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

Persson, Helena

2015-09


http://hdl.handle.net/10138/203739
https://doi.org/10.1016/j.jaci.2015.02.026

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles

Helena Persson, PhD,a,b,c Andrew T. Kwon, PhD,b,c,d Jordan A. Ramiłowski, PhD,b,c,d Gilad Silberberg, PhD,d Cilla Söderhäll, PhD,a,e Cristina Orsmark-Pietras, PhD,a,e Björn Nordlund, RN, PhD,e,f,g Jon R. Konradsen, MD, PhD,e,f,g Michiel J. L. de Hoon, PhD,b,c,e Erik Meilén, MD, PhD,e,f,h,i Yoshihide Hayashizaki, MD, PhD,b,i Gunilla Hedlin, MD, PhD,e,f,g Juha Kere, MD, PhD,a,e,k,l and Carsten O. Daub, PhD,a,b,c Stockholms region, 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 children1 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 children3 and can cause extensive personal suffering, early deterioration of lung function, and significant consumption of health care resources.1 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.3

Global gene expression in human asthma has been studied in isolated leukocyte populations,6,8 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...
was approved by the local ethics committee. Informed consent was obtained from all participating children and their parents, and the study was previously registered in the Swedish Search study; details on inclusion criteria and clinical examination are provided in Fig E1 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.

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, 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.25 Leukocyte cell counts are provided for individual samples in Table E2 in this article’s Online Repository at www.jacionline.org.

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

Table I

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>CA (n = 15)</th>
<th>SA (n = 13)</th>
<th>CTRL (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methacholine resp</td>
<td>4.7 ± 2.1</td>
<td>7.2 ± 3.5</td>
<td>3.6 ± 1.8</td>
</tr>
<tr>
<td>Neutrophil count</td>
<td>2000 ± 1000</td>
<td>3000 ± 1500</td>
<td>1500 ± 500</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>CTRLs</th>
<th>Children with CA</th>
<th>Children with SA</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subjects</td>
<td>9</td>
<td>15</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>11 (5)</td>
<td>13 (3)</td>
<td>14 (3)</td>
<td>NS</td>
</tr>
<tr>
<td>Female sex (%)</td>
<td>11</td>
<td>40</td>
<td>38</td>
<td>NS</td>
</tr>
<tr>
<td>BMI (%), median (IQR)</td>
<td>59 (36-66)</td>
<td>62 (32-76)</td>
<td>77 (61-93)</td>
<td>NS</td>
</tr>
<tr>
<td>ACT score, mean (SD)</td>
<td>NA</td>
<td>24 (1)</td>
<td>16 (3)</td>
<td></td>
</tr>
<tr>
<td>ICS (μg), median (IQR)</td>
<td>NA</td>
<td>320 (180-400)</td>
<td>800 (800-800)</td>
<td></td>
</tr>
<tr>
<td>FEV1 (% predicted), mean (SD)</td>
<td>NA</td>
<td>86 (11)</td>
<td>88 (22)</td>
<td>NS</td>
</tr>
<tr>
<td>DRSmethacholine, median (IQR)</td>
<td>NA</td>
<td>0.8 (0.2-8.0)</td>
<td>11.1 (1.7-61.0)</td>
<td>.015 (CA-SA)</td>
</tr>
<tr>
<td>FENO (ppb), median (IQR)</td>
<td>NA</td>
<td>13 (9-21)</td>
<td>18 (11-38)</td>
<td>NS</td>
</tr>
<tr>
<td>Total IgE (kUA/L), median (IQR)</td>
<td>NA</td>
<td>200 (66-795)</td>
<td>500 (120-1000)</td>
<td>NS</td>
</tr>
<tr>
<td>Atopic* (%)</td>
<td>NA</td>
<td>80</td>
<td>85</td>
<td>NS</td>
</tr>
<tr>
<td>Total WBC (10^9 × L-1), mean (SD)</td>
<td>5.8 (0.8)</td>
<td>5.9 (1.5)</td>
<td>6.7 (2.2)</td>
<td>NS</td>
</tr>
<tr>
<td>Eosinophils (10^5 × L-1), mean (SD)</td>
<td>0.2 (0.1-0.2)</td>
<td>0.2 (0.2-0.4)</td>
<td>0.4 (0.2-0.4)</td>
<td>NS</td>
</tr>
<tr>
<td>Lymphocytes (10^7 × L-1), mean (SD)</td>
<td>2.4 (0.6)</td>
<td>2.7 (0.6)</td>
<td>2.3 (0.7)</td>
<td>NS</td>
</tr>
<tr>
<td>Monocytes (10^5 × L-1), mean (SD)</td>
<td>0.4 (0.1)</td>
<td>0.5 (0.2)</td>
<td>0.4 (0.1)</td>
<td>NS</td>
</tr>
<tr>
<td>Neutrophils (10^5 × L-1), mean (SD)</td>
<td>2.6 (0.5)</td>
<td>2.5 (0.7)</td>
<td>3.5 (1.6)</td>
<td>.046 (CA-SA)</td>
</tr>
</tbody>
</table>

ACT, Asthma Control Test; BMI, body mass index; DRSmethacholine, 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,
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. We analyzed 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.

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.

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. 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
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.” 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 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)
Upregulated in children with SA vs those with CA

KEGG:04666  FcγR-mediated phagocytosis  8.992e-06  ARPC1B, ARPC2, ASAP1, DOCK2, FCGR2A, GAB2, LIMK2, LIN, MAPK1, MARCKS, PAK1, PIK3R5, PLCG2, PRKCB, PRKCD, VASP, WAS

KEGG:04810  Regulation of actin cytoskeleton  2.109e-04  ARPC1B, ARPC2, CYFIP2, FGJD3, IQGAP1, IQGAP2, ITGAL, LIMK2, MAPK1, MSN, MYL12A, PAK1, PAK2, PIK3R5, PIKFYVE, PXN, RHOA, SOS2, SH2, TIAM1, WAS

KEGG:04380  Osteoclast differentiation  0.002  FCGR2A, FOS, FOSL2, GAB2, JUNB, LILRR2, MAP2K6, MAPK1, NCF2, NCF4, NFATC1, PIK3R5, PLCG2, SP1, TGFBR2

KEGG:4670  Leukocyte transendothelial migration  0.003  GNA2, ITGAL, MSN, MYL12A, NCF2, NCF4, PIK3R3, PLCG2, PRKCB, PXN, RHOA, VASP

KEGG:4062  Chemokine signaling pathway  0.007  ARRB2, DOCK2, GNAI2, GNG2, LIN, MAPK1, PK1, PIK3R5, PREX1, RAF, RHOA, ROS2, TIAM1, WAS

KEGG:5140  Bacterial invasion of epithelial cells  0.011  CR1, FCGR2A, FOS, MAPK1, NCF2, NCF4, PRKCB, PTPN6, TLR4

KEGG:5100  Bacterial invasion of epithelial cells  0.021  ARRC1B, ARPC2, GAB1, HCLS1, PIK3R5, PXN, RHOA, HOG, WAS

KEGG:5150  Staphylococcus aureus infection  0.021  C5AR1, FCGR2A, FPR1, ITGAL, PTAFR, SELPLG

KEGG:4650  Leukocyte–mediated cytotoxicity  0.025  CD48, ITGAL, MAPK1, NFATC1, PK1, PIK3R5, PLCG2, PRKCB, PTPN6, SOS2, TNFRSF10C

Downregulated in children with SA vs those with CA

GO:0030852  Regulation of granulocyte differentiation  0.017  GFI1B, IKZF1, OGT, RARA

GO:0046649  Lymphocyte activation  0.044  ATM, BCL11B, CD3, CK, FBKP1A, GRAP2, IKZF1, IKZF3, ITPKB, LAT, MS4A1, PIK3R3, RARA, TRAC

Upregulated in children with SA vs CTRLs

KEGG:4730  Long-term depression  0.030  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, RPS21, RPS2, RPSA

KEGG:0045944  Positive regulation of transcription from RNA polymerase II promoter  0.028  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  0.010  FOS, FOSB, JUNB, NFKB1A, SPI1

GO:0006357  Regulation of transcription from RNA polymerase II promoter  1.168e-05  CSA, EGR1, EGR2, ETS2, FOS, FOSB, ID1, JUNB, KLF13, NFIL3, NFKB1A, PLK3, RXRA, SP1, TXNIP, ZFP36

GO:0006955  Immune response  0.007  AMICA1, AQP9, C5AR1, CSF1, DUSP6, EGR1, FO, NFKB1A, NFKBIA, PTEN, TXNIP

GO:0008219  Cell death  0.021  CSA, DUSP6, EGR3, EIF4G2, ID1, IRS2, NFKB1A, PLK3, PREX1, PTEN, RXRA, SRCN, TXNIP

*Adjusted \( P < 0.1 \) after adjustment for cell counts in the GLM analysis.

significantly associated with asthma in genome-wide association studies. HaploReg v2\(^2^8\) was used to identify linkage disequilibrium (LD) blocks for 62 SNPs with genome-wide significant association with asthma from 17 studies.\(^2^9\) 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)\(^3^0\) and IL-2 receptor \(\beta \) (IL2RB)\(^3^1\) 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 melanoma 2 (AIM2)\(3^2\) overlapped an LD block associated with asthma in subjects of African descent.\(^3^2\) Furthermore, 2 differentially expressed TCs were in the genes GRB2-associated binding protein 1 (GAB1)\(^3^3\) and RAR-related orphan receptor A (RORA),\(^3^1\) 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.\(^3^2\) 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\(^3^2\) and is predicted to change a binding motif for the peroxisome proliferator-activated receptors.\(^3^5,3^6\) 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
Differentially expressed novel TSSs (adjusted $P < .1$) selected for experimental validation by using RT-PCR and Sanger sequencing

<table>
<thead>
<tr>
<th>TSS</th>
<th>Gene</th>
<th>Mean tpm</th>
<th>Log fold change (CTRLs vs children with CA)</th>
<th>Log fold change (CTRLs vs children with SA)</th>
<th>Log fold change (Children with CA vs those with SA)</th>
<th>Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr2:24714142(+)</td>
<td>NCOA1</td>
<td>57</td>
<td>NS</td>
<td>NS</td>
<td>0.39</td>
<td>No</td>
</tr>
<tr>
<td>chr2:143828538(+)</td>
<td>ARHGAP15</td>
<td>175</td>
<td>NS</td>
<td>0.97</td>
<td>0.99</td>
<td>Yes</td>
</tr>
<tr>
<td>chr4:148973914(+)</td>
<td>ARHGAP10</td>
<td>3</td>
<td>NS</td>
<td>NS</td>
<td>$-1.02$</td>
<td>Yes</td>
</tr>
<tr>
<td>chr6:131370837(+)</td>
<td>ASAP1</td>
<td>105</td>
<td>NS</td>
<td>NS</td>
<td>0.57</td>
<td>No</td>
</tr>
<tr>
<td>chr9:117160251(+)</td>
<td>AKNA/RP11-9M16.2</td>
<td>5</td>
<td>NS</td>
<td>NS</td>
<td>0.63</td>
<td>Yes</td>
</tr>
<tr>
<td>chr13:46948966(+)</td>
<td>KJAA0226L</td>
<td>25</td>
<td>NS</td>
<td>0.94</td>
<td>NS</td>
<td>Yes</td>
</tr>
<tr>
<td>chr15:86087320(+)</td>
<td>AKAP13</td>
<td>30</td>
<td>NS</td>
<td>0.78</td>
<td>0.63</td>
<td>Yes</td>
</tr>
<tr>
<td>chr16:90023689(+)</td>
<td>DEF8</td>
<td>35</td>
<td>NS</td>
<td>NS</td>
<td>0.43</td>
<td>Yes</td>
</tr>
<tr>
<td>chr17:8316490(+)</td>
<td>NDELI</td>
<td>137</td>
<td>NS</td>
<td>0.61</td>
<td>0.68</td>
<td>Yes</td>
</tr>
<tr>
<td>chr6:105901520(+)</td>
<td>Possibly lincRNA</td>
<td>4</td>
<td>NS</td>
<td>NS</td>
<td>1.01</td>
<td>No</td>
</tr>
<tr>
<td>chr12:8605756(+)</td>
<td>Possibly lincRNA</td>
<td>5</td>
<td>NS</td>
<td>NS</td>
<td>$-1.01$</td>
<td>No</td>
</tr>
<tr>
<td>chr3:13431406(+)</td>
<td>KY</td>
<td>33</td>
<td>NS</td>
<td>NS</td>
<td>0.93</td>
<td>Yes</td>
</tr>
<tr>
<td>chr14:95652012(+)</td>
<td>CLMN/CTD-2240H23.2</td>
<td>19</td>
<td>NS</td>
<td>NS</td>
<td>0.58</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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 mathematical 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](http://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](http://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](http://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](http://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.
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, a comparison of log2 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. 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. 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. Irrespective of the differences, we see a number of advantages of using CAGE over microarrays: (i) 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. 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. 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.

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