Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY

EAACI

Allergy

ORIGINAL ARTICLE EPIDEMIOLOGY AND GENETICS

A common atopy-associated variant in the Th2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SP1 binding

A. Kretschmer[1], G. Möller[2], H. Lee[3,4,5,6], H. Laumen[3,4,5,6,7], C. von Toerne[8], K. Schramm[9,10], H. Prokisch[9,10], S. Eyerich[11], S. Wahl[1], H. Baurecht[12], A. Franke[13], M. Claussnitzer[3,4,5,6], K. Eyerich[14], A. Teumer[15], L. Milani[16], N. Klopp[1,17], S. M. Hauck[8], T. Illig[1,17], A. Peters[1,18], M. Waldenberger[1], J. Adamski[2,19], E. Reischl[1,*] & S. Weidinger[12,*]

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München; [2]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg; [3]Else Kroener-Fresenius-Center for Nutritional Medicine, Chair of Nutritional Medicine, Technische Universität München; [4]ZIEL – Research Center for Nutrition and Food Sciences, Technische Universität München; [5]Clinical Cooperation Group Nutrigenomics and type 2 diabetes at the Technische Universität München, Freising and the Helmholtz Zentrum München, Neuherberg; [6]DZD – German Center for Diabetes Research, Neuherberg; [7]Institute of Experimental Genetics, Helmholtz Zentrum München; [8]Research Unit Protein Science, Helmholtz Zentrum München; [9]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg; [10]Institute of Human Genetics, Technische Universität München; [11]Zentrum Allergie und Umwelt (ZAUM), Center for Allergy and Environment, Technische Universität and Helmholtz Zentrum München, Munich; [12]Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein; [13]Institute of Clinical Molecular Biology, University Hospital Schleswig-Holstein, Kiel; [14]Department of Dermatology and Allergology, Technische Universität München, Munich; [15]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; [16]Estonian Genome Center, University of Tartu, Tartu, Estonia; [17]Hannover Unified Biobank, Hannover Medical School, Hannover; [18]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg; [19]Chair of Experimental Genetics, Technische Universität München, Munich, Germany

To cite this article: Kretschmer A, Möller G, Lee H, Laumen H, von Toerne C, Schramm K, Prokisch H, Eyerich S, Wahl S, Baurecht H, Franke A, Claussnitzer M, Eyerich K, Teumer A, Milani L, Klopp N, Hauck SM, Illig T, Peters A, Waldenberger M, Adamski J, Reischl E, Weidinger S. A common atopy-associated variant in the Th2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SP1 binding. *Allergy* 2014; DOI:10.1111/all.12394.

Keywords
atopic diseases; functional SNP; *RAD50*; RHS7; rs2240032.

Correspondence
Stephan Weidinger, Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein, Campus Kiel, Schittenhelmstrasse 7, Kiel 24105, Germany.
Tel.: 0049-431-597-2623 or -2732
Fax: 0049-431-597-1815
E-mail: sweidinger@dermatology.uni-kiel.de

*Both authors are contributed equally.

Accepted for publication 12 February 2014

DOI:10.1111/all.12394

Edited by: Hans-Uwe Simon

Abstract

Background: Type 2 immune responses directed by Th2 cells and characterized by the signature cytokines IL4, IL5, and IL13 play major pathogenic roles in atopic diseases. Single nucleotide polymorphisms in the human Th2 cytokine locus in particular in a locus control region within the DNA repair gene *RAD50*, containing several *RAD50* DNase1-hypersensitive sites (RHS), have been robustly associated with atopic traits in genome-wide association studies (GWAS). Functional variants in *IL13* have been intensely studied, whereas no causative variants for the *IL13*-independent *RAD50* signal have been identified yet. This study aimed to characterize the functional impact of the atopy-associated polymorphism rs2240032 located in the human RHS7 on *cis*-regulatory activity and differential binding of transcription factors.
Methods: Differential transcription factor binding was analyzed by electrophoretic mobility shift assays (EMSAs) with Jurkat T-cell nuclear extracts. Identification of differentially binding factors was performed using mass spectrometry (LC-MS/MS). Reporter vector constructs carrying either the major or minor allele of rs2240032 were tested for regulating transcriptional activity in Jurkat and HeLa cells.
Results: The variant rs2240032 impacts transcriptional activity and allele-specific binding of SMAD3, SP1, and additional putative protein complex partners. We further demonstrate that rs2240032 is located in an RHS7 subunit which itself encompasses repressor activity and might be important for the fine-tuning of transcription regulation within this region.
Conclusion: The human RHS7 critically contributes to the regulation of gene transcription, and the common atopy-associated polymorphism rs2240032 impacts transcriptional activity and transcription factor binding.

Atopic diseases (atopic dermatitis, asthma, rhinitis) are strongly heritable and closely related traits which typically track in the same patients, same families and which are associated with Th2-mediated adaptive responses (1, 2). Besides the detection of disease-specific susceptibility loci, gene mapping approaches have identified loci shared between atopic and other immune-mediated diseases, which could represent important checkpoints in branching pathophysiological pathways involved in multiple diseases (3, 4). One of such regions is 5q31, where the genes encoding the Th2 cytokines *IL4*, *IL5*, and *IL13* are clustered with the constitutively expressed DNA repair gene *RAD50* (Fig. 1) (5). It is well established that single nucleotide polymorphisms (SNPs) in the *IL13* locus are associated with total IgE levels and atopic diseases (5). Interestingly, GWAS for asthma, total IgE, and atopic dermatitis identified strong and apparently *IL13*-independent signals from SNPs within the *RAD50* locus (6–9), although no functions have been described for the RAD50 protein for the development of atopic diseases. Follow-up studies combining the human-associated *RAD50* variants with functional characterization have not been performed though. Results from mouse models have shown the existence of a locus control region (LCR), which resides in a 25-kb region 3′ of *Rad50* and ~20-kb 5′ of *IL13*, the core of which is constituted by four *Rad50*-hypersensitive sites (RHS) in the introns 21 (RHS4-6) and 24 (RHS7) (10, 11). This LCR coordinates the expression of the neighboring genes *IL4* and *IL13* in Th2 cells via intrachromosomal DNA looping, which brings the regulatory elements in close proximity to the *IL4*, *IL5,* and *IL13* promoters (12). RHS7-deficient mice exhibit reduced Th2 cytokine expression (11). Deletion of the entire LCR leads to a dramatic reduction in Th2 cytokines as well as IgE levels and causes a reduced airway hyper-responsiveness in mice (10). Interestingly, the human conserved RHS7 sequence harbors a common SNP, rs2240032 (major allele C, minor allele T, minor allele frequency: 20%), which has robustly been associated with total serum IgE levels and asthma (6–8) and is in strong linkage disequilibrium (LD) with other atopy-associated SNPs.

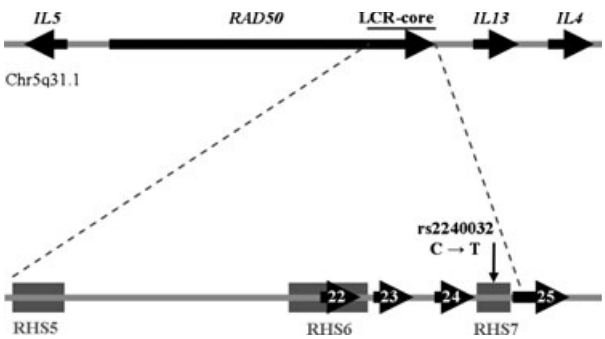

Figure 1 Schematic overview of the human Th2 cytokine locus and a zoom-in of the core of the locus control region at the 3′ end of *RAD50* containing the SNP rs2240032 (C→T) within the RHS7. Numbered arrows = exon with exon number, gray boxes = human conserved RHS.

Materials and methods

Linkage disequilibrium analysis and computational analysis of regulatory elements

To prioritize potentially functional *RAD50*-LCR SNPs for further analysis, we re-analyzed the complex LD pattern of atopy-associated SNPs at this locus using GWAS data from the two population-based German cohorts KORA (13) and PopGen (14) (n = 2551). After extensive quality control, we imputed them using IMPUTE2 (15) and the 1000 genomes reference panel (integrated variant set, release March 2012) (16). SNPs with low imputation quality (info score < 0.4), call rate <95%, deviation from HWE (P < 10E-8), or MAF < 5% were excluded. Linkage disequilibrium was calculated using GenABEL (17). In addition, the presence of transcription factor-binding sites at each SNP position was predicted by computational analysis using the Genomatix SNPInspector and MatInspector software (18). DNaseI-hypersensitive sites, histone marks for regulatory regions (H3K4Me1), ChIP-Seq data and sequence similarity between species were analyzed by using the 'ENCyclopedia of DNA Elements' (ENCODE) database (19) and the UCSC Genome Browser (20).

Cultured cells

T cells and epithelial cells are the main cell types implicated in the pathogenesis of atopic diseases. The human Jurkat T-cell line is widely used as surrogate cell type to investigate T-cell-specific effects (21, 22) and was used in our study to characterize mechanisms at the Th2 cytokine locus. In addition, the HeLa cell line was used, which is a commonly used nonlymphocytic, epithelial cell line, where T-cell-specific effects are not expected. Jurkat cells were cultivated in RPMI medium (Roswell Park Memorial Institute medium) supplemented with L-glutamine (Invitrogen, Carlsbad, CA, USA), and HeLa cells were cultivated in MEM (minimum essential medium). Both media were supplemented with 10% FBS (PAA Laboratories, Pasching, Austria), 100 IU/ml penicillin, and 100 μg/ml streptomycin (Gibco, Grand Island, NY, USA). Cell lines were grown at 37°C and 5% CO_2 in a humidified atmosphere.

Nuclear protein extraction and electrophoretic mobility shift assay

To investigate protein–DNA interaction, EMSA experiments with Cy5-labeled allele-specific oligonucleotides for rs2240032 were performed using nuclear protein extracts prepared from cells. Jurkat cells (DSMZ accession number: ACC282) and $CD2^+$ T cells were stimulated for 6 h with 25 ng/ml phorbol myristate acetate (PMA) and 500 ng/ml ionomycin. Nuclear extracts were prepared from stimulated Jurkat cells and $CD2^+$ T cells using the Nuclear Extraction Kit (Active Motif, Carlsbad, CA, USA) and the ProteoJET™ Cytoplasmic and Nuclear Protein Extraction Kit (Fermentas, Vilnius, Lithuania), respectively. Nuclear extracts from unstimulated HeLa cells were prepared using the NE-PER Kit® Nuclear

and Cytoplasmic Extraction reagents (Pierce, Rockford, IL, USA). Protein concentration was measured by the BCA Protein Assay Reagent (Thermo Scientific, Waltham, MA, USA). Cy5-labeled and unlabeled oligonucleotides containing the major or minor allele of rs2240032 or the consensus sequence for SP1, SMAD3/4 (harboring three binding motifs), or OCT1 were purchased from Metabion (Table S3). The EMSA procedure applied here has been described in detail elsewhere (23). For SP1 supershifts, 0.1 μg of SP1 antibody (Santa Cruz, sc-59) was added to the reaction mixture for 45 min at 4°C following the standard binding reaction. For SMAD3 supershifts, nuclear extract and 0.75 μg SMAD2/3 antibody (Santa Cruz, sc-6033) were incubated at room temperature for 30 min before the binding reaction was carried out. To assure specificity of antibody binding, isotype antibodies were used as controls. Experiments were performed at least twice.

Protein purification by DNA affinity

To substantiate the involvement of the candidate transcription factors and to identify further proteins involved in the observed protein–DNA complex, Jurkat nuclear proteins were purified by DNA affinity using oligonucleotides carrying the major C or minor T allele and the eluates were analyzed for DNA binding activity by EMSA. For DNA affinity purification of differentially binding proteins at rs2240032, 200 μl Dynabeads® M-280 Streptavidin (Invitrogen) was prepared according to the manufacturer's instructions and incubated overnight at 4°C with biotin-labeled oligonucleotides containing the same sequence as used for the EMSAs before. After washing with bind and wash buffer (5 mM Tris–HCl, 0.5 mM EDTA, 1 M NaCl), the beads were incubated for 1 h with 2 ng/μl biotin (Roth) and resuspended and incubated in 1 × EMSA-binding buffer without salt (4% v/v glycerol, 0.5 mM EDTA, 0.5 mM DTT, 10 mM Tris–HCl pH 7.5) for 20 min together with 2100 μg Jurkat nuclear extract, 2.5% CHAPS (Roth), and poly dI-dC (Roche Diagnostics, Rotkreuz, Risch, Switzerland). After washing three times with wash buffer (4% v/v glycerol, 0.5 mM EDTA, 0.5 mM DTT, 10 mM NaCl, 10 mM Tris–HCl pH 7.5), elution steps were performed with increasing amounts of NaCl. Electrophoretic mobility shift assays were performed with each eluate fraction to identify the elution step with differential binding pattern as starting material for the mass spectrometry.

Filter-aided sample preparation, nontargeted liquid chromatography–mass spectrometry, and protein identification

The protein eluates that were collected by affinity purification were analyzed by liquid chromatography–mass spectrometry (LC-MS/MS). Samples were prepared according to the filter-aided sample preparation (FASP) approach (24) using Microcon devices YM-30 (Millipore) and foregoing the salt elution step. The LC-MS/MS analysis was performed as described previously on a LTQ-Orbitrap XL (Thermo Scientific) (25) with the following adjustments: A nano trap column was used (300 μm inner diameter ×5 mm, packed with Acclaim PepMap100 C18, 5 μm, 100 Å; LC Packings) before separation by reversed-phase chromatography (PepMap, Sunnyvale, CA, USA, 25 cm, 75 μm ID, 2 μm/100 Å pore size, LC Packings) operated on a RSLC (Ultimate 3000, Dionex, Sunnyvale, CA, USA) with a nonlinear 300-min gradient using 2% acetonitrile in 0.1% formic acid in water (A) and 0.1% formic acid in 75% acetonitrile (B) at a flow rate of 300 nl/min. The gradient settings were: 5–270 min: 5–50% B, 270–275 min: 50–95% B, 280–285 min: 95% B, followed by equilibration for 15 min to starting conditions. From the MS prescan, the 10 most abundant peptide ions were selected for fragmentation and the dynamic exclusion was set to 60 s. For protein identification and label-free relative quantification, the RAW files (Thermo Xcalibur file format) were further analyzed using the Progenesis LC-MS software (version 4.0, Nonlinear, Newcastle upon Tyne, UK), as described previously (26, 27) with the following changes: Spectra were searched against the Ensembl human database (Release 66; 96 556 sequences). A Mascot-integrated decoy database search using the Percolator algorithm calculated an average peptide false discovery rate of <1% when searches were performed with a Percolator score cutoff of 15 and a significance threshold of $P < 0.05$. Peptide assignments were re-imported into Progenesis LC-MS. Normalized abundance of all unique peptides was summed up and allocated to the respective protein. Proteins with <2 peptides used for quantification were excluded. Networks of candidate proteins were analyzed by the STRING database (functional protein association networks).

Construction of plasmids and luciferase assays

Luciferase assays were carried out to characterize the regulatory impact of the RHS7 and regulatory alterations due to the polymorphism rs2240032. Minimal promoter luciferase assays were performed with three different RHS7 constructs containing no other SNPs than rs2240032. For the detection of any polymorphism-dependent regulatory change and to minimize effects arising from genomic overlay, a short RHS7 fragment (RHS7_150 bp) with rs2240032 at mid-position was tested. To identify the effect of the entire human RHS7 harboring rs2240032, the entire human RHS7 fragment (RHS7_1396 bp) was analyzed in the assay. To further characterize the impact of the short RHS7_150 bp fragment located inside the RHS7_1396 bp on the functionality of the entire RHS7 region, these 150 base pairs were deleted (RHS7Δ150 bp_1246 bp). DNA inserts for the luciferase assay were obtained by PCR amplification from human genomic DNA. The RHS7 without the central 150-bp fragment (RHS7_Δ150 bp_1246 bp) was obtained by fusion PCR. Primer sequences are summarized in Table S4. The PCR products were cloned via two restriction sites (Acc65I, Xho1) into the vector pGL4.23[luc2/minP] (Promega, Madison, WI, USA) which contains a minimal promoter. All constructs were verified by sequencing on an ABI3730 (Applied Biosystems, Foster City, CA, USA). The luciferase

assay has been described in detail elsewhere (28) and was carried out with the following adjustments: Jurkat and HeLa cells were seeded at a density of 4×10^5 and 1×10^5 cells/well, respectively. Jurkat cells were transfected with 2000 ng of plasmid per well using Jurkat Trans-IT (Mirus, Madison, WI, USA). HeLa cells were transfected with 1000 ng of plasmid per well using FuGENE6 (Roche Diagnostics). For normalization, 50 ng (for Jurkat) or 25 ng (for HeLa) of the vector pGL4.74 (Promega) was cotransfected. Luciferase activity was measured on the GloMax®-Multi Detection System (Promega). Kolmogorov–Smirnov tests showed no significant deviation from normal distribution of the ratio values. To identify statistically significant differences in promoter activity between the constructs, linear mixed-effects models were used [R package *nlme,* R version 2.15.1 (29)], which account for repeated measurements. *P*-values were corrected for multiple testing using the Bonferroni method.

Results

Polymorphism rs2240032 is located in a 76-kb region of strong LD and in a RHS with high regulatory activity

Analysis of the LD pattern of the gene region showed that all *RAD50* SNPs reported to be associated with atopic traits (Table S1) are highly correlated ($r^2 > 0.97$) as are atopy-associated *IL13* SNPs ($r^2 = 0.998$) located in a second moderate LD block ($0.40 < r^2 < 0.44$) (Fig. 2). This confirms the already known complex pattern of LD in the region. Murine *Rad50*-hypersensitive sites (RHS) were shown to coincide with peaks of evolutionarily conserved DNA sequences (30). Only three of the atopy-associated *RAD50* SNPs are located in conserved human RHS. Bioinformatic analysis indicated that all three RHS SNPs are potentially functional but only rs2240032 maps to a region that harbors all considered regulatory elements (Table S2). Furthermore, rs2240032 is the only one located in RHS7, which had been described to be essential for T-cell differentiation (11). In line with observations in mice, own experiments show that RHS7 is the site with the highest regulatory impact in human Jurkat cells (622% enhancing activity as compared to empty vector, Figure S1). Therefore, this polymorphism was chosen for further analyses.

rs2240032 within RHS7 modulates transcription factor binding *in vitro*

Analysis of protein–DNA interactions by EMSA in Jurkat cells revealed that the major C allele of rs2240032 has a higher affinity for the protein–DNA complex as compared to the minor T allele (Fig. 3, lanes 3 + 10). Competition experiments with increasing concentrations of the unlabeled oligonucleotides containing the major or minor allele led to disappearance of the protein–DNA complex, an effect that was even more pronounced when competing with the major allele (Fig. 3, lanes 4–7, 11–14). The addition of an unlabeled negative control OCT1 oligonucleotide did not result in competition and disappearance of the protein–DNA complex (Fig. 3, lanes 8 + 9, 15 + 16). These results indicate that the observed protein complex has a higher binding affinity to the major allele as compared to the minor allele. Electrophoretic mobility shift assay experiments with HeLa nuclear extracts did not show differences in the binding pattern comparing the major and the minor allele, indicating cell-type-specific allelic protein–DNA interaction at rs2240032 (Fig. S2). To confirm that Jurkat cells are an appropriate surrogate for human T cells, the EMSA experiment was also performed with nuclear extracts from human primary $CD2^+$ T cells. As in the Jurkat cell line, the major C allele revealed a stronger binding affinity for the protein complex than the minor T allele (Fig. S3).

The rs2240032 risk allele alters binding of a protein complex containing SMAD3 and SP1

The analysis for the prediction of differences in transcription factor binding using the MatInspector and SNPInspector tool (Genomatix) revealed that the minor T allele of

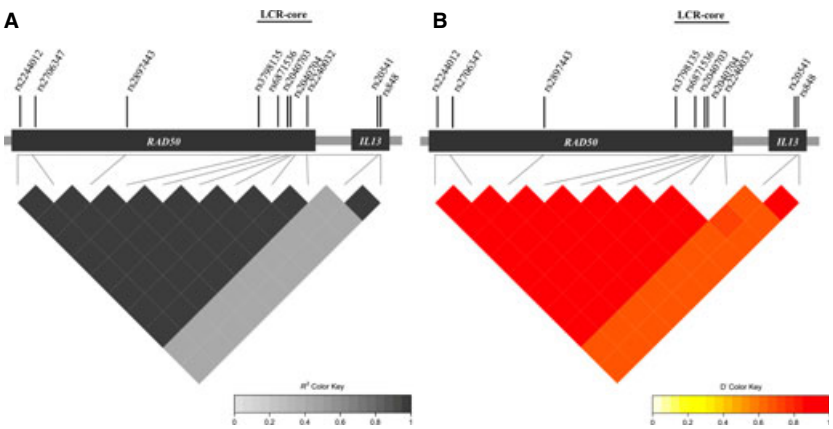

Figure 2 Linkage disequilibrium structure from KORA and PopGen genome-wide association studies data ($n = 2551$). Atopy-associated SNPs within RAD50 ($r^2 > 0.97$) and two atopy-associated reference SNPs in IL13 ($r^2 = 0.998$) are in strong LD. Both LD blocks are in moderate LD with each other ($0.40 < r^2 < 0.44$).

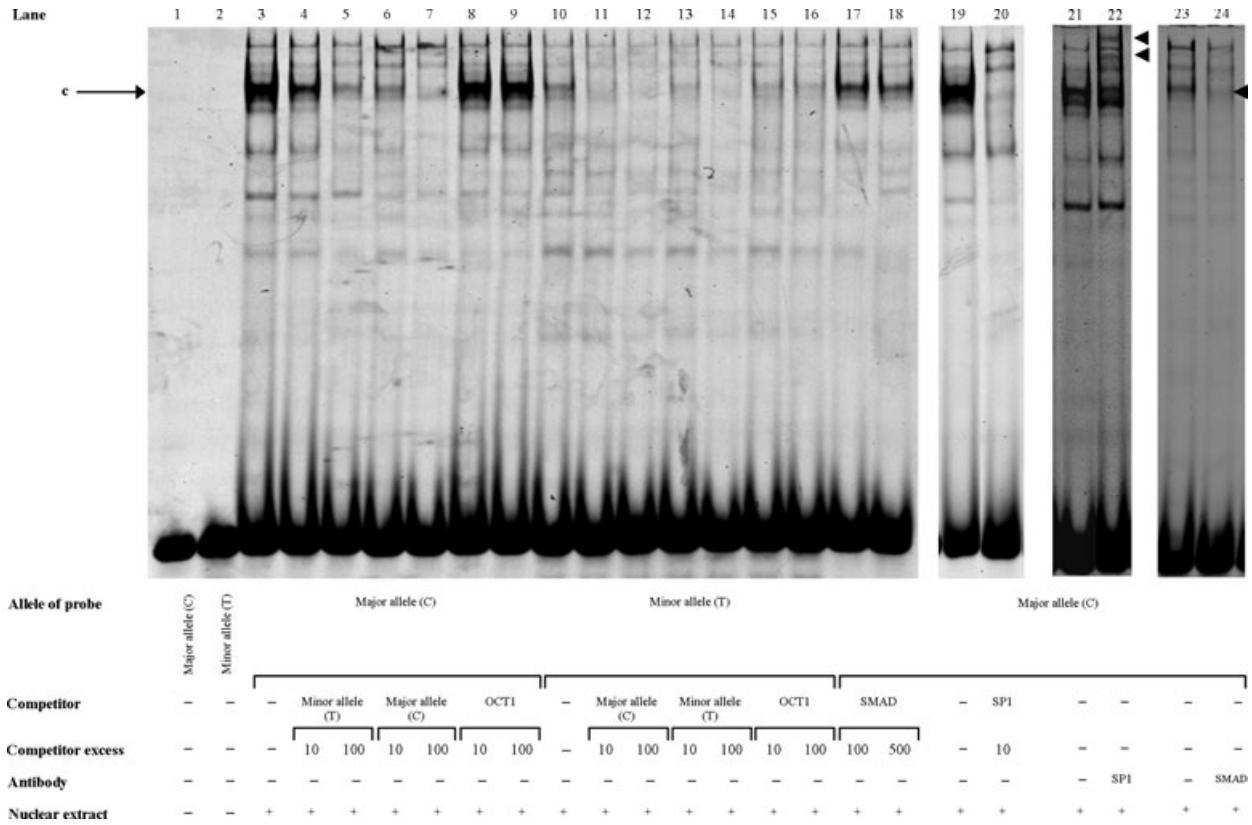

Figure 3 Allele-specific protein–DNA interaction of rs2240032 within the RHS7 and identification of SMAD3 and SP1 as binding proteins in Jurkat nuclear extracts. Labeled oligonucleotides carrying the major allele (C) or minor allele (T) of rs2240032 revealed differential protein binding (lanes 3, 10, 19, 21, 23) (c, indicated by arrow). Competition was performed with unlabeled oligonucleotide DNA (lanes 4–9, 11–18, 20). Supershift with an SP1 antibody resulted in two additional bands (lane 22), and supershift with an SMAD2/3 antibody resulted in band intensity decrease (lane 24) (indicated by arrowheads). Oligonucleotides without nuclear extract did not reveal any protein–DNA binding (lanes 1–2).

rs2240032 results in the loss of a core SMAD3 transcription factor-binding site (TFBS) on the complementary DNA strand (GTCT → ATCT). SMAD3 and SP1 are known to act cooperatively in a complex (31), and indeed, we furthermore found an SP1-binding site (Fig. 4). The SP1-binding site is less conserved and only changed by the SNP in the conserved region surrounding the core-binding motif (gGGGCgggg → gGGGCgagg). To show whether the protein–DNA complex which is destabilized by the minor allele of rs2240032 contains the SMAD3 and SP1 transcription factors, competition and supershift assays were performed. Increasing amounts of an unlabeled SMAD3/4 competitor DNA decreased the intensity of the major allele protein–DNA complex (Fig. 3, lanes 3, 17–18). However, in contrast to competition with the major allele, complete competition of the major allele protein–DNA complex was not achieved with the amounts of SMAD3/4 competitor DNA used. This might be explained by the exclusive presence of a SMAD3/4 consensus sequence within the SMAD3/4 competitor DNA, whereas the major allele competitor also contains a SP1-binding motif and thus inhibits both proteins from complex formation on the labeled oligonucleotide. In contrast, corresponding amounts of OCT1 competitor DNA as negative control did not modify protein–DNA complexes at all (Fig. S4). Further, supporting the specificity of SMAD3 binding, supershift experiments using a SMAD2/3 antibody resulted in a reproducible disappearance of protein–DNA complexes without a visible additional supershift (Fig. 3, lane 24). Competition with an unlabeled SP1 oligonucleotide revealed a strong decrease in protein–DNA complexes even with low amounts of the competitor DNA (10×) (Fig. 3, lanes 19–20). Supershift experiments with an SP1 antibody resulted in two additional upper bands (Fig. 3, lane 22). The formation of two supershifted SP1 complexes has been observed earlier (32) and may be explained by antibody-SP1 binding on the one hand and binding of the antibody-SP1 together with associated complex proteins on the other hand. Affinity-purified eluates from Jurkat cells that showed differences in DNA affinity purification using the major C or the minor T allele were used for protein identification by LC-MS/MS. Electrophoretic mobility shift assay experiments revealed an additional protein–DNA complex in eluate E300 when comparing the bands of each eluate of the DNA affinity purification from reactions with either the

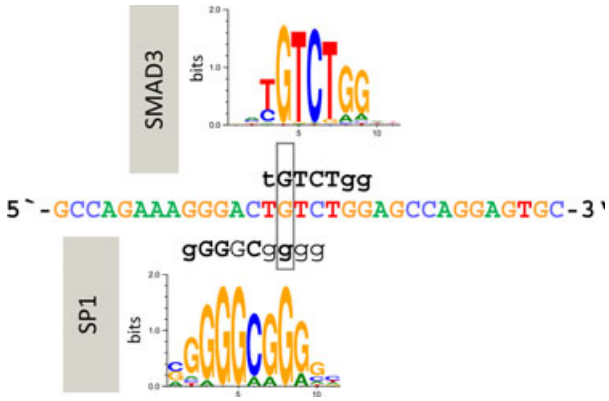

Figure 4 Binding motifs for SMAD3 and SP1 at rs2240032 (C→T) are located inside the human RHS7. The reverse complementary DNA sequence containing the major allele used for EMSA is depicted (chr5: 131, 977, 110–131, 977, 141; GRCh37/hg19). Position weight matrices (PWM) and IUPAC consensus sequences for SMAD3 and SP1, identified with the MatInspector software, are shown. The PWM height of each stack represents the conservation at each position. The variable PWM font size is in proportion to the nucleotide frequency within the consensus sequence. Gray boxed nucleotides = position of rs2240032, black upper-case letters = 'core sequence' (highest conserved, consecutive positions of the matrix), black lower-case letters = 'conserved binding motif' (high matrix conservation), bits = measure of sequence conservation.

major or the minor allele (Fig. S5). The observed differences between the major and the minor allele reflect the initial EMSA experiments (Fig. 3). Mass spectrometric analysis revealed a difference in SMAD3 abundance between the major and the minor allele eluate, confirming the results obtained by competition and supershift experiments. Further identified proteins in the protein–DNA complex were SMAD4, SKI, SKIL, TFAM, ZNF48, NCL, EXOSC6,7,9, and WDR76 (Table 1). SMAD3, SMAD4, SKI, SKIL, and TFAM provided the biggest difference between the major and the minor allele with maximum fold changes ranging from 3.1 to 76.6 (Table 1).

The rs2240032 risk allele influences human RHS7 regulatory function

Having shown that the rs2240032 allele affects the formation of a protein–DNA complex which includes SMAD3, SP1, and additional proteins, we next analyzed whether the variant also modifies the transcriptional activity mediated by the SNP-surrounding RHS7 genomic sequence. In Jurkat cells, the RHS7_150 bp fragment served as a repressor, whereas the RHS7_1396 bp fragment acted as a strong enhancer irrespective of the allele. The deletion of the RHS7_150 bp region from the RHS7_1396 bp fragment resulted in even stronger luciferase promoter activity compared to the RHS7_1396 bp_Ma fragment (fold change: 1.4). Although both alleles for both constructs were either repressing or enhancing, there were significant allele-specific differences for both the RHS7_150 bp and the RHS7_1396 bp fragment: The minor allele exhibited less repressor activity in the RHS7_150 bp fragment (fold change: 2.6) and conclusively higher enhancer activity in the RHS7_1396 bp fragment (fold change: 1.1) compared to the major allele (Fig. 5A). In HeLa cells, both the RHS7_150 bp and the RHS7_1396 bp_Ma fragment showed weak transcriptional activity on a minimal promoter; however, the strong enhancer effect of the RHS7_1396 bp fragment in the Jurkat cell line could not be observed in HeLa cells, which suggests that this effect is T cell specific. After deletion of the 150 base pairs from the entire RHS7, the RHS7Δ150 bp_1246bp fragment lost its

Table 1 Putative complex members binding to the major allele of rs2240032: significant top hits of the mass spectrometry results ranked according to their *P*-values

ENSEMBL Accession	Peptide count	Peptides used for quantification	Confidence score*	ANOVA (*P*)	Max fold change (major to minor)†	Description
ENSP00000259119	4	4	110	<0.001	23, 5	SKIL
ENSP00000341551	3	3	141	<0.001	25, 7	SMAD4
ENSP00000367797	8	8	278	<0.001	4, 2	SKI
ENSP00000332973	2	1	58	0.001	3, 1	SMAD3
ENSP00000315476	2	2	69	0.006	1, 3	EXOSC4
ENSP00000378776	6	1	159	0.007	76, 6	TFAM
ENSP00000324056	1	1	18	0.01	1, 8	ZNF48
ENSP00000318195	18	17	796	0.02	1, 2	NCL
ENSP00000398597	5	5	205	0.02	1, 3	EXOSC6
ENSP00000263795	2	2	84	0.02	1, 6	WDR76
ENSP00000265564	7	7	247	0.02	1, 3	EXOSC7
ENSP00000243498	4	4	181	0.03	1, 3	EXOSC9
ENSP00000261692	2	2	42	0.03	1, 3	CDK2AP1

*The confidence score of a protein is defined as the summed up peptide confidence scores of unique peptides passing the Mascot filter. Confidence score of unique peptides from the Mascot filter describes the likelihood that the detected sequences correspond to the real protein sequence.

†Depending on quantified proteins and can be exaggerated.

regulatory potential on a minimal promoter, indicating that the RHS7_150 bp fragment is necessary for transcriptional activity in HeLa cells. No genotype-specific effect was observed for the RHS7_150 bp fragment in HeLa cells, whereas when the minor T allele was present, the RHS7_1396 bp fragment lost its transcriptional activity (Fig. 5B).

Discussion

The human cytokine gene cluster on 5q31 is one of the best established genetic susceptibility loci for atopic diseases with multiple associated DNA variants (5). In mice, the expression of the Th2 cytokines encoded in this region is orchestrated by multiple *cis*-regulatory elements including a LCR, which is located in the introns of the *Rad50* gene and which is composed of the four DNaseI-hypersensitive sites RHS4-7. During Th2 differentiation, the murine Th2 LCR undergoes epigenetic changes and interacts with the promoters of Th2 cytokine genes while looping out the *Rad50* promoter in order to govern Th2 cytokine expression (33, 34). Previous association and functional studies have shown that the properties of the regulatory elements in the Th2 locus are further modified by genetic variants, in particular polymorphisms at the *IL13* locus (5, 21). Multiple additional association signals in *RAD50* introns have not been characterized on the functional level so far. Re-analysis of GWAS data showed that all *RAD50* SNPs associated with atopic traits are strongly correlated and only in moderate LD with known functional *IL13* variants, thus allowing no prioritization. Likewise, bioinformatic analysis predicted alterations in transcription factor binding, a high regulatory potential, and DNaseI sensitivity in Th2 cells for three of the atopy-associated SNPs (Table S2). Of these, rs2240032 is the only one located in the well-characterized RHS7, which was described to be essential for T-cell differentiation in mice (11) and in line with this

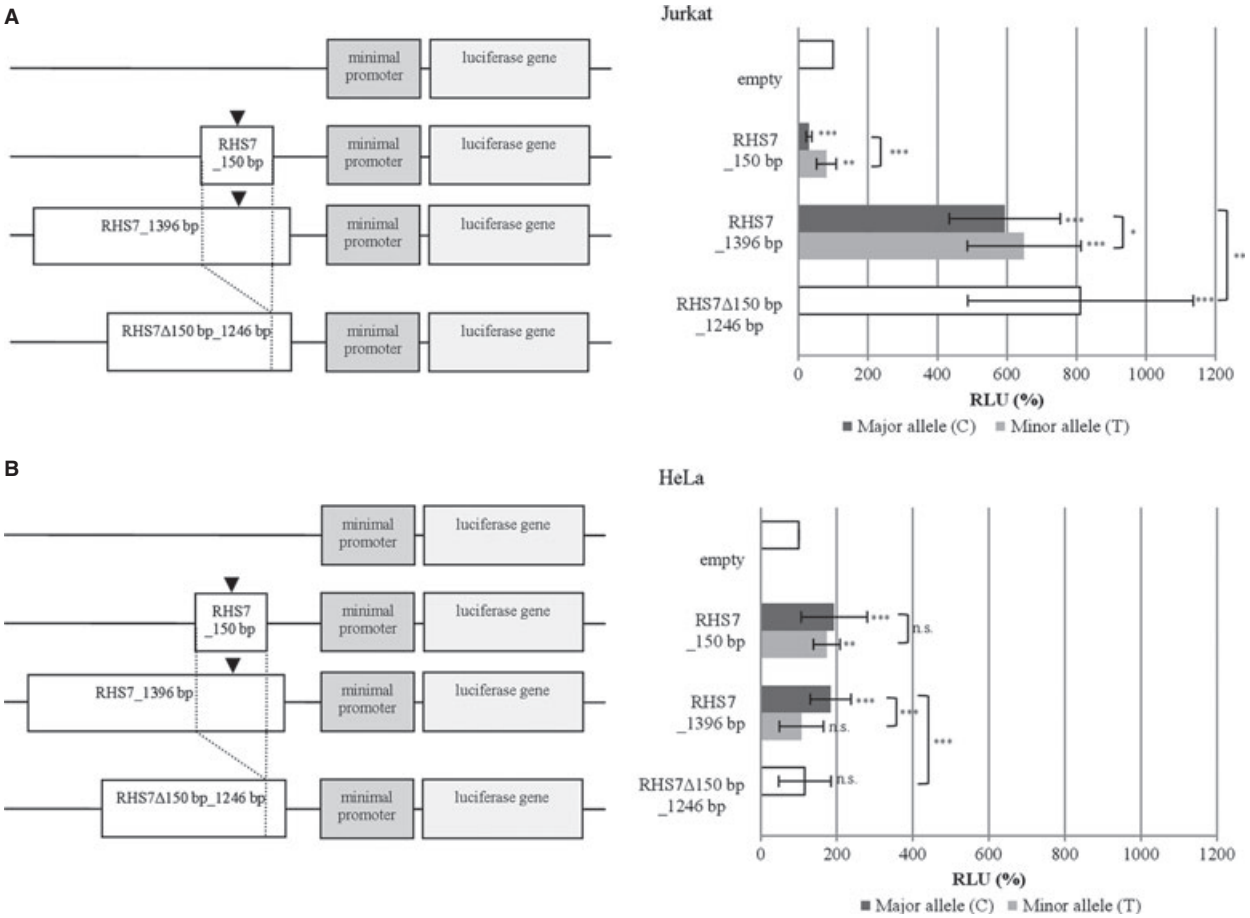

Figure 5 Genotype-specific regulation of minimal promoter activity depending on the rs2240032 variant located inside a regulatory subunit of the human RHS7. Assays were performed in (A) Jurkat T cells and (B) HeLa epithelial cells with minimal promoter vectors without any inserts (empty), vectors with RHS7_150 bp, vectors with RHS7_1396 bp, and vectors with RHS7Δ150 bp_1246 bp inserts. *P*-values are given for each construct in comparison with the empty luciferase vector (unless otherwise noted by brackets). *P*-values were derived from linear mixed-effects models and corrected for multiple testing using the Bonferroni method. Arrowhead = position of rs2240032, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, n.s. = *P*-value not significant; RLU = relative light unit.

showed high regulatory impact in human Jurkat cells in our study. Therefore, this polymorphism was investigated in more detail.

Using an EMSA approach, a T-cell-type-specific, allele-dependent differential binding of nuclear proteins to the polymorphic site rs2240032 was found in the Jurkat T-cell line as well as in human $CD2^+$ T cells, but not in HeLa epithelial cells, indicating that Jurkat cells are an appropriate surrogate cell line for our purpose and that the observed effect is T cell specific. In particular, a strong band in the major allele fraction, which appeared in the upper part of the gel and therefore possibly indicates a large protein complex, revealed reduced binding of the protein complex to the minor T-risk allele. Computational analysis, competition, and supershift experiments showed SMAD3 and SP1 binding to the polymorphic site. However, SMAD3 competition experiments needed a high excess of competitor DNA, and supershift experiments led to a slight decrease in band intensity without a visible supershift, possibly indicating that SMAD3 is tightly bound in a protein–protein complex and not freely available for antibody or competitor DNA binding. Besides SMAD3, the transcription factor SP1 was identified as an additional member of the protein–DNA complex. Surprisingly, even small amounts of SP1 competitor DNA dramatically reduced the observed protein–DNA complex binding. Based on mass spectrometry results and a STRING network analysis (35), it can be hypothesized that the complex that binds to the major C allele sequence with higher affinity compared to the minor T-allele sequence may consist of a SMAD–SKI complex and an EXOSC complex linked by SP1 and SKIV2L2 (Fig. S6). SP1 and SMAD proteins have been described to cooperate at promoter sequences to induce the transcription of several genes in response to TGFβ signaling (31). The specific SNP effect observed here might be rather mediated by SMAD3 because its 'core-binding motif' is directly disrupted by the SNP (tGTCTgg → tATCTgg), whereas the SP1 'core-binding motif' is unaffected by the SNP and only the 'conserved binding sequence' is altered (gGGGCgggg → gGGGCgagg) (Fig. 4). The complete loss of protein–DNA interaction in the EMSA after SP1 competition indicates that SP1 may be essential to stabilize the entire protein complex. SP1 was not detected by mass spectrometry, possibly due to amounts of SP1 being below the detection levels.

SMAD transcription factors are well-described players in transforming growth factor-β (TGFβ) pathways, which regulate growth, differentiation, function of T cells, B cells, macrophages, and natural killer cells (36, 37). Several lines of evidence point toward a role of SMAD transcription factors in atopic diseases. In humans, SMAD3/4 mRNA expression levels are increased in the skin of healthy controls as compared to atopic dermatitis lesional skin (38). SMAD3 polymorphisms have been associated with asthma (7, 39) and atopic dermatitis (40). Further, Smad3-knockout mice express elevated levels of IL4, IL5, and IL13 and are susceptible to asthma (41), whereas the expression of Th1 chemokines is not affected (42). Additionally, these mice show that enhanced allergen-induced Smad3 regulated skin inflammation and IgE production (43). The interaction of SMADs with a variety of SMAD-binding cofactors is essential for high affinity and specificity of target gene regulation (44). It is assumed that at the TFBS investigated here, SMAD3 forms a complex with SMAD4 and interacts with other transcription factors, including SP1. SMAD3/4–SKI complexes are known to be repressors for TGFβ-signaling (45), but might also interfere with other pathways.

The results presented here indicate that the rs2240032 risk allele impedes SMAD3 and SP1 binding, which in turn increases luciferase promoter activity in a cell-type-specific manner. It was also shown that the human RHS7 is a strong Jurkat-specific enhancer and contains different regulatory units (subregulatory regions), which may control the entire RHS7 region. Thus, in case the identified subregulatory region containing the SNP is deleted, the promoter activation is even stronger, indicating that this region has regulatory potential for the entire region in Jurkat cells. Allele-specific differences in reporter gene activation were stronger for the RHS7_150 bp fragment (fold change: 2.6) compared to the RHS7_1396 bp fragment (fold change: 1.1), which is most probably due to the immense enhancing activity and the larger genomic context that leads to masking of the allelic effects in the longer fragment. In order to minimize this genomic overlay and analyze the specific effect of a single gene variant, typically small fragments are used in reporter gene assays to analyze allelic effects (e.g., 46) although this of course does not necessarily represent the situation of the complete genomic context *in vivo*. The observed differences in the EMSA and luciferase experiments for Jurkat and HeLa cells might be explained by the presence of different transcription and cofactors as well as different expression levels and activation states of them within the two cell lines. Data from the ENCODE project for Th1 and Th2 reveal DNA accessibility for the RHS7 in both subtypes at the position of rs2240032. The presence of the RHS7 DNaseI-hypersensitive site in human cells may also support the possibility of a looping out of the LCR containing the RHS7 comparable to a similar mechanism described in mice (12), which might enable promoter activity control of neighboring genes. Our hypothesis is supported by the results of whole-blood expression profile analysis of 740 German individuals from the KORA F4 population-based study analyzed on the Illumina HumanHT-12 v3 BeadChip (47), where increased IL4 expression levels were observed for the rs2240032 risk allele genotype ($P = 0.011$, $\beta = 0.028$) (data not shown). Further, a genotype-dependent effect on IL4 expression was also observed in the whole blood of individuals of the MH96 adult population (Michael Kabesch, personal communication). However, to firmly establish the specific genotype effect on expression, investigations in large collections of more homogenous tissues, in particular T cells, are needed.

In summary, our study shows that RHS7 exerts a cell-type-specific regulatory function in humans, and demonstrates that the common allergy-associated variant rs2240032 has a regulatory function on transcriptional activity and alters binding of SMAD3 and other nuclear transcription factors. Our results provide valuable insights into mechanisms

contributing to the regulation of Th2 cytokine expression by the identification of the causative variant rs2240032, which might be one of several causative variants located in a region with complex LD. It is tempting to speculate that in this region, gene variants together with epigenetic mechanisms fundamentally skew adaptive immune response patterns toward a Th2 phenotype. Hence, further studying the interaction between genetic and epigenetic alterations at this locus is intriguing and important.

Acknowledgments

The authors thank Nadine Lindemann and Viola Maag for technical support and Dr. Christian Lindermayr from the Institute for Biochemical Plant pathology, Helmholtz Zentrum München, for the use of the Typhoon Trio+ Scanner.

Funding

The work was supported by the Fritz Thyssen Stiftung (Az. 10.11.1.225), the German Ministry of Education and Research (BMBF), the National Genome Research Network (NGFN, 01GS 0818, 01GS 0812), a Heisenberg Professorship of the German Research Council Deutsche Forschungsgesellschaft (DFG, WE 2678/7-1 to S.W.), the Deutsche Forschungsgesellschaft (DFG) Cluster of Excellence 'Inflammation at Interfaces', the Graduate School of Information Science in Health (GSISH) of the Technical University Muenchen (to A.K., H.B.), the Else Kroener-Fresenius Foundation Bad Homburg v. d. H (to H.L., M.C., H.L), the Virtual Institute 'Molecular basis of glucose regulation and type 2 diabetes' from the Helmholtz Zentrum Muenchen (to M.C.), the Clinical Cooperation Group 'Nutrigenomics and type 2 diabetes' from the Helmholtz Zentrum Muenchen (to H.L., M.C., H.L) and the Technical University Muenchen (to A.K., H.L., M.C., H.L).

Author contributions

A.K. designed and performed EMSAs, luciferase assays, and protein purification by DNA affinity, carried out the computational analyses, and drafted the manuscript. S.H. and C.T. performed LC-MS/MS and protein identification. K.S., H.P., L.M., A.T., and A.F. supported this work by population-based whole-blood expression profile analysis. S.W. performed the statistical analysis for the luciferase assays, and H.B. performed the LD analysis. H.L., H.L., and M.C. contributed knowledge concerning EMSAs and protein purification by DNA affinity and helped to interpret the results. G.M., M.W., A.P., N.K., T.I., S.E., and K.E. critically interpreted the data and revised the manuscript. J.A., E.R., and S.W. supervised the work and revised the article.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Luciferase assays with several subunits of the core LCR revealed cell-type specific and site specific effects on a minimal promoter.

Figure S2. HeLa nuclear extract does not show allele-specific molecular interaction at rs2240032.

Figure S3. Allele-specific protein-DNA interaction of rs2240032 within the RHS7 in human primary $CD2^+$ T-cell nuclear extracts.

Figure S4. Competition with a negative control does not affect band intensities for rs2240032.

Figure S5. DNA-affinity purification of differential binding proteins for mass spectrometry.

Figure S6. Hypothetical protein network based on mass spectrometry results.

Table S1. Summary of *RAD50* polymorphisms associated with atopic traits used for LD analysis (without imputed SNPs) and all associated SNPs for rs2240032.

Table S2. Bioinformatic analysis of atopy-associated *RAD50-SNPs*.

Table S3. Sequences of oligonucleotide used for EMSA.

Table S4. Primer sequence and size of the fragments used for the PCR fragments to generate RHS7 fragments.

References

1. Lloyd CM, Hessel EM. Functions of T cells in asthma: more than just T(H)2 cells. *Nat Rev Immunol* 2010;**10**:838–848.
2. Bieber T, Cork M, Reitamo S. Atopic dermatitis: a candidate for disease-modifying strategy. *Allergy* 2012;**67**:969–975.
3. Ellinghaus D, Baurecht H, Esparza-Gordillo J, Rodriguez E, Matanovic A, Marenholz I et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat Genet* 2013;**45**: 808–812.
4. Weidinger S, Baurecht H, Naumann A, Novak N. Genome-wide association studies on IgE regulation: are genetics of IgE also genetics of atopic disease? *Curr Opin Allergy Clin Immunol* 2010;**10**:408–417.
5. Vercelli D. Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* 2008;**8**:169–182.
6. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA et al. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 2010;**125**:328–335.
7. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;**363**:1211–1221.
8. Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, Klopp N et al. Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. *PLoS Genet* 2008;**4**:e1000166.

9. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet* 2012;**44**:187–192.
10. Koh BH, Hwang SS, Kim JY, Lee W, Kang MJ, Lee CG et al. Th2 LCR is essential for regulation of Th2 cytokine genes and for pathogenesis of allergic asthma. *Proc Natl Acad Sci USA* 2010;**107**:10614–10619.
11. Lee GR, Spilianakis CG, Flavell RA. Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. *Nat Immunol* 2005;**6**:42–48.
12. Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 2004;**5**:1017–1027.
13. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005;**67**(Suppl 1):S26–S30.
14. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006;**9**:55–61.
15. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–913.
16. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.
17. GenABEL Project Developers. GenABEL: Genome-Wide SNP Association Analysis. 2013; Available from: http://CRAN.R-project.org/package=GenABEL.
18. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 2005;**21**:2933–2942.
19. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
20. Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B et al. 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res* 2007;**17**:1797–1808.
21. Kiesler P, Shakya A, Tantin D, Vercelli D. An allergy-associated polymorphism in a novel regulatory element enhances IL13 expression. *Hum Mol Genet* 2009;**18**:4513–4520.
22. Kozuka T, Sugita M, Shetzline S, Gewirtz AM, Nakata Y. c-Myb and GATA-3 cooperatively regulate IL-13 expression via conserved GATA-3 response element and recruit mixed lineage leukemia (MLL) for histone modification of the IL-13 locus. *J Immunol* 2011;**187**:5974–5982.
23. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE* 2013;**8**:e63812.
24. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods* 2009;**6**:359–362.
25. Hauck SM, Dietter J, Kramer RL, Hofmaier F, Zipplies JK, Amann B et al. Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol Cell Proteomics* 2010;**9**:2292–2305.
26. Merl J, Ueffing M, Hauck SM, von Toerne C. Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Muller cells. *Proteomics* 2012;**12**:10.
27. von Toerne C, Kahle M, Schafer A, Ispiryan R, Blindert M, Hrabe De Angelis M et al. Apoe, Mbl2, and Psp plasma protein levels correlate with diabetic phenotype in NZO mice–an optimized rapid workflow for SRM-based quantification. *J Proteome Res* 2013;**12**:1331–1343.
28. Lattka E, Eggers S, Moeller G, Heim K, Weber M, Mehta D et al. A common FADS2 promoter polymorphism increases promoter activity and facilitates binding of transcription factor ELK1. *J Lipid Res* 2010;**51**:182–191.
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing, 2013. Available from: http://www.R-project.org/.
30. Lee DU, Rao A. Molecular analysis of a locus control region in the T helper 2 cytokine gene cluster: a target for STAT6 but not GATA3. *Proc Natl Acad Sci USA* 2004;**101**:16010–16015.
31. Poncelet AC, Schnaper HW. Sp1 and Smad proteins cooperate to mediate transforming growth factor-beta 1-induced alpha 2(I) collagen expression in human glomerular mesangial cells. *J Biol Chem* 2001;**276**:6983–6992.
32. Xu J, Rogers MB. Modulation of Bone Morphogenetic Protein (BMP) 2 gene expression by Sp1 transcription factors. *Gene* 2007;**392**:221–229.
33. Wilson CB, Rowell E, Sekimata M. Epigenetic control of T-helper-cell differentiation. *Nat Rev Immunol* 2009;**9**:91–105.
34. Martino DJ, Prescott SL. Silent mysteries: epigenetic paradigms could hold the key to conquering the epidemic of allergy and immune disease. *Allergy* 2010;**65**:7–15.
35. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;**39** (Database issue):D561–D568.
36. Letterio JJ, Roberts AB. Regulation of immune responses by TGF-beta. *Annu Rev Immunol* 1998;**16**:137–161.
37. Yang YC, Zhang N, Van Crombruggen K, Hu GH, Hong SL, Bachert C. Transforming growth factor-beta1 in inflammatory airway disease: a key for understanding inflammation and remodeling. *Allergy* 2012;**67**:1193–1202.
38. Gambichler T, Tomi NS, Skrygan M, Altmeyer P, Kreuter A. Alterations of TGF-beta/Smad mRNA expression in atopic dermatitis following narrow-band ultraviolet B phototherapy: results of a pilot study. *J Dermatol Sci* 2006;**44**:56–58.
39. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* 2011;**7**:e1002170.
40. Otsuka K, Takeshita S, Enomoto H, Takahashi T, Hirota T, Tamari M et al. SMAD3 as an atopic dermatitis susceptibility gene in the Japanese population. *J Dermatol Sci* 2009;**55**:200–202.
41. Anthoni M, Wang G, Leino MS, Lauerma AI, Alenius HT, Wolff HJ. Smad3 -signalling and Th2 cytokines in normal mouse airways and in a mouse model of asthma. *Int J Biol Sci* 2007;**3**:477–485.
42. Anthoni M, Fyhrquist-Vanni N, Wolff H, Alenius H, Lauerma A. Transforming growth factor-beta/Smad3 signalling regulates inflammatory responses in a murine model of contact hypersensitivity. *Br J Dermatol* 2008;**159**:546–554.
43. Anthoni M, Wang G, Deng C, Wolff HJ, Lauerma AI, Alenius HT. Smad3 signal transducer regulates skin inflammation and specific IgE response in murine model of atopic dermatitis. *J Invest Dermatol* 2007;**127**:1923–1929.
44. Schmierer B, Hill CS. TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol* 2007;**8**:970–982.
45. Le Scolan E, Luo K. *Ski, SnoN, and Akt as Negative Regulators of Smad Activity:*

Balancing Cell Death and Cell Survival. New York, NY: Humana Press, 2008.
46. Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T et al. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 2003;**35**:341–348.
47. Mehta D, Heim K, Herder C, Carstensen M, Eckstein G, Schurmann C et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur J Hum Genet* 2013;**21**: 48–54.