Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles

Helena Persson, PhD,[a]‡ Andrew T. Kwon, PhD,[b,c] Jordan A. Ramilowski, PhD,[b,c] Gilad Silberberg, PhD,[d] Cilla Söderhäll, PhD,[a,e] Christina Orsmark-Pietras, PhD,[a]* Björn Nordlund, RN, PhD,[e,f,g] Jon R. Konradsen, MD, PhD,[e,f,g] Michiel J. L. de Hoon, PhD,[b,c] Erik Melén, MD, PhD,[e,h,i] Yoshihide Hayashizaki, MD, PhD,[b,j] Gunilla Hedlin, MD, PhD,[e,f,g] Juha Kere, MD, PhD,[a,e,k,l] and Carsten O. Daub, PhD[a,b,c] *Stockholm, Sweden, Yokohama and Wako, Japan, and Helsinki, Finland*

Background: Children with problematic severe asthma have poor disease control despite high doses of inhaled corticosteroids and additional therapy, leading to personal suffering, early deterioration of lung function, and significant consumption of health care resources. If no exacerbating factors, such as smoking or allergies, are found after extensive investigation, these children are given a diagnosis of therapy-resistant (or therapy-refractory) asthma (SA).

Objective: We sought to deepen our understanding of childhood SA by analyzing gene expression and modeling the underlying regulatory transcription factor networks in peripheral blood leukocytes.

Methods: Gene expression was analyzed by using Cap Analysis of Gene Expression in children with SA (n = 13), children with controlled persistent asthma (n = 15), and age-matched healthy control subjects (n = 9). Cap Analysis of Gene Expression sequencing detects the transcription start sites of known and novel mRNAs and noncoding RNAs.

Results: Sample groups could be separated by hierarchical clustering on 1305 differentially expressed transcription start sites, including 816 known genes and several novel transcripts. Ten of 13 tested novel transcripts were validated by means of RT-PCR and Sanger sequencing. Expression of RAR-related orphan receptor A *(RORA)*, which has been linked to asthma in genome-wide association studies, was significantly upregulated in patients with SA. Gene network modeling revealed decreased glucocorticoid receptor signaling and increased activity of the mitogen-activated protein kinase and Jun kinase cascades in patients with SA.

Conclusion: Circulating leukocytes from children with controlled asthma and those with SA have distinct gene expression profiles, demonstrating the possible development of specific molecular biomarkers and supporting the need for novel therapeutic approaches. (J Allergy Clin Immunol 2015;136:638-48.)

Key words: *Therapy-resistant asthma, childhood asthma, peripheral blood leukocytes, transcriptome, long noncoding RNA*

Asthma is the most common chronic disease in children[1] and can be defined as mild, moderate, or severe depending on the extent of medication needed to control symptoms.[2] Problematic severe asthma is characterized by poor disease control, even with high doses of inhaled corticosteroids and additional therapies. It affects approximately 5% of all asthmatic children[3] and can cause extensive personal suffering, early deterioration of lung function, and significant consumption of health care resources.[4] Problematic severe asthma in children can often be explained by exposure to exacerbating factors in the environment, such as smoking or allergens, but therapy-resistant (or therapy-refractory) asthma (SA) is believed to exist in a subgroup of children in whom no such factors are found, despite extensive investigation.[5]

Global gene expression in human asthma has been studied in isolated leukocyte populations,[6-9] bronchial and epithelial biopsy specimens,[10-14] and nasal lavage samples.[15] These studies have provided important mechanistic insights for mild-to-moderate and atopic asthma, but few studies have addressed severe

From [a]the Department of Biosciences and Nutrition and Center for Innovative Medicine (CIMED), [d]the Unit of Computational Medicine, Department of Medicine, [e]the Centre for Allergy Research, [g]the Department of Women's and Children's Health, and [h]the Institute of Environmental Medicine, Karolinska Institutet, Stockholm; [b]Omics Science Center,§ RIKEN Yokohama Institute, Yokohama; [c]the Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama; and [f]Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm; [i]Sachs' Children's Hospital, Stockholm; [j]the Preventive Medicine and Diagnosis Innovation Program, RIKEN Research Cluster for Innovation, Wako; [k]the Folkhälsan Institute of Genetics, Helsinki; and [l]the Research Programs Unit, University of Helsinki.

*Dr Orsmark-Pietras is currently affiliated with the Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden.

‡Dr Persson is currently affiliated with the Division of Oncology and Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden.

§The RIKEN Omics Science Center ceased to exist as of April 1, 2013, because of RIKEN reorganization.

Supported by a research grant for the RIKEN Omics Science Center from MEXT (to Y.H.) and grants from the Swedish Foundation for Strategic Research (RBc08-0027) and the Swedish Research Council (2009-5091; to J.K.). Patient sample collection was supported by grants from the Freemason Child House Foundation in Stockholm, the Konsul Th. C. Bergh's Foundation, the Swedish Asthma and Allergy Association's Research Foundation, the Centre for Allergy Research at Karolinska Institutet, the Swedish Heart-Lung Foundation, Karolinska Institutet, and the Bernard Osher Initiative for Research on Severe Asthma.

Disclosure of potential conflict of interest: J. R. Konradsen has received research support from Novartis; has received lecture fees from Novartis, Thermo Fisher Scientific, and Meda; and has received travel support from Thermo Fisher Scientific. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication December 19, 2013; revised November 23, 2014; accepted for publication February 3, 2015.

Available online April 9, 2015.

Corresponding author: Juha Kere, MD, PhD, Karolinska Institutet, Department of Biosciences and Nutrition and Center for Innovative Medicine (CIMED), Hälsovägen 7-9, SE-141 83 Huddinge, Sweden. E-mail: juha.kere@ki.se.

0091-6749

© 2015 The Authors. Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

http://dx.doi.org/10.1016/j.jaci.2015.02.026

Abbreviations used

CA: Controlled persistent asthma
CAGE: Cap Analysis of Gene Expression
CTRL: Healthy control subject
GLM: Generalized linear model
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes
LD: Linkage disequilibrium
lincRNA: Long intergenic noncoding RNA
MAPK: Mitogen-activated protein kinase
MARA: Motif Activity Response Analysis
NK: Natural killer
RORA: RAR-related orphan receptor A
SA: Therapy-resistant asthma
SNP: Single nucleotide polymorphism
TC: Tag cluster
TF: Transcription factor
TSS: Transcription start site

asthma[16] or SA,[17] for which few clinically well-characterized cohorts are available.[18]

Our aim in the present study was to identify the gene pathways that are deregulated in children with SA. Therefore we analyzed gene expression in peripheral blood leukocytes from children with SA (n = 13) and controlled persistent asthma (CA; n = 15), as well as age-matched healthy control subjects (CTRLs; n = 9). We used Cap Analysis of Gene Expression (CAGE), a technology based on next-generation sequencing that quantitatively measures gene expression for individual transcription start sites (TSSs), including mRNAs and long noncoding RNAs.[19] The asthmatic children were selected from Swedish Search, a national multicenter, cross-sectional study that aims to identify children with problematic severe asthma and characterize those who are difficult to treat or resistant to therapy.[20] Although these samples had previously been analyzed with expression microarrays,[21] we reasoned that the promoter-level resolution of CAGE sequencing data and the unbiased detection of novel genes and alternative promoters could provide new and interesting information.

METHODS

Children with CA (n = 15) and SA (n = 13) were selected from the Swedish Search study; details on inclusion criteria and clinical examination procedures are described elsewhere.[20,21] CTRLs (n = 9) were recruited at Astrid Lindgren Children's Hospital, Stockholm, Sweden, from children admitted for elective surgical procedures unrelated to asthma. Buffy coat isolation from blood samples and RNA extraction were described previously.[21] For statistical analysis of clinical parameters, normally distributed data are summarized by means and SDs and compared by using the Student *t* test. Nonnormally distributed data are presented as medians and interquartile ranges and compared with the Mann-Whitney *U* test. A comprehensive description of the methods used in this study is provided in the Methods section in this article's Online Repository at www.jacionline.org. Primers for validation of novel TSSs are listed in Table E1 in this article's Online Repository at www.jacionline.org. Informed consent was obtained from all participating children and their parents, and the study was approved by the local ethics committee.

RESULTS

We analyzed the transcriptome of peripheral blood leukocytes from children with CA (n = 15) and those with SA (n = 13), as well as age-matched CTRLs (n = 9) using CAGE sequencing. The asthmatic children were selected from the Swedish Search study,[20] and all cases of SA had been classified as SA in the absence of any identified aggravating factors. Most clinical variables did not differ between the CA and SA groups, but children with SA had significantly stronger responses to methacholine provocation and increased blood neutrophil counts compared with those with CA (Table I), as previously reported in children with severe asthma.[22] Leukocyte cell counts are provided for individual samples in Table E2 in this article's Online Repository at www.jacionline.org. A flow chart illustrating the different analyses performed in this article is provided in Fig E1 in this article's Online Repository at www.jacionline.org.

CAGE measures gene expression at the TSS of known and novel transcripts

CAGE sequencing detects the 5′ end of transcripts with a 5′ 7-methylguanosine cap, such as mRNAs and many long noncoding RNAs. Use of the HeliScope single-molecule sequencer (Helicos BioSciences, Cambridge, Mass) bypasses the need for PCR amplification of sample libraries to produce highly quantitative expression data.[19] Expression for a particular TSS is estimated by grouping nearby sequencing reads (tags) after mapping to the human genome to tag clusters (TCs) and counting the number of reads within each cluster. Sequencing library sizes are shown in Table E2. We limited expression analysis to a set of 45635 TCs that were detected by at least 3 reads in at least 9 samples (ie, the size of the smallest sample group). When compared with GENCODE[23] transcript annotation, 49% of the TCs mapped within 500 bp of a known TSS, and another 24% mapped in exons of known genes. A further 27% of the TCs were not located within exons of known coding or noncoding genes; these might represent novel genes or transcript isoforms. TC annotation is summarized in Fig E2 in this article's Online Repository at www.jacionline.org.

Children with CA and those with SA have distinct gene expression profiles and cluster separately

We performed differential expression analysis for all pairwise comparisons: children with CA versus CTRLs, children with SA versus CTRLs, and children with SA versus those with CA. A total of 1305 TCs were differentially expressed with a Benjamini-Hochberg–adjusted *P* value of less than .1 in any comparison, and 1029 of these mapped to 816 unique gene symbols (ie, some genes had >1 cluster). The remaining TCs were intergenic, intronic, or antisense to known transcripts. Hierarchical clustering of the samples based on differentially expressed TCs clearly separated the sample groups, with the exception of a single CTRL sample that clustered with the SA group (Fig 1). Statistically significant TCs with GENCODE transcript annotation are provided in Tables E3-E5 in this article's Online Repository at www.jacionline.org.

These results imply that differences in the underlying disease biology of CA and SA are reflected as distinct molecular fingerprints. However, because we measured gene expression for a complex population of leukocytes, the profiles also reflect differences in cell composition between sample groups, such as the significantly increased neutrophil counts in children with

TABLE I. Clinical characteristics of the included samples

	CTRLs	Children with CA	Children with SA	*P* value
No. of subjects	9	15	13	
Age (y), mean (SD)	11 (5)	13 (3)	14 (3)	NS
Female sex (%)	11	40	38	NS
BMI (%), median (IQR)	59 (36-66)	62 (32-76)	77 (61-93)	NS
ACT score, mean (SD)	NA	24 (1)	16 (3)	
ICS (μg), median (IQR)	NA	320 (180-400)	800 (800-800)	
FEV_1 (% predicted), mean (SD)	NA	86 (11)	88 (22)	NS
$DRS_{methacholine}$, median (IQR)	NA	0.8 (0.2-8.0)	11.1 (1.7-61.0)	.015 (CA-SA)†
FENO (ppb), median (IQR)	NA	13 (9-21)	18 (11-38)	NS
Total IgE (kU_A/L), median (IQR)	NA	200 (66-795)	500 (120-1000)	NS
Atopic* (%)	NA	80	85	NS
Total WBC ($10^9 \times$ L-1), mean (SD)	5.8 (0.8)	5.9 (1.5)	6.7 (2.2)	NS
Eosinophils ($10^9 \times$ L-1), mean (SD)	0.2 (0.1-0.2)	0.2 (0.2-0.4)	0.4 (0.2-0.4)	NS
Lymphocytes ($10^9 \times$ L-1), mean (SD)	2.4 (0.6)	2.7 (0.6)	2.3 (0.7)	NS
Monocytes ($10^9 \times$ L-1), mean (SD)	0.4 (0.1)	0.5 (0.2)	0.4 (0.1)	NS
Neutrophils ($10^9 \times$ L-1), mean (SD)	2.6 (0.5)	2.5 (0.7)	3.5 (1.6)	.046 (CA-SA)‡

ACT, Asthma Control Test; *BMI*, body mass index; *$DRS_{methacholine}$*, slope of the dose-response curve for methacholine provocation; *FENO*, fraction of exhaled nitric oxide; *ICS*, inhaled corticosteroids; *IQR*, Interquartile range; *NA*, not available; *NS*, not significant.
*Atopy is defined as a specific IgE level of greater than 0.35 kU/L to inhalant allergens (Phadiatop).
†Mann-Whitney *U* test.
‡Student *t* test.

SA (Table I). Therefore we analyzed differential expression further using the generalized linear model (GLM) methods in edgeR,[24] applying a model that incorporated both sample groups and neutrophil, eosinophil, and lymphocyte blood counts as fractions of the total number of white blood cells. This adjusted the expression levels for differences in cell counts among samples but also removed TCs that varied with both sample groups and neutrophil counts, resulting in a conservative set of TCs that were differentially expressed, irrespective of cell composition. A total of 680 TCs were differentially expressed, with an adjusted *P* value of less than .1 in any comparison, and 503 of these mapped to 460 unique gene symbols. Hierarchical clustering and Venn diagrams illustrating the overlap between the 2 differential expression analyses are shown in Figs E3 and E4 in this article's Online Repository at www.jacionline.org, respectively. Statistically significant TCs with GENCODE transcript annotation are provided in Tables E6-E8 in this article's Online Repository at www.jacionline.org.

Because our patient groups were comparatively small, we assessed statistical power for our current sample size and false discovery rates and obtained consistent power predictions of between 0.4 and 0.5. By randomizing sample group labels, we found the rate of false-positive results among differentially expressed TCs to be practically zero. Combining these analyses (see Figs E5 and E6 in this article's Online Repository at www.jacionline.org), we conclude that although identification of marker genes between disease groups is somewhat limited by the statistical power at these sample sizes, we can still identify a stable set of biomarker genes.

Genes differentially expressed between children with SA and those with CA also distinguish severe from nonsevere asthma in adults

We tested these results on an independent set of publicly available gene expression microarray data for $CD4^+$ and $CD8^+$ T cells from adults with severe and nonsevere asthma.[16] The original study found activation of $CD8^+$ T cells in patients with severe asthma but only minor differences for $CD4^+$ cells and no significant difference between patients with nonsevere asthma and CTRLs. We selected all available microarray probes for GENCODE transcripts with a differentially expressed TC in our analysis (1537 probes for 698 of 1305 TCs from pairwise comparisons between sample groups) and clustered the expression profiles of the 2 T-cell types separately. Although our set of differentially expressed transcripts was derived from mixed leukocytes from children, it could still separate $CD8^+$ T cells from adults with severe and those with nonsevere asthma, and results were similar for differentially expressed transcripts from the GLM analysis (see Figs E7 and E8 in this article's Online Repository at www.jacionline.org). Bootstrapping analysis of the stability of the separation between groups is shown in Fig E9 in this article's Online Repository at www.jacionline.org. These results suggest that the differences in gene expression reflect the underlying disease dysregulation and that there are molecular similarities between SA in children and severe asthma in adults. It also hints that some of the differential expression we see is likely to depend on $CD8^+$ rather than $CD4^+$ T cells. Finally, it shows that results from CAGE sequencing can be transferred to another technology and that these transcripts are applicable as markers in another sample set.

Increased activity of natural killer cells and genes involved in leukocyte migration in children with SA

To explore the biology underlying the differentially expressed TCs, we annotated them with gene symbols and searched for Gene Ontology (GO) term and pathway (Kyoto Encyclopedia of Genes and Genomes [KEGG]) enrichment using GOseq.[25] Table II lists a selection of significant terms (adjusted $P <$ .05) and their associated genes. Because some of these might result from cell count differences, we also analyzed enrichment for the GLM analysis of differential expression. Few GO terms and KEGG pathways were significantly enriched in this analysis, partly because of the smaller number of differentially expressed genes. Most likely, this is also because many

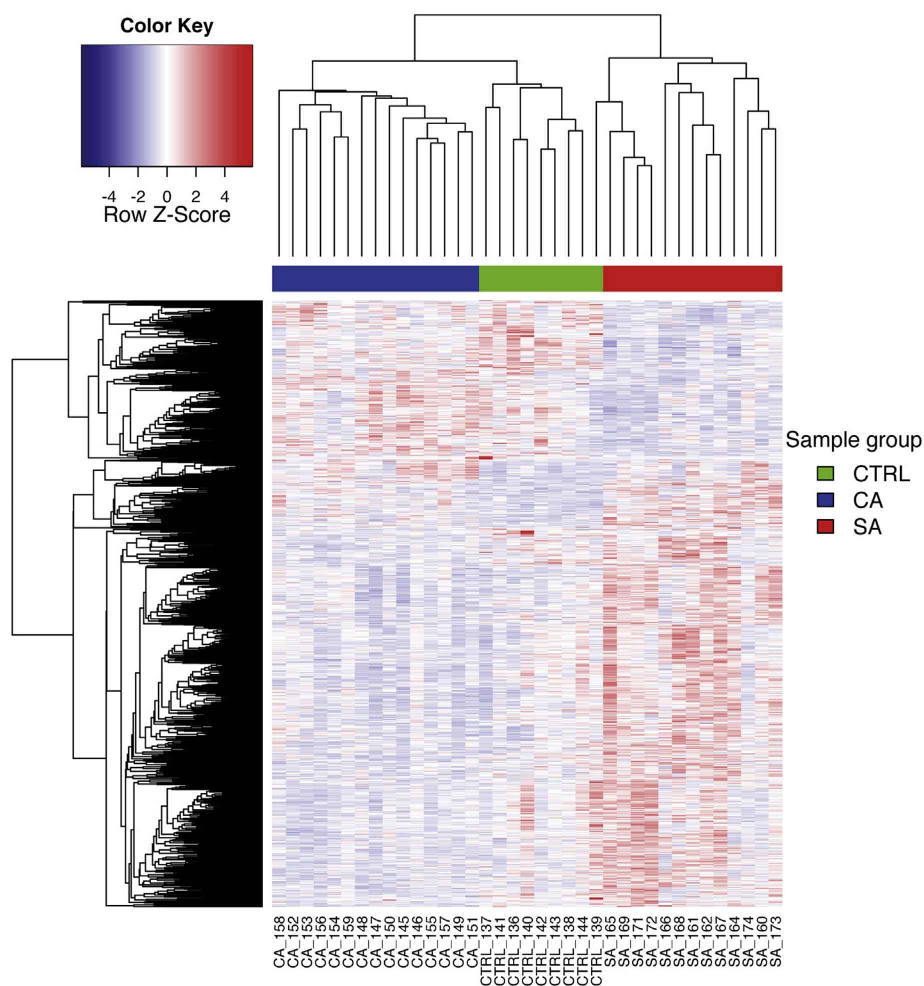

FIG 1. Hierarchical clustering based on 1305 TCs differentially expressed in any of the 3 pairwise comparisons of CTRLs versus children with CA, CTRLs versus children with SA, or children with CA versus those with SA. The distance measure for clustering of samples and TCs (expression in normalized tags per million reads *[tpm]*) was Spearman rank correlation.

differences found before adjusting for leukocyte cell counts represent differences in sample composition. Categories that retain significance (adjusted $P < .1$) are marked in Table II. Both children with CA and those with SA displayed activation of genes involved in natural killer (NK) cell–mediated cytotoxicity, but the differentially expressed genes differed. Fig E10 in this article's Online Repository at www.jacionline.org shows differences in NK cell signaling between the 2 patient groups in a KEGG pathway diagram.

Experimental validation of novel transcripts and alternative TSSs

Using CAGE sequencing, we could also detect novel transcripts and alternative TSSs. Combined bioinformatic analysis and manual curation identified a high-confidence set of 13 candidate novel TSSs for experimental validation (Table III). Ten of these could be detected by using RT-PCR for 2 CTRL samples, and all 10 were further confirmed by means of Sanger sequencing (see Fig E11 in this article's Online Repository at www.jacionline.org). Fig 2 shows examples of validated novel TSSs with mean expression across samples and supporting sequence data, including an alternative first exon for Rho GTPase activating protein 15 *(ARHGAP15)*, which has been called a "master negative regulator of neutrophil functions."[26] Both known and novel TSSs were upregulated in children with SA versus CTRLs, irrespective of adjustment for sample leukocyte composition, and also between children with SA and those with CA in the classical analysis of differential expression (see Fig E12 in this article's Online Repository at www.jacionline.org). The novel first exon for *ARHGAP15* is spliced to the second exon of the annotated transcript upstream of the open reading frame. Expression in sorted leukocyte populations from the FANTOM5 project[27] suggests that most of the novel TSSs are expressed in neutrophils and eosinophils (see Fig E13 in this article's Online Repository at www.jacionline.org).

Comparison of differentially expressed TCs with genetic associations in asthmatic patients

We compared the genomic coordinates of our differentially expressed TCs with single nucleotide polymorphisms (SNPs)

TABLE II. Representative GO terms and KEGG pathways enriched among differentially expressed genes

GO/KEGG ID	GO/KEGG name	Adjusted *P* value	Genes
Upregulated in children with SA vs those with CA			
KEGG:04666	FcγR-mediated phagocytosis	8.992e-06	*ARPC1B, ARPC2, ASAP1, DOCK2, FCGR2A, GAB2, LIMK2, LYN, MAPK1, MARCKS, PAK1, PIK3R5, PLCG2, PRKCB, PRKCD, VASP, WAS*
KEGG:04810*	Regulation of actin cytoskeleton	2.109e-04	*ARPC1B, ARPC2, CYFIP2, FGD3, IQGAP1, IQGAP2, ITGAL, LIMK2, MAPK1, MSN, MYL12A, PAK1, PAK2, PIK3R5, PIKFYVE, PXN, RHOA, SOS2, SSH2, TIAM2, WAS*
KEGG:04380	Osteoclast differentiation	.002	*FCGR2A, FOS, FOSL2, GAB2, JUNB, LILRA2, MAP2K6, MAPK1, NCF2, NCF4, NFATC1, PIK3R5, PLCG2, SPI1, TGFBR2*
KEGG:4670*	Leukocyte transendothelial migration	.003	*GNAI2, ITGAL, MSN, MYL12A, NCF2, NCF4, PIK3R5, PLCG2, PRKCB, PXN, RHOA, VASP*
KEGG:4062	Chemokine signaling pathway	.007	*ARRB2, DOCK2, GNAI2, GNG2, LYN, MAPK1, PAK1, PIK3R5, PREX1, PRKCB, PRKCD, PXN, RHOA, SOS2, TIAM2, WAS*
KEGG:5140	Leishmaniasis	.011	*CR1, FCGR2A, FOS, MAPK1, NCF2, NCF4, PRKCB, PTPN6, TLR4*
KEGG:4662	B-cell receptor signaling pathway	.021	*DAPP1, FOS, LYN, MAPK1, NFATC1, PIK3R5, PLCG2, PRKCB, PTPN6, SOS2*
KEGG:5100	Bacterial invasion of epithelial cells	.021	*ARPC1B, ARPC2, GAB1, HCLS1, PIK3R5, PXN, RHOA, RHOG, WAS*
KEGG:5150	*Staphylococcus aureus* infection	.021	*C5AR1, FCGR2A, FPR1, ITGAL, PTAFR, SELPLG*
KEGG:4650*	NK cell–mediated cytotoxicity	.025	*CD48, ITGAL, MAPK1, NFATC1, PAK1, PIK3R5, PLCG2, PRKCB, PTPN6, SOS2, TNFRSF10C*
Downregulated in children with SA vs those with CA			
GO:0030852	Regulation of granulocyte differentiation	.017	*GFI1B, IKZF1, OGT, RARA*
GO:0046649	Lymphocyte activation	.044	*ATM, BCL11B, CD3G, CSK, FKBP1A, GRAP2, IKZF1, IKZF3, ITPKB, LAT, MS4A1, PIK3R1, RARA, TRAC*
Upregulated in children with SA vs CTRLs			
KEGG:4730	Long-term depression	.030	*GNAQ, IGF1R, ITPR2, PLCB1, PRKCB*
Downregulated in children with SA vs CTRLs			
KEGG:3010*	Ribosome	1.399e-04	*RPL11, RPL13, RPL27, RPL27A, RPL3, RPL35, RPL35A, RPL37A, RPLP2, RPS11, RPS21, RPS27, RPSA*
GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	.028	*BCL11B, DDX17, EGR1, EGR2, FOXO3, IKZF1, MED10, NFIX, SOX4, UTF1*
Upregulated in children with CA vs CTRLs			
KEGG:4650*	NK cell–mediated cytotoxicity	6.332e-06	*CD247, GZMB, NCR3, PRF1, SH2D1B, ZAP70*
Downregulated in children with CA vs CTRLs			
KEGG:4380	Osteoclast differentiation	.010	*FOS, FOSB, JUNB, NFKBIA, SPI1*
GO:0006357	Regulation of transcription from RNA polymerase II promoter	1.168e-05	*CSDA, EGR1, EGR2, ETS2, FOS, FOSB, ID1, JUNB, KLF13, NFIL3, NFKBIA, PLK3, RXRA, SPI1, TXNIP, ZFP36*
GO:0006955	Immune response	.007	*AMICA1, AQP9, C5AR1, CD83, DUSP6, EGR1, FOS, NFIL3, NFKBIA, PTEN, TXNIP*
GO:0008219	Cell death	.021	*CSDA, DUSP6, EGR3, EIF4G2, ID1, IRS2, NFKBIA, PLK3, PREX1, PTEN, RXRA, SRGN, TXNIP*

*Adjusted $P < .1$ after adjustment for cell counts in the GLM analysis.

significantly associated with asthma in genome-wide association studies. HaploReg v2[28] was used to identify linkage disequilibrium (LD) blocks for 62 SNPs with genome-wide significant association with asthma from 17 studies.[29] Two differentially expressed TCs from the classical differential expression analysis overlapped LD blocks associated with asthma in European populations that were located in the IL-6 receptor *(IL6R)*[30] and IL-2 receptor β *(IL2RB)*[31] genes (see Tables E9 and E10 in this article's Online Repository at www.jacionline.org). Both TCs were upregulated 3-fold in children with SA compared with those with CA but were not differentially expressed between children with CA and CTRLs. A third TC in absent in melanoma 2 *(AIM2)* overlapped an LD block associated with asthma in subjects of African descent.[32] Furthermore, 2 differentially expressed TCs were in the genes GRB2-associated binding protein 1 *(GAB1)*[33] and RAR-related orphan receptor A *(RORA)*,[31] which have been previously associated with asthma, although not within the association blocks. Only the TC in *RORA* was significantly differentially expressed in the cell count–adjusted (GLM) analysis, which also included 2 differentially expressed TCs in *PYHIN1* and *IFI16*, both of which were located in the same LD block as *AIM2*.[32] Expression levels for these clusters are shown in Fig E14 in this article's Online Repository at www.jacionline.org. Additionally, an SNP 225 bp upstream of the cluster in *IL2RB*, rs228954, overlaps a DNase I hypersensitivity signal[34] and is predicted to change a binding motif for the peroxisome proliferator-activated receptors.[35,36] Interestingly, rs2284033, an SNP located 1.6 kb upstream of the cluster, is predicted to alter a binding site for *RORA*. The small overlap between differentially expressed TCs and associated genetic variants supports the concept that transcriptional

TABLE III. Differentially expressed novel TSSs (adjusted $P < .1$) selected for experimental validation by using RT-PCR and Sanger sequencing

TSS	Gene	Mean tpm	Log fold change			Validated
			CTRLs vs children with CA	CTRLs vs children with SA	Children with CA vs those with SA	
Alternative TSSs						
chr2:24714142(+)	*NCOA1*	57	NS	NS	0.39	No
chr2:143828538(+)*	*ARHGAP15*	175	NS	0.97	0.99	Yes
chr4:148973914(+)	*ARHGAP10*	3	NS	NS	−1.02	Yes
chr8:131370837(−)	*ASAP1*	105	NS	NS	0.57	No
chr9:117160251(−)	*AKNA/RP11-9M16.2*	5	NS	NS	0.63	Yes
chr13:46948966(−)	*KIAA0226L*	25	NS	0.94	NS	Yes
chr15:86087320(+)*	*AKAP13*	30	NS	0.78	0.63	Yes
chr16:90023689(+)	*DEF8*	35	NS	NS	0.43	Yes
chr17:8316490(+)*	*NDEL1*	137	NS	0.61	0.68	Yes
Novel transcripts						
chr6:105901520(+)	Possibly lincRNA	4	NS	NS	−1.01	No
chr12:8605756(−)*	Possibly lincRNA	5	NS	NS	0.93	Yes
Antisense transcripts						
chr3:134314106(+)	*KY*	33	NS	NS	0.58	Yes
chr14:95652012(+)	*CLMN/CTD-2240H23.2*	19	NS	NS	0.52	Yes

Coordinates represent the first nucleotide of the transcript, with the genomic strand in parentheses.
NS, Not significant; *tpm*, tags per million reads.
*Significantly differentially expressed also after adjustment for cell counts.

dysregulation typically occurs downstream of the causal factors at disease onset.[37]

Distinct transcription factor networks regulate gene expression in children with CA and those with SA

Because CAGE measures gene expression by sequencing TSSs, the data are well suited for promoter analysis. We applied Motif Activity Response Analysis (MARA)[38] to model the regulatory networks underlying gene expression profiles. MARA combines prediction of transcription factor (TF) binding sites with mathematic modeling of changes in gene expression to calculate the global importance of TF binding sites for driving gene expression in a given sample. We analyzed proximal promoter regions from 300 bp upstream to 100 bp downstream of TCs and a wider promoter region from 1 kb upstream to 200 bp downstream. We limited the analysis to TCs near known TSSs plus our high-confidence set of novel TSSs and again used linear models to adjust for sample leukocyte composition. This produces a conservative estimate of differences between sample groups, excluding also motifs varying with both group and cell counts. This adjustment is especially important because we aim to infer regulatory relationships between TFs and target genes occurring within a cell. The top 10 significant motifs (adjusted $P < .1$) from pairwise comparisons of motif activities between sample groups are shown in Table IV for the wider promoter region (see Table E11 in this article's Online Repository at www.jacionline.org for the proximal promoter region). Box plots illustrating motif activities within sample groups are shown in Figs E15-E17 in this article's Online Repository at www.jacionline.org. Expression profiles of the TFs corresponding to the motifs in Table IV are shown in Fig E18 in this article's Online Repository at www.jacionline.org for sorted leukocyte populations from the FANTOM5 project.[27] Fig 3 shows the regulatory network for motifs with significantly different activity between children with CA and those with SA. Differences in gene expression between these 2 types of asthma appear to be determined by 2 internally highly connected TF modules. One of these has higher activity in children with SA than in those with CA and includes motifs for TFs, such as GTF2I, HIC1, PAX5, and the MYF and TFAP2 families. The other module, which includes motifs for TFs, such as ZNF148 and the NFY family, has decreased activity in children with SA compared with those with CA. Notably, motif activity for the glucocorticoid receptor (NR3C1) was also lower in children with SA. Regulatory networks for the proximal promoter region and the other (children with CA vs CTRLs and children with SA vs CTRLs) are shown in Figs E19-E23 in this article's Online Repository at www.jacionline.org.

GO and pathway analyses highlight the roles of individual TFs

To understand the function of these TFs in our samples, we analyzed GO term and KEGG pathway enrichment for their predicted target gene sets. For example, the motif for ZNF238 was associated with genes involved in mitogen-activated protein kinase (MAPK) signaling, including nuclear factor of kappa light polypeptide gene enhancer in B cells 1 *(NFKB1)*. This motif has increased activity in children with CA compared with CTRLs and even higher activity in children with SA. The proximal promoter regions of genes involved in the MAPK and Jun kinase cascades were also enriched for the TFAP2B motif, which has increased activity in children with SA versus those with CA. Interestingly, the ELK1,4_GABP{A,B1} motif, which has decreased motif activity in both children with CA and those with SA versus CTRLs, was found in the promoters of a large number of ribosomal proteins, as well as proteins involved in RNA transport and the spliceosome.

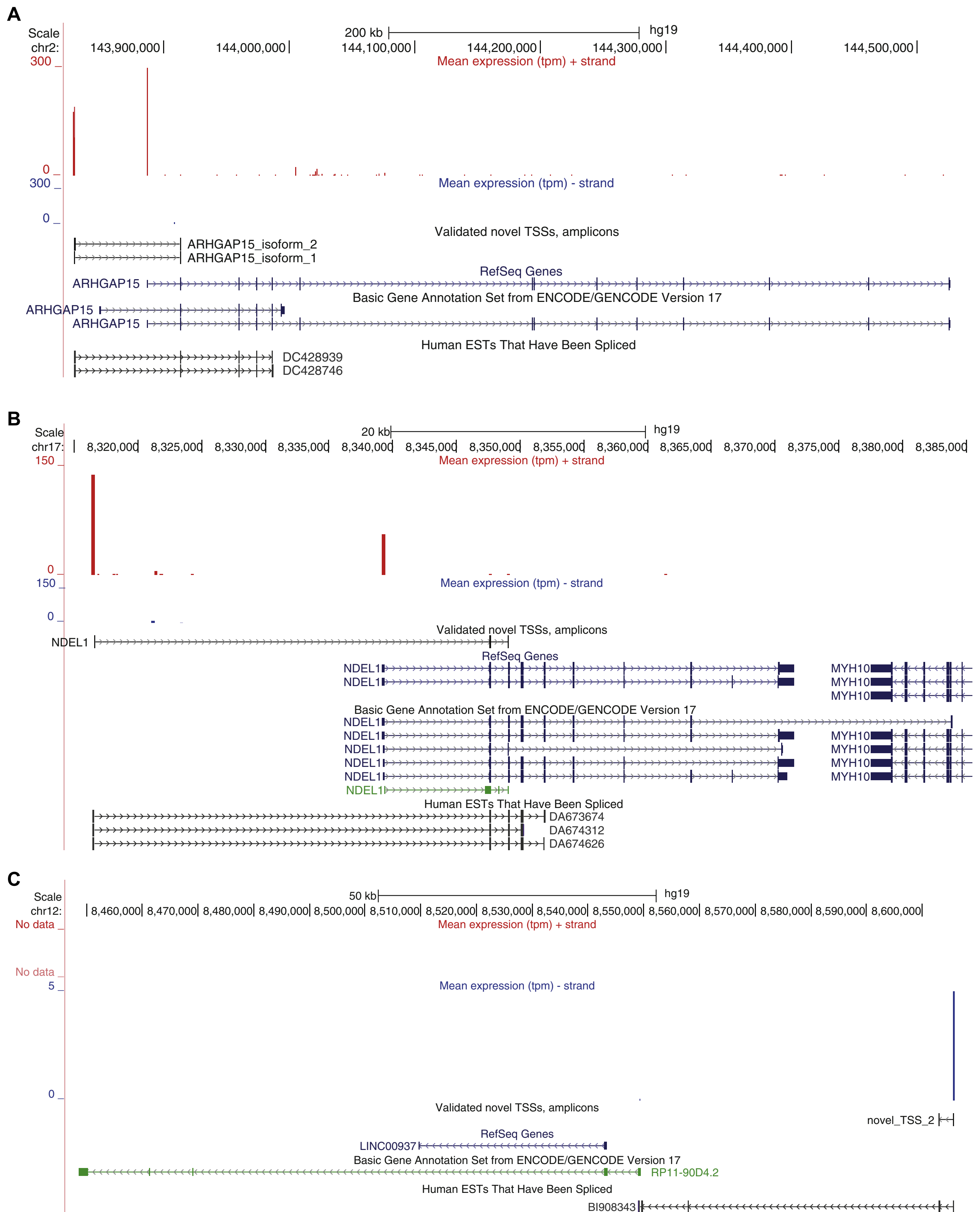

FIG 2. Expression and support for 3 differentially expressed novel TSSs validated by means of RT-PCR and Sanger sequencing. Only spliced expressed sequence tags *(ESTs)* supporting the novel TSS are shown. **A,** Alternative first exon for *ARHGAP15* used by 2 novel isoforms. **B,** Alternative first exon for *NDEL1*. **C,** Novel transcript, possibly an isoform of the lincRNA *LINC00937*. *tpm*, Tags per million reads.

TABLE IV. Top 10 significant TF motifs in the promoter region (−1000 to +200 bp) from MARA for the pairwise comparisons between sample groups

Motif	Score	Adjusted *P* value
Children with SA vs those with CA		
NKX6-1,2	−3.51	3.13E-04
XCPE1{core}	−5.79	1.16E-03
ZNF148	−5.37	1.84E-03
NFY{A,B,C}	−5.70	4.21E-03
MYF family	5.53	1.06E-04
BREu{core}	6.06	1.88E-03
ESR1	5.47	2.50E-03
MAFB	6.22	2.69E-03
ZBTB6	5.44	4.14E-03
ZNF238	4.47	5.37E-03
Children with CA vs CTRLs		
TLX1..3_NFIC{dimer}	−6.17	3.11E-03
TFCP2	−7.22	6.10E-03
MAFB	−5.03	9.53E-03
ELK1,4_GABP{A,B1}	−6.45	1.01E-02
HMX1	−5.18	1.17E-02
GATA4	−5.79	2.19E-02
STAT5{A,B}	7.77	1.14E-03
RUNX1..3	5.53	1.44E-03
LMO2	5.69	2.97E-03
ZNF238	5.40	2.11E-02
Children with SA vs CTRLs		
TLX1..3_NFIC{dimer}	−6.69	2.75E-03
LMO2	−5.61	4.40E-03
TFCP2	−9.32	6.43E-03
ELK1,4_GABP{A,B1}	−5.60	1.19E-02
MAFB	−6.21	1.53E-02
AHR_ARNT_ARNT2	−6.01	2.26E-02
STAT5{A,B}	5.18	1.49E-03
RUNX1..3	7.75	1.62E-03
ZNF238	4.47	1.22E-02
HMX1	5.21	1.71E-02

A negative score implies lower motif activity in the first group, and a positive score implies higher activity (see the Methods section in this article's Online Repository). *{core}*, Core promoter element; *{dimer}*, protein dimer.

DISCUSSION

We analyzed gene expression using CAGE in peripheral blood leukocytes from children with CA, children with SA, and CTRLs. We found that they have distinct gene expression profiles with implications for disease biology and that these differences are partly explained by the underlying regulatory TF networks. Gene network analysis identified decreased glucocorticoid receptor signaling and increased activity of the MAPK and Jun kinase cascades in children with SA. Many putative novel transcripts and TSSs were differentially expressed, and 77% (10/13) of those selected for experimental validation could be detected by using RT-PCR.

Pairwise comparison between sample groups identified 1305 significantly differentially expressed TCs, 1029 of which matched 816 known genes. This is similar to the number of differentially expressed genes in microarray data previously reported for a cohort in which these samples were included.[21] Although the microarray study only identified activation of NK cell–mediated cytotoxicity in children with CA, we here found upregulation of different gene sets within this pathway for both children with CA and those with SA compared with CTRLs. Because the increased blood neutrophil count in children with SA is likely to affect gene expression profiles, we also analyzed the CAGE data using GLMs to account for differences in leukocyte composition. Some processes, including upregulation of NK cell–mediated cytotoxicity and leukocyte migration, as well as downregulation of ribosomal proteins, in children with SA remained significant, even after adjustment.

The overlap of significantly differentially expressed genes between CAGE and microarray analysis was surprisingly small (approximately 5%). Although earlier work also found relatively large differences between these methods,[39] a comparison of $\log_2$ fold changes irrespective of *P* values demonstrated a reassuring overrepresentation of concordant versus discordant changes in gene expression (see Table E12 in this article's Online Repository at www.jacionline.org). One contrasting finding is the upregulation of bitter taste receptor genes in children with SA in the microarray study.[21] These genes are localized in a cluster on chromosome 12, overlapping the *PRR4/PRH1* genes, and alternative splicing of the *PRR4/PRH1* transcripts results in multiple isoforms encompassing different bitter taste receptor genes.[23] We found a few CAGE tags within this locus, but a strong TSS signal closely upstream of *PRR4/PRH1*, which might be the promoter for the overlapping bitter taste receptors. In agreement with the microarray study, we found that this promoter is significantly overexpressed in children with SA compared with CTRLs.[21] Irrespective of the differences, we see a number of advantages of using CAGE over microarrays: (1) a markedly improved separation of sample groups by differentially expressed TCs, as shown by using hierarchical clustering; (ii) the ability to explore transcriptional complexity, leading to identification of differentially expressed novel TSSs; and (iii) sensitive detection and reliable quantification of coding and noncoding transcripts, where especially the latter are often lowly expressed.

The current literature contains little data on severe asthma for comparison, but a recent study on adults reported activation of CD8$^+$ but not CD4$^+$ T cells in patients with severe asthma, whereas patients with nonsevere asthma did not differ from CTRLs.[16] The CD8$^+$ T cells from patients with severe and those with nonsevere asthma could also be separated by clustering on both of our sets of differentially expressed marker transcripts. Interestingly, both of the cited studies reported extensive deregulation of noncoding RNAs in patients with severe asthma.[16,21] Three of our 10 validated novel transcripts are putative long noncoding RNAs: one TSS upstream of an annotated long intergenic noncoding RNA (lincRNA) and 2 antisense transcripts. The first TSS was upregulated in children with SA, even after adjustment for cell counts. GENCODE-annotated noncoding RNAs upregulated in patients with SA included nuclear paraspeckle assembly transcript 1 *(NEAT1)* and *PINT (AC058791.2)*, a lincRNA recently associated with p53 signaling.[40]

Because asthma primarily affects the airways, it would be informative to analyze gene expression in cells from the respiratory tract, including inflammatory cells, bronchial epithelium, and smooth muscle. Bronchoscopies, including bronchial biopsy specimens, were not needed for the diagnosis of asthma, and the procedure would neither have been easy nor ethically defensible for these children, especially not for patients with severe asthma. Here we instead used total leukocytes from peripheral blood, a sample source that is easy

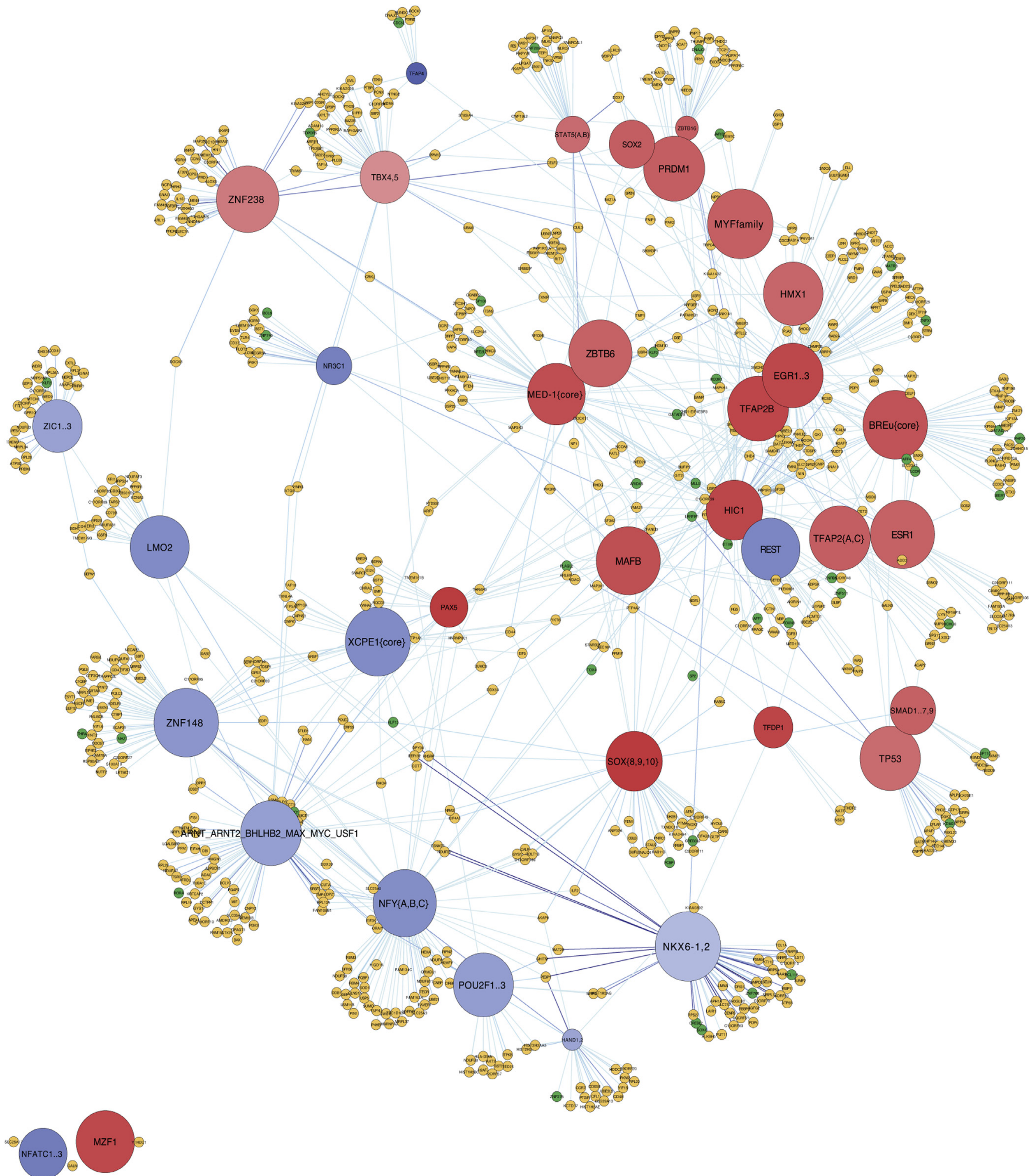

FIG 3. MARA identifies the TF network underlying differential gene expression between children with CA and those with SA, which are here significant motifs in the promoter region (−1000 to +200 bp). Node color corresponds to fold change (*blue*, lower motif activity in patients with SA; *red*, higher activity), and node size is inversely proportional to *P* value. Edge color reflects gene-specific motif activity (darker equals higher activity [ie, stronger regulatory effect]). Target genes are represented by *green circles* for TFs, and *yellow circles* for other genes.

to obtain and store and well suited for biomarker development. These data provide a wealth of information on the blood transcriptome that could be expanded, such as through studies of alternative promoter use and functional testing of novel transcripts and transcript isoforms. However, the different results from conventional and GLM differential expression analysis stress the importance of appropriately handling differences in sample cell composition when sorted cell populations are unavailable.

In summary, the clear separation of children with CA and those with SA shows the promise of gene expression profiling for the development of biomarkers given precisely posed clinical questions and appropriate validation cohorts. Differences in the disease biology of the 2 groups, as revealed by differential expression of both known and novel transcripts, support the need for novel therapeutic approaches.

We thank Dr Efthymios Motakis, Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan, for his expert advice on statistical analysis using GLMs.

Key messages

- **Circulating leukocytes from children with CA and those with SA have distinct gene expression profiles.**
- **Gene network analysis identified decreased glucocorticoid receptor signaling and increased activity of the MAPK and Jun kinase cascades in children with SA.**
- **The results suggest that it would be possible to develop molecular biomarkers for SA and support the need for novel therapeutic approaches in this group of patients.**

REFERENCES

1. Chronic respiratory diseases: asthma. Geneva (Switzerland): World Health Organization; 2013. Available at: http://www.who.int/respiratory/asthma/en/index.html. Accessed November 15, 2013.
2. Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma (GINA); 2012. Available at: http://www.ginasthma.org/. Accessed September 19, 2013.
3. Lang A, Carlsen KH, Haaland G, Devulapalli CS, Munthe-Kaas M, Mowinckel P, et al. Severe asthma in childhood: assessed in 10 year olds in a birth cohort study. Allergy 2008;63:1054-60.
4. Fleming L, Wilson N, Bush A. Difficult to control asthma in children. Curr Opin Allergy Clin Immunol 2007;7:190-5.
5. Hedlin G, Bush A, Lodrup Carlsen K, Wennergren G, De Benedictis FM, Melen E, et al. Problematic severe asthma in children, not one problem but many: a GA2LEN initiative. Eur Respir J 2010;36:196-201.
6. Aoki T, Matsumoto Y, Hirata K, Ochiai K, Okada M, Ichikawa K, et al. Expression profiling of genes related to asthma exacerbations. Clin Exp Allergy 2009;39:213-21.
7. Baines KJ, Simpson JL, Bowden NA, Scott RJ, Gibson PG. Differential gene expression and cytokine production from neutrophils in asthma phenotypes. Eur Respir J 2010;35:522-31.
8. Bjornsdottir US, Holgate ST, Reddy PS, Hill AA, McKee CM, Csimma CI, et al. Pathways activated during human asthma exacerbation as revealed by gene expression patterns in blood. PLoS One 2011;6:e21902.
9. Chen E, Miller GE, Walker HA, Arevalo JM, Sung CY, Cole SW. Genome-wide transcriptional profiling linked to social class in asthma. Thorax 2009;64:38-43.
10. Bochkov YA, Hanson KM, Keles S, Brockman-Schneider RA, Jarjour NN, Gern JE. Rhinovirus-induced modulation of gene expression in bronchial epithelial cells from subjects with asthma. Mucosal Immunol 2010;3:69-80.
11. Chamberland A, Madore AM, Tremblay K, Laviolette M, Laprise C. A comparison of two sets of microarray experiments to define allergic asthma expression pattern. Exp Lung Res 2009;35:399-410.
12. Choy DF, Modrek B, Abbas AR, Kummerfeld S, Clark HF, Wu LC, et al. Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways. J Immunol 2011;186:1861-9.
13. Kicic A, Hallstrand TS, Sutanto EN, Stevens PT, Kobor MS, Taplin C, et al. Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium. Am J Respir Crit Care Med 2010;181:889-98.
14. Laprise C, Sladek R, Ponton A, Bernier MC, Hudson TJ, Laviolette M. Functional classes of bronchial mucosa genes that are differentially expressed in asthma. BMC Genomics 2004;5:21.
15. Bosco A, Ehteshami S, Panyala S, Martinez FD. Interferon regulatory factor 7 is a major hub connecting interferon-mediated responses in virus-induced asthma exacerbations in vivo. J Allergy Clin Immunol 2012;129:88-94.
16. Tsitsiou E, Williams AE, Moschos SA, Patel K, Rossios C, Jiang X, et al. Transcriptome analysis shows activation of circulating CD8+ T cells in patients with severe asthma. J Allergy Clin Immunol 2012;129:95-103.
17. Hakonarson H, Bjornsdottir US, Halapi E, Bradfield J, Zink F, Mouy M, et al. Profiling of genes expressed in peripheral blood mononuclear cells predicts glucocorticoid sensitivity in asthma patients. Proc Natl Acad Sci U S A 2005;102:14789-94.
18. Melen E, Pershagen G. Pathophysiology of asthma: lessons from genetic research with particular focus on severe asthma. J Intern Med 2012;272:108-20.
19. Kanamori-Katayama M, Itoh M, Kawaji H, Lassmann T, Katayama S, Kojima M, et al. Unamplified cap analysis of gene expression on a single-molecule sequencer. Genome Res 2011;21:1150-9.
20. Konradsen JR, Nordlund B, Lidegran M, Pedroletti C, Gronlund H, van Hage M, et al. Problematic severe asthma: a proposed approach to identifying children who are severely resistant to therapy. Pediatr Allergy Immunol 2011;22:9-18.
21. Orsmark-Pietras C, James A, Konradsen JR, Nordlund B, Soderhall C, Pulkkinen V, et al. Transcriptome analysis reveals upregulation of bitter taste receptors in severe asthmatics. Eur Respir J 2013;42:65-78.
22. Just J, Gouvis-Echraghi R, Rouve S, Wanin S, Moreau D, Annesi-Maesano I. Two novel, severe asthma phenotypes identified during childhood using a clustering approach. Eur Respir J 2012;40:55-60.
23. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 2012;22:1760-74.
24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
25. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 2010;11:R14.
26. Costa C, Germena G, Martin-Conte EL, Molineris I, Bosco E, Marengo S, et al. The RacGAP ArhGAP15 is a master negative regulator of neutrophil functions. Blood 2011;118:1099-108.
27. FANTOM Consortium, RIKEN PMI, CLST (DGT). A promoter-level mammalian expression atlas. Nature 2014;507:462-70.
28. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930-4.
29. Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, et al. A catalog of published genome-wide association studies. Available at: http://www.genome.gov/gwastudies. Accessed May 7, 2013.
30. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet 2011;378:1006-14.
31. 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;363:1211-21.
32. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011;43:887-92.
33. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat Genet 2011;43:893-6.
34. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 2011;9:e1001046.
35. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 2003;31:374-8.
36. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 2010;38:D105-10.

37. Hudson NJ, Dalrymple BP, Reverter A. Beyond differential expression: the quest for causal mutations and effector molecules. BMC Genomics 2012;13:356.
38. Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet 2009;41:553-62.
39. Vitezic M, Lassmann T, Forrest AR, Suzuki M, Tomaru Y, Kawai J, et al. Building promoter aware transcriptional regulatory networks using siRNA perturbation and deepCAGE. Nucleic Acids Res 2010;38:8141-8.
40. Marin-Bejar O, Marchese FP, Athie A, Sanchez Y, Gonzalez J, Segura V, et al. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol 2013;14:R104.

Reviewer Board

The JACI Reviewer Board is selected by the Editor-in-Chief to acknowledge a strong record of service to the Journal. The Editors thank the current members of our Reviewer Board:

Daniel C. Adelman, MD
Ioana Agache, MD
Zuhair K. Ballas, MD, PhD
Donald H. Beezhold, PhD
Bruce G. Bender, PhD
M. Cecilia Berin, PhD
Malcolm N. Blumenthal, MD
Larry Borish, MD
Eric B. Brandt, PhD
Albrecht Bufe, MD, PhD
Javier Chinen, MD, PhD
Donald W. Cockcroft, MD, FRCP(C)
Scott P. Commins, MD, PhD
Ronina A. Covar, MD
Azzeddine Dakhama, PhD
Stephen C. Dreskin, MD, PhD
Kirk M. Druey, MD
Markus J. Ege, MD, MPH
Thomas Eiwegger, MD
Matthew Greenhawt, MD, MBA, MSc
Gabriele Grunig, DMV, PhD
David P. Huston, MD
Daniel Jackson, MD
John M. Kelso, MD
Alex KleinJan, PhD
Daphne Koinis-Mitchell, PhD
Fu-Tong Liu, MD, PhD
Eric M. Macy, MD
Bruce D. Mazer, MD
Akio Mori, MD, PhD
Ariel Munitz, PhD
Oscar Palomares, PhD
Lars K. Poulsen, PhD
Stephen I. Rosenfeld, MD
Ulla Seppälä, PhD
Jonathan M. Spergel, MD, PhD
P. Brock Williams, PhD
Darryl C. Zeldin, MD
Bruce L. Zuraw, MD