and more often had mothers who smoked during pregnancy (32 versus 15%) (**Supplementary Tables 1** and **2**). Controls consisted of 2,511 individuals of Danish descent without asthma who were previously genotyped (Illumina Human610-Quad v1.0 BeadChip). We analyzed association between disease and 124,514 SNPs genotyped in both cases and controls, and we accounted for population stratification by multidimensional scaling. The genomic inflation factor was 1.04. The genome-wide association analysis detected an excess of association signals beyond those expected by chance (**Supplementary Fig. 2**), and SNPs from five regions reached genome-wide significance ($P < 5 \times 10^{-8}$; **Fig. 1** and **Supplementary Fig. 3**). The top SNPs from the five loci were rs2305480 in *GSDMB* (odds ratio (OR) = 2.28, $P = 1.3 \times 10^{-48}$), rs928413 near *IL33* (OR = 1.50, $P = 4.2 \times 10^{-13}$), rs6871536 in *RAD50* (OR = 1.44, $P = 1.7 \times 10^{-9}$), rs1558641 in *IL1RL1* (OR = 1.56, $P = 6.6 \times 10^{-9}$) and rs6967330 in *CDHR3* (OR = 1.45, $P = 1.4 \times 10^{-8}$) (**Table 1**). Validation of results for the top SNPs by regenotyping of cases and use of an alternative control population gave similar results (**Supplementary Tables 3** and **4**).

Association analyses in the discovery cohort stratified on number of asthma-related hospitalizations showed higher OR with increasing number of hospitalizations for all five SNPs (**Table 2**). There was no significant interaction between the top SNPs and no effect modification by sex.

We first sought replication in the childhood-onset stratum (with onset before 16 years of age) from a previous GWAS of asthma including 14,503 individuals conducted by the GABRIEL Consortium[11] (**Supplementary Table 5**), which showed evidence of association for all 5 of the genome-wide significant loci reported here (**Table 1**). The *CDHR3* locus was the only locus that had not previously been associated with asthma or any other atopic trait. We therefore followed up the top SNP from this locus (rs6967330) by further replication in a total of 3,975 children from 2 birth cohorts of European ancestry ($COPSAC_{2000}$ and the Manchester Asthma and Allergy Study (MAAS)) and in 1 cohort with a population of mixed ancestry (Generation R). There was evidence for association with asthma before the age of 6 years in combined analyses of the three birth cohorts and in the combined replication sets (**Table 1**, **Supplementary Fig. 4** and **Supplementary Table 6**), as well as in a subsample including the 980 individuals with non-European ancestry (**Supplementary Table 6**).

Phenotype-specific replication was possible in the $COPSAC_{2000}$ and MAAS birth cohorts with prospective registration of acute asthma hospitalizations and exacerbations from birth to 6 years of age in a

Table 1 Discovery and replication results for the five genome-wide significant loci in the discovery analyses

Chr.	SNP effect allele	Nearest gene	Distance to gene (bp)	Effect allele frequency	Stage	OR (95% CI)	*P* value (fixed-effects model)[a]	*P* value (random effects model)	*P* heterogeneity
17	rs2305480[G]	*GSDMB*	0	0.60	Discovery	2.28 (2.04–2.55)	1.3×10^{-48}	–	–
					Replication 1	1.32 (1.23–1.39)	**6.4×10^{-23}**	**6.4×10^{-23}**	0.86
9	rs928413[G]	*IL33*	2,418	0.28	Discovery	1.50 (1.34–1.67)	4.2×10^{-13}	–	–
					Replication 1	1.24 (1.17–1.32)	**8.8×10^{-13}**	**2.5×10^{-6}**	0.007
5	rs6871536[C]	*RAD50*	0	0.22	Discovery	1.44 (1.28–1.62)	1.8×10^{-9}	–	–
					Replication 1	1.17 (1.10–1.25)	**7.6×10^{-7}**	**7.6×10^{-7}**	0.54
2	rs1558641[G]	*IL1R1*	0	0.85	Discovery	1.56 (1.34–1.81)	6.6×10^{-9}	–	–
					Replication 1	1.11 (1.04–1.19)	**0.003**	**0.003**	0.75
7	rs6967330[A]	*CDHR3*	0	0.19	Discovery	1.45 (1.28–1.66)	1.4×10^{-8}	–	–
					Replication 1	1.18 (1.10–1.27)	**3.0×10^{-6}**	**1.3×10^{-4}**	0.04
					Replication 2	1.40 (1.16–1.67)	**3.2×10^{-4}**	**3.2×10^{-4}**	0.87
					Replications 1 + 2	1.21 (1.13–1.29)	**1.6×10^{-8}**	**2.6×10^{-6}**	0.05
					Discovery + replications 1 + 2	1.26 (1.18–1.33)	**2.7×10^{-14}**	**2.7×10^{-7}**	0.02

Replication *P* values are shown in bold if significant after Bonferroni correction for the five loci tested ($P < 0.01$). Replication 1 results are from a previously published large-scale GWAS of asthma (asthma onset before 16 years; subanalysis of ref. 11). Replication 2 results are from the $COPSAC_{2000}$, MAAS and Generation R cohorts (asthma onset before 6 years). Chr., chromosome.

[a]A fixed-effects model was not applied in the discovery analysis.

© 2013 Nature America, Inc. All rights reserved.

Table 2 Association results for the five genome-wide significant and replicated top SNPs stratified on number of hospitalizations for asthma or acute bronchitis from 0–6 years of age in the discovery cohort

		Number of asthma-related hospitalizations				
		2 *n* = 272	3 *n* = 228	4–5 *n* = 277	6 or more *n* = 358	Association between number of hospitalizations and genotype
SNP effect allele	Nearest gene	OR (95% CI) *P* value	OR (95% CI) *P* value	OR (95% CI) *P* value	OR (95% CI) *P* value	*P* value[a]
rs2305480[G]	*GSDMB*	1.87 (1.54–2.26) 1.5×10^{-10}	2.24 (1.81–2.78) 2.1×10^{-13}	2.24 (1.83–2.73) 1.7×10^{-15}	2.72 (2.26–3.28) 3.5×10^{-27}	0.002
rs928413[G]	*IL33*	1.32 (1.09–1.61) 0.005	1.22 (0.98–1.50) 0.07	1.47 (1.21–1.79) 8.5×10^{-5}	1.91 (1.61–2.26) 6.2×10^{-14}	2.4×10^{-4}
rs6871536[C]	*RAD50*	1.31 (1.06–1.61) 0.01	1.26 (1.00–1.59) 0.05	1.45 (1.18–1.78) 3.6×10^{-4}	1.58 (1.31–1.89) 1.3×10^{-6}	0.09
rs1558641[G]	*IL1R1*	1.53 (1.16–2.02) 0.002	1.20 (0.91–1.57) 0.20	1.32 (1.02–1.71) 0.04	2.19 (1.66–2.90) 3.2×10^{-8}	0.02
rs6967330[A]	*CDHR3*	1.23 (0.98–1,56) 0.07	1.37 (1.07–1.75) 0.01	1.42 (1.13–1.78) 0.003	1.63 (1.33–1.97) 1.6×10^{-6}	0.04

Only the 1,135 children with full follow-up were included. The number of controls was 2,511 for all analyses.
[a]Mantel-Haenszel test for linear association.

© 2013 Nature America, Inc. All rights reserved.

total of 1,091 children. The rs6967330 risk allele (A) was associated with greater risk of asthma hospitalizations (hazards ratio (HR) = 1.7 (95% confidence interval (CI) = 1.2–2.4), $P = 0.002$) and severe exacerbations (HR = 1.4 (95% CI = 1.1–1.9), $P = 0.007$) in combined analyses (**Fig. 2**, **Supplementary Fig. 5** and **Supplementary Table 6**).

In $COPSAC_{2000}$, we observed a trend in the direction of increased neonatal bronchial responsiveness associated with the rs6967330 risk allele ($P = 0.10$) (**Supplementary Table 7**). There was no association with eczema in any of the three birth cohorts, and data on allergic sensitization were inconsistent (**Supplementary Table 6**).

The top SNP at the *CDHR3* locus (rs6967330) is a nonsynonymous coding SNP, where the risk allele (A), corresponding to the minor allele, results in an amino acid change from cysteine to tyrosine at position 529. This SNP is the only known nonsynonymous variant in this linkage disequilibrium (LD) region, but there are other variants located within Encyclopedia of DNA Elements (ENCODE)-predicted regulatory regions that are in moderate to high LD ($r^2 > 0.5$) with the sentinel SNP (**Supplementary Table 8**). Two SNPs with partial LD ($r^2 = 0.71$ and 0.58) were also associated with asthma in the discovery analysis but with less statistical significance. A similar association pattern with rs6967330 as the top SNP was observed in the GABRIEL (replication) study (**Supplementary Fig. 6**) and in the Generation R (replication) subsample of individuals with non-European ancestry (**Supplementary Fig. 7**), suggesting that rs6967330 might be the causal gene variant at this locus.

We investigated the potential functional consequences of the top variant in *CDHR3* (rs6967330; p.Cys529Tyr) by generating an expression construct encoding tagged human CDHR3 and introducing the mutation encoding p.Cys529Tyr (A allele at rs6967330 resulting in mutation of cysteine 529 to tyrosine) by site-directed mutagenesis. We transfected the constructs for wild-type and mutant CDHR3 into 293T cells. Consistent results from six independent experiments involving flow cytometry ($n = 3$) (**Supplementary Fig. 8**) and immunofluorescence staining ($n = 3$) (**Supplementary Fig. 9**) showed that the wild-type protein was expressed at very low levels at the cell surface, whereas the Cys529Tyr mutant showed a marked increase in cell surface expression (**Supplementary Note**). These results support the possibility that rs6967330 represents the causal variant at this locus. A recent study[17] reported that a SNP (rs17152490) in high LD ($r^2 = 0.69$) with our top SNP was associated with lung expression of *CDHR3*, further supporting a functional role for this locus.

CDHR3 is a transmembrane protein with six extracellular cadherin domains. Protein structure modeling showed that the risk-associated alteration (p.Cys529Tyr) was located at the interface between two membrane-proximal cadherin domains, D5 and D6 (**Fig. 3**). Interestingly, Cys592 and Cys566, which are expected to form a disulfide bridge within D6, are close to Cys529 in D5, and the short distance between them could allow disulfide rearrangement (for the wild-type, non-risk cysteine variant). The location of the variant residue at the domain interface suggests that the variant residue may interfere with interdomain stabilization, overall protein stability, folding or conformation, in agreement with the observation in our experimental studies of altered cell surface expression.

The biological function of CDHR3 is unknown, but it belongs to the cadherin family of transmembrane proteins involved in homologous cell adhesion and important for several cellular processes, including epithelial polarity, cell-cell interaction and differentiation[18]. Other members of the cadherin family have been associated with asthma

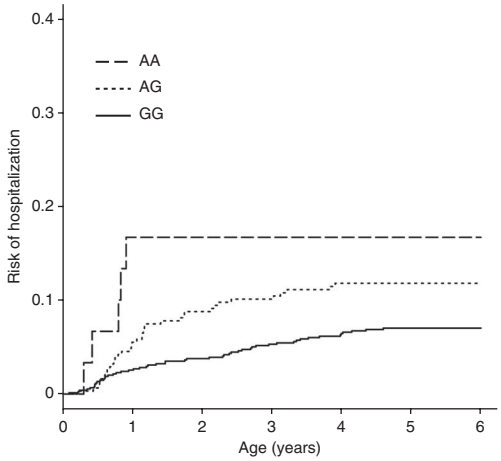

Figure 2 Cumulative risk of asthma hospitalization during the first 6 years of life stratified on *CDHR3* (rs6967330) genotype. Data are from combined analysis of the $COPSAC_{2000}$ and MAAS birth cohorts (replication), including a total of 1,091 children, of whom 92 were hospitalized for asthma. Genotype distribution was as follows: AA, 30 individuals; AG, 312 individuals; GG, 749 individuals. The *P* value for the association between genotype and risk of hospitalization was 0.002 (Cox regression analysis using an additive genetic model).

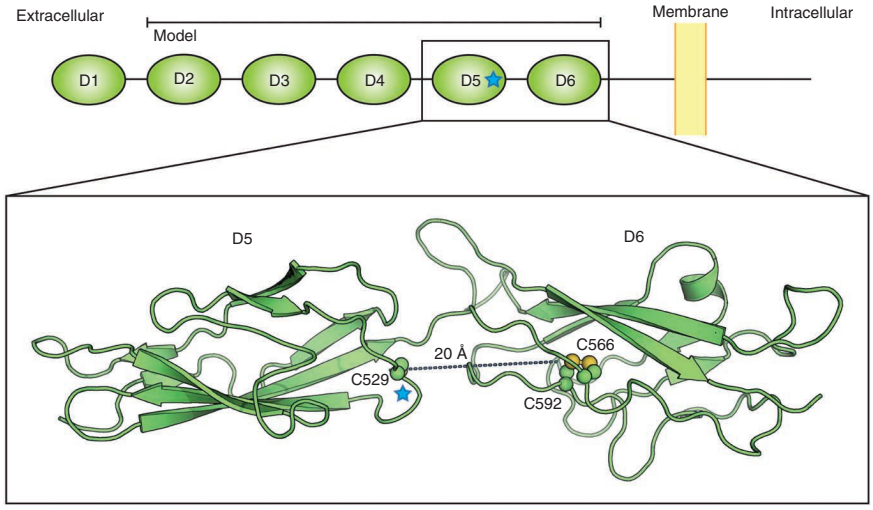

Figure 3 Overview of the CDHR3 protein model. The model covers cadherin domains 2–6 (D2–D6) and is based on the structure of the entire mouse N-cadherin ectodomain (Protein Data Bank (PDB) 3Q2W; domains 1–5). The location of the alteration at position 529 is indicated with a blue star. The distance between residue 529 and the disulfide bridge in D6 (between residues 566 and 592) is approximately 20 Å.

and related traits, including E-cadherin[19] and protocadherin-1 (ref. 20).

We demonstrated protein expression of CDHR3 in bronchial epithelium from adults and in fetal lung tissue (**Supplementary Fig. 10**). *CDHR3* was previously found to be highly expressed in normal human lung tissue[21] and specifically in the bronchial epithelium[22]. *CDHR3* (probe 235650_at) was upregulated by tenfold in differentiating epithelial cells (with a rank of 123 out of more than 47,000 transcripts ranked by magnitude of upregulation)[23] and seems to be highly expressed in the developing human lung[24].

There is an increasing focus on the role of the airway epithelium in asthma pathogenesis. Structural or functional abnormalities in the epithelium may increase susceptibility to environmental stimuli by exaggerating immune responses and structural changes in underlying tissues and increasing airway reactivity[25]. Epithelial integrity is dependent on the interaction of proteins in cell-cell junction complexes, including adhesion molecules. Studies have shown impaired tight junction function[26] and reduced E-cadherin expression[27] in the airway epithelium of individuals with asthma. *CDHR3* is a plausible candidate gene for asthma because of its high level of expression in the airway epithelium and the known role of cadherins in cell adhesion and interaction. Most asthma exacerbations in children are caused by respiratory infections, predominantly common viral infections such as rhinovirus[28], but bacterial infection may also have a role[29], as well as exposure to air pollution[30]. It is therefore plausible that *CDHR3* variation increases susceptibility to respiratory infections or other airway irritants through impaired epithelial integrity and/or disordered repair processes.

Interestingly, the *CDHR3* asthma risk allele is the ancestral allele. Public data from protein databases suggest that humans are unique among 36 other vertebrate species in having the derived (non-risk) allele resulting in a cysteine at position 529 (**Supplementary Table 9**), which is now the wild-type allele in most human populations (Human Genome Diversity Project (HGDP) selection browser; see URLs). This finding suggests that the risk (ancestral) allele, associated with increased surface expression of CDHR3, may have been advantageous during early human evolution. This phenomenon in which the ancestral allele is the risk allele is known for other common diseases and may reflect a shift from a beneficial to a deleterious effect for a particular allele as a result of a changing environment[31].

The *CDHR3* variant seems to be associated with an asthma phenotype of early onset, as demonstrated by the strongest replication of association in the GABRIEL stratum with asthma onset before 16 years of age (**Supplementary Table 10**) and in the second replication including children with asthma onset before 6 years of age (**Table 1**). Increased risk was already demonstrated in the first year of life (**Fig. 2**), particularly in children who were homozygous for the risk allele (A). This finding is in line with the tendency toward association of increased airway reactivity in neonates with the risk allele and findings of *CDHR3* expression in the fetal lung. *CDHR3* variation also seems to be more strongly associated with an asthma phenotype with exacerbations (**Supplementary Table 6**), particularly with recurrent exacerbations (**Table 2** and **Supplementary Table 6**).

The top locus in this study, on chromosome 17q12-21, has consistently been associated with childhood-onset asthma[11,13]. The effect size in the present study is remarkably high, with an OR of 2.3 that increases to 2.7 for the children with the highest number of exacerbations. This finding suggests a key role for this locus in severe exacerbations in early childhood, in line with a previous report from the $COPSAC_{2000}$ birth cohort study[32].

Genome-wide significant association with asthma has previously been shown for variants in or near *IL33*, *RAD50-IL13* and *IL1RL1* (refs. 11,33). The fact that the top loci in our study were generally shared with previous GWAS of asthma suggests that early-onset asthma with severe exacerbations is at least partly driven by multiple common variants in the same genes that contribute to asthma without severe exacerbations.

The sample size of the present GWAS was less than one-fifth that of the largest published GWAS of asthma (GABRIEL)[11], and, yet, we found a similar number of genome-wide significant loci, similar statistical significance and considerably larger effect estimates. Further increasing phenotypic specificity by stratified analysis in the 358 children with the highest number of exacerbations resulted in an additional increase in effect estimates, with ORs between 1.6 and 2.7 per risk allele, and strong statistical significance. Effect estimates were also higher than previously reported when replicating the exact top SNP from the GABRIEL study (**Supplementary Table 11**). This finding demonstrates that specific phenotyping is a helpful approach in the search for asthma susceptibility genes. The narrow age criteria (2–6 years) for disease may be an important phenotypic characteristic, as heritability has been demonstrated to be higher for early-onset asthma[34].

The method of case identification through national registries allowed us to define a specific and rare phenotype of repeated acute hospitalizations in young children from 2 to 6 years of age, which, to our knowledge, has not previously been done in a GWAS. One limitation of this study is that we had relatively poor genome-wide coverage (approximately 125,000 SNPs).

In conclusion, our results demonstrate the strength of specific phenotyping in genetic studies of asthma. Future research focusing on understanding the role of *CDHR3* variants in the development of asthma and severe exacerbations may increase understanding and improve treatment of this clinically important disease entity.

© 2013 Nature America, Inc. All rights reserved.

URLs. HGDP selection browser data for rs6967330, http://hgdp.uchicago.edu/cgi-bin/alfreqs.cgi?pos=105445687&chr=chr7&rs=rs6967330&imp=false.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

A full list of acknowledgments for each study is given in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

K.B. was the main author responsible for designing the study, analyzing and interpreting data, writing the manuscript and directing the work. He had full access to the data and final responsibility for the decision to submit the work for publication. H.B. contributed to design of the study, analysis of data and writing of the manuscript. P.S. and H.H. contributed to design of the study and analysis of data in relation to whole-genome genotyping. K.N. performed the GWAS analysis and contributed to regional imputation. E.K.-M., A. Sevelsted, M.A.R., R.Y. and R.G. contributed to data analysis. J.M.M., S.B.-G. and D.T. directed and contributed to regional imputation and data analyses. M.V.H. and D.M.H. were responsible for subject identification, collection of dried blood spots and DNA extraction and amplification. K.B., E.K.-M., L.J.M., R.F. and A.M. contributed to data acquisition. T.B. performed modeling of the CDHR3 protein structure. L.P., C.H. and E.A.N. were responsible for data from the discovery control cohort. H.H. and M.E.M. were responsible for the functional studies of the CDHR3 variant involving flow cytometry. A.H., D.E.S. and D.E.D. were responsible for the experimental studies involving immunofluorescence staining. A. Simpson, A.C. and D.B. were responsible for data from the MAAS cohort. H.T.d.D., L.D. and V.W.V.J. were responsible for data from the Generation R cohort. G.F.-T., P.M.L. and J.W.H. were responsible for the studies of lung tissue. P.F.T. studied the evolutionary aspects of the *CDHR3* risk variant (rs6967330). All coauthors provided important intellectual input to the study and approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Kocevar, V.S. *et al.* Variations in pediatric asthma hospitalization rates and costs between and within Nordic countries. *Chest* **125**, 1680–1684 (2004).
2. Lozano, P., Sullivan, S.D., Smith, D.H. & Weiss, K.B. The economic burden of asthma in US children: estimates from the National Medical Expenditure Survey. *J. Allergy Clin. Immunol.* **104**, 957–963 (1999).
3. Matterne, U., Schmitt, J., Diepgen, T.L. & Apfelbacher, C. Children and adolescents' health-related quality of life in relation to eczema, asthma and hay fever: results from a population-based cross-sectional study. *Qual. Life Res.* **20**, 1295–1305 (2011).
4. Smith, D.H. *et al.* A national estimate of the economic costs of asthma. *Am. J. Respir. Crit. Care Med.* **156**, 787–793 (1997).
5. Bush, A. Practice imperfect—treatment for wheezing in preschoolers. *N. Engl. J. Med.* **360**, 409–410 (2009).
6. 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).
7. van Beijsterveldt, C.E. & Boomsma, D.I. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur. Respir. J.* **29**, 516–521 (2007).
8. Ferreira, M.A. *et al.* Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (2011).
9. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
10. Himes, B.E. *et al.* Genome-wide association analysis identifies *PDE4D* as an asthma-susceptibility gene. *Am. J. Hum. Genet.* **84**, 581–593 (2009).
11. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
12. Sleiman, P.M. *et al.* Variants of *DENND1B* associated with asthma in children. *N. Engl. J. Med.* **362**, 36–44 (2010).
13. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* **43**, 887–892 (2011).
14. Anderson, G.P. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* **372**, 1107–1119 (2008).
15. Hollegaard, M.V. *et al.* Genome-wide scans using archived neonatal dried blood spot samples. *BMC Genomics* **10**, 297 (2009).
16. Hollegaard, M.V. *et al.* Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet.* **12**, 58 (2011).
17. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8**, e1003029 (2012).
18. Hulpiau, P. & van Roy, F. Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41**, 349–369 (2009).
19. Nawijn, M.C., Hackett, T.L., Postma, D.S., van Oosterhout, A.J. & Heijink, I.H. E-cadherin: gatekeeper of airway mucosa and allergic sensitization. *Trends Immunol.* **32**, 248–255 (2011).
20. Koppelman, G.H. *et al.* Identification of *PCDH1* as a novel susceptibility gene for bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **180**, 929–935 (2009).
21. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
22. McCall, M.N., Uppal, K., Jaffee, H.A., Zilliox, M.J. & Irizarry, R.A. The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res.* **39**, D1011–D1015 (2011).
23. Ross, A.J., Dailey, L.A., Brighton, L.E. & Devlin, R.B. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **37**, 169–185 (2007).
24. Kho, A.T. *et al.* Transcriptomic analysis of human lung development. *Am. J. Respir. Crit. Care Med.* **181**, 54–63 (2010).
25. Holgate, S.T. The sentinel role of the airway epithelium in asthma pathogenesis. *Immunol. Rev.* **242**, 205–219 (2011).
26. Xiao, C. *et al.* Defective epithelial barrier function in asthma. *J. Allergy Clin. Immunol.* **128**, 549–556 (2011).
27. de Boer, W.I. *et al.* Altered expression of epithelial junctional proteins in atopic asthma: possible role in inflammation. *Can. J. Physiol. Pharmacol.* **86**, 105–112 (2008).
28. Johnston, S.L. *et al.* Community study of role of viral infections in exacerbations of asthma in 9–11 year old children. *Br. Med. J.* **310**, 1225–1229 (1995).
29. Bisgaard, H. *et al.* Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. *Br. Med. J.* **341**, c4978 (2010).
30. Iskandar, A. *et al.* Coarse and fine particles but not ultrafine particles in urban air trigger hospital admission for asthma in children. *Thorax* **67**, 252–257 (2012).
31. Di Rienzo, A. & Hudson, R.R. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet.* **21**, 596–601 (2005).
32. Bisgaard, H. *et al.* Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am. J. Respir. Crit. Care Med.* **179**, 179–185 (2009).
33. Li, X. *et al.* Genome-wide association study of asthma identifies *RAD50-IL13* and *HLA-DR/DQ* regions. *J. Allergy Clin. Immunol.* **125**, 328–335 (2010).
34. Thomsen, S.F., Duffy, D.L., Kyvik, K.O. & Backer, V. Genetic influence on the age at onset of asthma: a twin study. *J. Allergy Clin. Immunol.* **126**, 626–630 (2010).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

The individual studies are described in further detail in the **Supplementary Note**.

$COPSAC_{exacerbation}$ cohort (GWAS). This is a register-based cohort of children with asthma who were identified and characterized from national health registries. The study was approved by the Ethics Committee for Copenhagen (H-B-2998-103) and the Danish Data Protection Agency (2008-41-2622). According to Danish law, research ethics committees can grant exemption from obtaining informed consent for research projects based on biobank material under certain circumstances. For this study, such an exemption was granted (H-B-2998-103).

Case selection. Children with repeated acute hospitalizations (cases) were identified in the Danish National Patient Register covering all diagnoses of discharges from Danish hospitals[35]. Information on birth-related events was obtained from the national birth register. Inclusion criteria were at least two acute hospitalizations for asthma (ICD8-codes 493, ICD-10 codes J45-46) from 2 to 6 years of age (both years included). Duration of hospitalization had to be more than 1 d, and two hospitalizations had to be separated by at least 6 months. Exclusion criteria were side diagnosis during hospitalization, registered chronic diagnosis considered to affect risk of hospitalization for asthma, low birth weight (<2.5 kg) or gestational age of under 36 weeks at birth. Cases were further characterized with respect to the number of hospitalizations from asthma and acute bronchitis and for concurrent atopy.

DNA sampling and genotyping of cases. DNA was obtained from blood spots sampled as part of the Danish neonatal screening program and stored in the Danish Neonatal Screening Biobank[36]. Two disks, each 3.2 mm in diameter, were punched from each blood spot. DNA was extracted, and the whole genome for each individual sample was amplified in triplicate as previously described[15,16]. Cases were genotyped on the Affymetrix Axiom CEU array (567,090 SNPs). Top SNPs from the five genome-wide significant loci were regenotyped with the PCR KASPar genotyping system (KBiosciences) to validate the results (**Supplementary Table 3**). Two additional SNPs in the proximity of the newly discovered *CDHR3* variant were genotyped for further exploration of the region encompassing it.

Controls. The control population was randomly drawn from two large Danish cohorts: the Danish National Birth Cohort (females) and the Copenhagen draft board examinations (males). Individuals who indicated in a questionnaire that they had physician-diagnosed asthma were excluded. Genome-wide genotyping had previously been performed as part of the Genomics of Overweight in Young Adults (GOYA) study[37] on the Illumina Human610-Quad v1.0 BeadChip (545,350 SNPs). Potential bias introduced by differences in chemistry between the different platforms used for cases and controls (Affymetrix and Illumina, respectively) was investigated by also using control data from the Wellcome Trust Case Control Consortium 2 (WTCCC2) project that performed genotyping on an Affymetrix platform (Affymetrix 6.0) (**Supplementary Table 4**).

Replication in a previously published GWAS. Replication of the five genome-wide significant loci from the discovery analysis was sought in publically available data from a GWAS performed by the GABRIEL Consortium[11]. This replication included 19 studies of childhood-onset asthma (onset before 16 years of age) with a total of 6,783 cases and 7,720 controls.

Replication in birth cohorts for the *CDHR3* top SNP. *The $COPSAC_{2000}$ replication cohort*. Replication and phenotypic characterization of the *CDHR3* risk locus were sought in the $COPSAC_{2000}$ cohort, a prospective clinical study of a birth cohort of 411 children. This cohort is not overlapping with the $COPSAC_{exacerbation}$ discovery study. The $COPSAC_{2000}$ cohort study was approved by the Ethics Committee for Copenhagen (KF 01-289/96) and the Danish Data Protection Agency (2008-41-1754), and informed consent was obtained from both parents of each child. All mothers had a history of a doctor's diagnosis of asthma after 7 years of age. Newborns were enrolled in the first month of life, as previously described in detail[38–40]. This cohort is characterized by deep phenotyping during close clinical follow-up. Doctors employed in the clinical research unit were acting primary physicians for the children from the cohort and diagnosed and treated respiratory and skin symptoms, and asthmatic symptoms were recorded in daily diaries[41].

Acute, severe exacerbations from birth to 6 years of age were defined as requiring the use of oral prednisolone or high-dose inhaled corticosteroid for wheezy symptoms, prescribed at the discretion of the doctor in the clinical research unit, or by acute hospitalization at a local hospital for such symptoms[32]. Asthma from birth to 7 years of age was diagnosed on the basis of predefined algorithms of symptoms and response to treatment, as previously described[40].

Neonatal spirometry and analysis of neonatal bronchial responsiveness to methacholine were carried out by 4 weeks of age, applying the raised volume, rapid thoracic compression technique. Lung function was measured by spirometry in the child's seventh year of life. Specific airway resistance (sR_{aw}) was measured at 4 and 6 years by whole-body plethysmography. Bronchial responsiveness at ages 4 and 6 years was determined as the relative change in sR_{aw} after hyperventilation of cold, dry air.

Allergic sensitization against common inhalant allergens was determined at 6 years of age by measurement of serum-specific IgE levels. Atopic dermatitis was diagnosed using the Hanifin-Rajka criteria[42] from birth to 7 years of age.

High-throughput genome-wide SNP genotyping was performed using the Illumina Infinium II HumanHap550 v1, v3 or Quad BeadChip platform at the Children's Hospital of Philadelphia's Center for Applied Genomics. We excluded SNPs with call rate of <95%, minor allele frequency (MAF) of <1% or Hardy-Weinberg equilibrium P value of $<1 \times 10^{-5}$. rs6967330 was a genotyped SNP on this array.

MAAS replication cohort. The Manchester Asthma and Allergy Study is a population-based birth cohort described in detail elsewhere[43]. Subjects were recruited prenatally and were followed prospectively. The study was approved by the local research ethics committee (South Manchester, reference 03/SM/400). Parents gave written informed consent. Participants attended follow-up at ages 1, 3 and 5 years of age.

For asthma, validated questionnaires were administered by interviewers to collect information on parentally reported symptoms, physician-diagnosed asthma and treatments received. 'Current wheeze and asthma treatment' was defined as parentally reported wheeze in the past 12 months. 'Asthma ever' was defined as positive if, at any given time point, two of three responses were positive to the following questions: "Has your child wheezed within the past 12 months?", "Does your child currently take asthma medication?" or "Has a doctor ever told you that your child has asthma?" Controls were defined as children with none of these symptoms.

For exacerbations, a pediatrician extracted data from primary-care medical records, including information on diagnosis with wheeze and/or asthma, all prescriptions (including inhaled corticosteroids (ICS) and β2 agonists), unscheduled visits and hospital admissions for asthma and/or wheeze during the first 8 years of life. Following American Thoracic Society guidelines, we defined asthma exacerbations by either admission to a hospital or an emergency department visit and/or by receipt of oral corticosteroids for at least 3 d[44].

DNA samples were genotyped on the Illumina Human610-Quad BeadChip. Genotypes were called using the Illumina GenCall application, following the manufacturer's instructions. Quality control criteria for samples included call rate of greater than 97%, exclusion of samples with outlier autosomal heterozygosity and sex validation. We excluded SNPs with call rate of <95%, Hardy-Weinberg equilibrium P value of $>5.9 \times 10^{-7}$ and MAF of <0.005. We then performed a look-up for SNP rs6967330, which showed a genotyping success rate of 100% and a Hardy-Weinberg equilibrium P value of 0.4164.

Generation R replication cohort. The Generation R Study is a population-based prospective cohort study of pregnant women and their children from fetal life onward in Rotterdam, The Netherlands[45]. The study protocol was approved by the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam (MEC 217.595/2002/20). Written informed consent was obtained from all mothers and biological fathers or legal guardians. Information on wheezing, asthma and eczema was collected for the children by questionnaires at the ages of 1 to 4 and 6 years[46]. Questions about wheezing included: "Has your child had problems with a wheezing chest during the last year? (never, 1–3 times, >4 times) (age 1 to 4 years)" and "Did your child ever suffer from chest wheezing? (never, 1–3 times,

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2830

>4 times) (age 6 years)." Questions about asthma included: "Has a doctor diagnosed your child as having asthma during the past year? (yes, no) (age 2 and 4 years)" and "Was your child ever diagnosed with asthma by a doctor? (yes, no) (age 3 and 6 years)." On the basis of the last obtained questionnaire, we grouped children as having 'asthma ever before 6 years of age'. Reported asthma at 2, 3 or 4 years of age was used to reclassify children included in this group where appropriate. We then recategorized children as those with an asthma diagnosis before 3 years of age and at 3 years of age or older. Reported numbers of wheezing episodes at 1 and 2 years of age and at 3 to 6 years of age, respectively, were used to reclassify asthma diagnosis before and at 3 years of age into 'asthma diagnosis or ≥3 episodes of wheezing before 3 years of age'. Questions about eczema included: "Has a doctor diagnosed your child as having eczema during the past year? (yes, no) (age 1 to 4 years)" and "Was your child ever diagnosed with eczema by a doctor? (yes, no) (6 years)." As with asthma, we grouped children into those with 'eczema ever before 6 years of age' on the basis of the last obtained questionnaire and used reported eczema at 1 or 4 years of age to reclassify children included in this group where appropriate.

Samples were genotyped using Illumina Infinium II HumanHap610 Quad arrays, following standard manufacturer's protocols. Intensity files were analyzed using BeadStudio Genotyping Module software v.3.2.32, and genotypes were called using default cluster files. Any sample with a call rate of less than 97.5%, excess autosomal heterozygosity (F < mean – 4 s.d.) or mismatch between called and phenotypic sex was excluded. rs6967330 was a genotyped SNP in this set. Individuals identified as genetic outliers by identity-by-state (IBS) clustering analysis (>3 s.d. away from the mean for the HapMap CEU population (Utah residents of Northern and Western European ancestry)) were considered to have non-European ancestry. Ancestry determination analysis included genomic data from all Generation R individuals merged with data for three reference panels from Phase 2 of the HapMap Project (YRI (Yoruba from Ibadan, Nigeria), CHB + JPT (Han Chinese in Beijing, China, and Japanese in Tokyo, Japan) and CEU). Analysis of association between an asthma or eczema phenotype and GWAS SNPs was carried out using a regression framework, adjusting for population stratification in the Generation R cohort using MACH2QTL, as implemented in GRIMP. Ten genomic principal components obtained after the application of SNP quality exclusion criteria and LD pruning were used to adjust for population substructure in the combined population, four principal components were used for the European subpopulation and eight principal components were used for the non-European subpopulation. Individuals were grouped as having European (n = 1,962; 64.5%) or non-European (n = 1,078; 35.5%) ancestry on the basis of genetic ancestry. On the basis of information on the country of birth of parents and grandparents obtained by questionnaires, the largest non-European ancestry groups included individuals of Turkish (5.4%), Surinamese (4.6%), Dutch Antillean (4.0%), Moroccan (2.9%) and Cape Verdean (2.3%) origin.

Statistical analyses. *Genome-wide association analysis.* Quality control was carried out separately on cases and controls. This included filtering on SNP call rate (>99%) and sample call rate (>98%) and tests for excess heterozygosity, deviation from Hardy-Weinberg equilibrium, sex mismatch and familial relatedness. Non-European individuals were excluded on the basis of deviation from the HapMap CEU reference panel (release 22). Indication of population stratification or genotyping bias was tested by multidimensional scaling (MDS) after quality control. This analysis showed evidence of association with disease status for the first seven MDS components, and these were therefore included as covariates in the association analysis. Additional analyses including the first 100 MDS components did not materially alter the results. Merged data for SNPs present on both arrays after quality control were used for association testing with PLINK (v. 1.07) using a logistic additive model, adjusting for the first seven MDS components. Additional quality control was performed for genome-wide significant SNPs after association analysis, including a test for genotyping batch effects, resulting in the removal of one genome-wide significant SNP with strong evidence of batch-related genotyping error.

Functional annotation for the SNPs in LD (r^2 > 0.5) with the *CDHR3* top SNP (rs6967330) was obtained from the RefSeq track downloaded from the UCSC Genome Browser. SNPs were associated with regulatory elements by HaploReg[47] in terms of predicted ENCODE chromatin state (promoter and enhancer histone modification signals) and DNase I hypersensitivity (**Supplementary Table 8**).

Regional imputation was performed to describe the identified loci from the discovery analysis (**Supplementary Fig. 3**) as well as reported loci from the previous largest published GWAS (GABRIEL)[11] (**Supplementary Table 11**). We used two-step genotype imputation as described[48]. We used the SHAPEIT algorithm to prephase the haplotypes[49] and then used IMPUEv2 software for the imputation of unknown genotypes[50] separately in cases and controls. We used the 1000 Genomes Project reference panel[51] (April 2012 version). We used a strict cutoff (info of 0.88), which, according to our analyses, provides an allelic dosage R^2 correlation between real and imputed genotypes of greater than 0.8 and shows an optimal balance between sufficient accuracy and power[52]. We then compared the resulting allelic frequencies using SNPTEST 2.4.1 (ref. 53).

CDHR3 protein expression in experimental models. The top SNP at the *CDHR3* locus is a nonsynonymous SNP (encoding p.Cys529Tyr). To determine the functional consequences of the p.Cys529Tyr variant, we generated expression constructs encoding tagged human CDHR3 protein, and the mutation encoding the p.Cys529Tyr alteration was introduced by site-directed mutagenesis. Plasmids encoding wild-type or mutant *CDHR3* or empty vector were transfected into 293T cells, and cells were monitored for surface and intracellular expression of CDHR3 by flow cytometry. 293T cells were from the American Type Culture Collection (ATCC), catalog number CRL-3216. They were recently tested for mycoplasma contamination but were not authenticated. For protein blotting, cells expressing CDHR3 proteins were lysed, and whole-cell lysates were separated by SDS-PAGE under reducing or non-reducing conditions, transferred to PVDF membranes and blotted for Flag (anti-Flag antibody, clone M2 (Agilent Technologies, 200470-21) at a dilution of 1:2,000). For immunofluorescence and confocal microscopy, 293T cells were grown on glass coverslips in DMEM with 3 mM glutamine and 10% heat-inactivated FBS at 37 °C and 5% CO_2 before and for 2 d after transfection with expression constructs for Flag-tagged wild-type CDHR3 and CDHR3 Cys529Tyr using TransIT 2020 reagent according to a standard protocol (Mirus Bio). Cells were obtained and used at a low passage from ATCC and had recently been tested for mycoplasma. Cells were incubated in 10% serum-containing culture medium plus primary anti-Flag mouse antibodies (F3165, Sigma; 1:300 dilution) for 1 h at 37 °C before being washed briefly with culture medium. Cells were then stained with secondary rabbit anti-mouse antibodies (F0261, Daco; 1:600 dilution) conjugated with fluorescein isothiocyanate (FITC) with incubation at 37 °C for 30 min and washed with culture medium before PBS. Afterward, cells were fixed in 2% paraformaldehyde for 15 min, washed with PBS and permeabilized in 0.2% Triton X-100 in PBS for 5 min, washed and incubated with Cy3-conjugated mouse anti-Flag antibody (Cy3-labeled F3165, Sigma; 1:300 dilution). Finally, cells were mounted with ProLong Gold antifade reagent with DAPI (Invitrogen). Images were acquired using a Leica DMI 6000-B confocal microscope (Leica Microsystems) with 40× magnification and were processed in Photoshop (Adobe Systems). Experiments were performed in triplicate (independent transfections) for both flow cytometry and immunofluorescence staining. Data presented (**Supplementary Figs. 8** and **9**) were chosen as being representative of the repeated experiments.

CDHR3 protein structure modeling. A homology model of CDHR3 domains 2–6 (residues 141–681) was generated using the HHpred server[54]. The model was based on the structure of mouse N-cadherin (PDB 3Q2W) domains 1–5. A disulfide bridge was manually introduced in the final model between the structurally adjacent residues Cys566 and Cys592, as this corresponds to a disulfide bridge commonly observed in cadherin domains.

35. Lynge, E., Sandegaard, J.L. & Rebolj, M. The Danish National Patient Register. *Scand. J. Public Health* **39**, 30–33 (2011).
36. Nørgaard-Pedersen, B. & Hougaard, D.M. Storage policies and use of the Danish Newborn Screening Biobank. *J. Inherit. Metab. Dis.* **30**, 530–536 (2007).
37. Paternoster, L. *et al.* Genome-wide population-based association study of extremely overweight young adults—the GOYA study. *PLoS One* **6**, e24303 (2011).
38. Bisgaard, H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann. Allergy Asthma Immunol.* **93**, 381–389 (2004).

© 2013 Nature America, Inc. All rights reserved.

39. Bisgaard, H., Hermansen, M.N., Loland, L., Halkjaer, L.B. & Buchvald, F. Intermittent inhaled corticosteroids in infants with episodic wheezing. *N. Engl. J. Med.* **354**, 1998–2005 (2006).
40. Bisgaard, H. *et al.* Childhood asthma after bacterial colonization of the airway in neonates. *N. Engl. J. Med.* **357**, 1487–1495 (2007).
41. Bisgaard, H., Pipper, C.B. & Bonnelykke, K. Endotyping early childhood asthma by quantitative symptom assessment. *J. Allergy Clin. Immunol.* **127**, 1155–1164 (2011).
42. Hanifin, J.M. & Rajka, G. Diagnostic features of atopic dermatitis. *Acta Derm. Venereol.* **92**, 44–47 (1980).
43. Lowe, L. *et al.* Specific airway resistance in 3-year-old children: a prospective cohort study. *Lancet* **359**, 1904–1908 (2002).
44. Reddel, H.K. *et al.* An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am. J. Respir. Crit. Care Med.* **180**, 59–99 (2009).
45. Jaddoe, V.W. *et al.* The Generation R Study Biobank: a resource for epidemiological studies in children and their parents. *Eur. J. Epidemiol.* **22**, 917–923 (2007).
46. Jaddoe, V.W. *et al.* The Generation R Study: design and cohort update 2012. *Eur. J. Epidemiol.* **27**, 739–756 (2012).
47. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
48. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
49. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
50. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
51. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
52. Auer, P.L. *et al.* Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *Am. J. Hum. Genet.* **91**, 794–808 (2012).
53. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
54. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2830