Immune cell transcriptome datasets reveal novel leukocyte subset–specific genes and genes associated with allergic processes

Sue M. Liu, PhD,[a,b,c] Ramnik Xavier, MD, PhD,[d] Kim L. Good, BSc, Hons,[a,e] Tatyana Chtanova, PhD,[a,b,c] Rebecca Newton, PhD,[a] Mary Sisavanh, BSc, Hons,[a,b,c] Sabine Zimmer, PhD,[a,b] Chaoyang Deng, PhD,[f] Diego G. Silva, MD, PhD,[g] Melinda J. Frost, BSc, Hons,[a,b] Stuart G. Tangye, PhD,[a,e] Michael S. Rolph, PhD,[a,b,c] and Charles R. Mackay, PhD[a,b,c] *Sydney, Camperdown, and Canberra, Australia, and Boston and Cambridge, Mass*

Background: The precise function of various resting and activated leukocyte subsets remains unclear. For instance, mast cells, basophils, and eosinophils play important roles in allergic inflammation but also participate in other immunologic responses. One strategy to understand leukocyte subset function is to define the expression and function of subset-restricted molecules.
Objective: To use a microarray dataset and bioinformatics strategies to identify novel leukocyte markers as well as genes associated with allergic or innate responses.
Methods: By using Affymetrix microarrays, we generated an immune transcriptome dataset composed of gene profiles from all of the major leukocyte subsets, including rare enigmatic subsets such as mast cells, basophils, and plasma cells. We also assessed whether analysis of genes expressed commonly by certain groups of leukocytes, such as allergic leukocytes, might identify genes associated with particular responses.
Results: Transcripts highly restricted to a single leukocyte subset were readily identified (>2000 subset-specific transcripts), many of which have not been associated previously with leukocyte functions. Transcripts expressed exclusively by allergy-related leukocytes revealed well known as well as novel molecules, many of which presumably contribute to allergic responses. Likewise, Nearest Neighbor Analysis of genes coexpressed with Toll-like receptors identified genes of potential relevance for innate immunity.
Conclusion: Gene profiles from all of the major human leukocyte subsets provide a powerful means to identify genes associated with single leukocyte subsets, or different types of immune response.
Clinical implications: A comprehensive dataset of gene expression profiles of human leukocytes should provide new targets or biomarkers for human inflammatory diseases.
(J Allergy Clin Immunol 2006;118:496-503.)

Key words: *Allergy, gene profiling, leukocyte-specific, Toll-like receptors*

Host responses to pathogens involve cooperation between numerous leukocyte types, each with distinct functional properties. Over the period of the past 30 years, a principle strategy used to understand leukocyte subset function has been the characterization of molecules expressed by these subsets. Since the 1980s, mAbs have been instrumental for identifying leukocyte subsets and revealing the function of molecules and cell types. Recently, however, sequencing the human genome and development of gene microarray technologies have provided new means of identifying genes expressed and regulated in leukocyte subsets. Transcriptional profiles of a range of leukocytes at various stages of differentiation and activation have been reported, including naive and activated T cells,[1] T_H1 and T_H2,[2] follicular B T_H,[3] various B-cell subsets and lymphomas,[4-6] mast cells,[7,8] eosinophils,[9,10] macrophages, dendritic cells (DCs),[11-13] and natural killer (NK) cells.[14] To date, most microarray studies have involved just a single cell type and only a limited number of variables, such as type of stimulation or stage of differentiation. In addition, gene microarrays have been used to establish signatures for different types of immune responses[15] or different aspects of cell physiology. For instance, specific sets of genes are characteristically transcribed in response to cellular activation or proliferation,[16] and signature genes have been

From [a]the Immunology and Inflammation Research Program, Garvan Institute of Medical Research, Sydney; [b]the Cooperative Research Centre for Asthma, University of Sydney; [c]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Sydney; [d]the Department of Medicine and Molecular Biology, Massachusetts General Hospital, Boston; [e]the Centenary Institute of Cancer Medicine and Cell Biology, Camperdown; [f]the Center for Cancer Research, Massachusetts Institute of Technology, Cambridge; and [g]the John Curtin School of Medical Research, Australian National University, Canberra.
Supported by the National Health and Medical Research Council and the Cooperative Research Centre for Asthma. S. Liu is supported by National Health and Medical Research Council Dora Lush postgraduate and CRC for Asthma scholarships.
Disclosure of potential conflict of interest: The authors have declared that they have no conflict of interest.
Received for publication December 15, 2005; revised April 19, 2006; accepted for publication April 24, 2006.
Available online June 22, 2006.
Reprint requests: Charles R. Mackay, PhD, Immunology and Inflammation Research Program, Garvan Institute of Medical Research, 384 Victoria Street, Sydney, 2010, Australia. E-mail: c.mackay@garvan.org.au.
0091-6749/$32.00

doi:10.1016/j.jaci.2006.04.040

Abbreviations used
DC: Dendritic cell
MACS: Magnetic-activated cell sorting
NK: Natural killer
PC: Plasma cell
PMA: Phorbol 12-myristate 13-acetate
TLR: Toll-like receptor

described for allergic airway inflammation,[17,18] innate immune responses to pathogens,[19] and numerous cancers.[5]

Leukocyte subsets express certain molecules exclusively, and these usually underlie the specific function of that subset. For instance, B cells express immunoglobulin, mast cells express mast cell tryptase, and eosinophils express eosinophil peroxidase. Similarly, many molecules are expressed by groups of leukocyte subsets, particularly those that reflect the common role of a molecule in a particular immune response. For instance, the chemokine receptor CCR7 facilitates migration of T cells, B cells, and DCs to lymphoid tissues, enabling the interaction of these cells in primary and secondary immune responses.[20,21] Similarly, molecules such as IL-4, IL-13, CCR3, and chemoattractant receptor preferentially expressed on T_H2 cells are preferentially expressed by allergic-type leukocytes and are strongly implicated in the pathogenesis of allergic diseases.[22-24]

We used oligonucleotide microarrays to generate gene transcription profiles for virtually all of the major human leukocyte types. This comprehensive dataset enabled us to identify leukocyte subset-restricted genes using simple bioinformatics strategies. We used this dataset to assess gene expression in groups of leukocytes and, as an example, assessed genes commonly expressed in leukocytes associated with allergic responses (mast cells, basophils, and eosinophils) or coexpressed with Toll-like receptors (TLRs) and, hence, likely to be involved in innate immunity. The description herein of a comprehensive and publicly available immune transcriptome dataset and established methods for data analysis should prove a valuable resource for the identification of molecules and pathways associated with different immune responses.

METHODS

Leukocyte isolation, differentiation, and activation

All experiments were approved by the Human Research Ethics Committee, St Vincent's Hospital (Sydney, Australia). Leukocyte subsets were purified from blood from healthy volunteers. Isolation of central and effector memory T cells from PBMCs,[3] and of plasma cells (PCs), naive B cells, and memory B cells that expressed IgM (nonswitched) or IgG/A/E (isotype switched) from spleen was previously described.[25-27] PC chips were run using pooled cRNA from as many as 5 donors. $CD19^+$ B cells and $CD16^+CD56^+$ NK cells were sorted from PBMCs using a FACSVantage SE DiVa (BD Biosciences, San Jose, Calif). Doublets and aggregated cells were excluded to minimize contaminating cells.

For basophil isolation, $CD3^+$ cells and $CD14^+$ cells were depleted from PBMCs by magnetic-activated cell sorting (MACS) (Miltenyi Biotec, Auburn, Calif). Resultant cells were stained with 5E8 anti-CCR3 antibody and labeled with Cy5-conjugated donkey antimouse IgG (Jackson Laboratories, Bar Harbor, Me). $CCR3^{hi}$, forward-scatter and side-scatter low cells were collected. RNA was pooled from 2 donors for each microarray analysis.

Macrophages and DCs were differentiated from peripheral blood monocytes isolated using CD14 positive selection by MACS. Macrophages were differentiated in the presence of 500 U/mL GM-CSF (BD Biosciences) for 11 days and then activated for 4 hours with 100 ng/mL LPS (Sigma, St Louis, Mo). DCs were differentiated in the presence of 800 U/mL IL-4 (BD Biosciences) and 1500 U/mL GM-CSF and cultured for 5 days. Nonadherent immature DCs were harvested after day 5, either for RNA isolation or activation by adding 100 ng/mL LPS for 48 hours. Macrophages were identified as $CD14^{hi}$/HLA-DR^{int}/$CD40^{int}$/$CD80^-$/$CD86^+$, immature DCs were identified as $CD14^-$/HLA-DR^{int}/$CD1a^+$/$CD40^+$/$CD80^-$/$CD83^-$/$CD86^+$, and mature DCs were identified as $CD14^-$/HLA-DR^{hi}/$CD1a^+$/$CD40^{hi}$/$CD80^+$/$CD83^+$/$CD86^{hi}$ by flow cytometry.

Granulocytes were isolated using a 70%/80% isotonic Percoll (Amersham, Buckinghamshire, United Kingdom) gradient.[28,29] Erythrocytes were eliminated using the Whole Blood Erythrocyte Lyzing Kit (R&D Systems, Minneapolis, Minn). Eosinophils were removed by incubation with anti-CCR3 followed by depletion using an antimouse secondary antibody conjugated to MACS beads. RNA was extracted from resting neutrophils and neutrophils stimulated for 1 hour with 100 ng/mL LPS.

Eosinophil isolation from peripheral blood granulocytes was adapted from published methods.[29] Neutrophils were depleted from the granulocyte population using anti-CD16 MACS beads. Eosinophils were then incubated with phorbol 12-myristate 13-acetate (PMA; 50 ng/mL), or with control medium for 2 hours. RNA extracted from 3 donors was then pooled for 1 microarray experiment, and a sample from a fourth donor was used for a second experiment.

Mast cells were derived from human cord blood obtained from the Sydney Cord Blood Bank and differentiated in the presence of IL-6 (BD Biosciences), IL-10 (BD Biosciences), and stem cell factor[30] (a gift from Amgen). Mature mast cells were activated by FcεRI cross-linking with human IgE anti-nitrophenyl (Serotec, Oxford, United Kingdom) followed by 2 hours with mouse antihuman IgE (Serotec) as previously described.[31]

RNA extraction and GeneChip hybridizations

HG-U133A and B GeneChips (Affymetrix, Santa Clara, Calif) were each hybridized with cRNA synthesized from RNA from single donors, except for experiments involving PCs, eosinophils, and basophils. RNA was isolated using the RNeasy Total RNA Isolation Kit (Qiagen, Chatsworth, Calif) or TRIzol reagent (Invitrogen Life Technologies, Mt Waverley, Australia). cRNA was prepared, and GeneChips were hybridized and scanned as previously described.[3]

Leukocyte signature analysis

Leukocyte-specific genes were defined as genes constitutively expressed by a particular leukocyte type and consistently absent in all other leukocyte subsets profiled. Only those genes showing similar patterns of expression in duplicate GeneChip experiments were considered for further analysis—that is, genes were considered to be present only if they were consistently detected. To increase the stringency of genes defined to be unique to a certain leukocyte type, genes present in naive and memory B cells (Good et al, Unpublished data, March 2004), γδ T cells, follicular B T_H, T_H1 and T_H2, and naive and central memory T-cell subsets[32] were subtracted from the initial

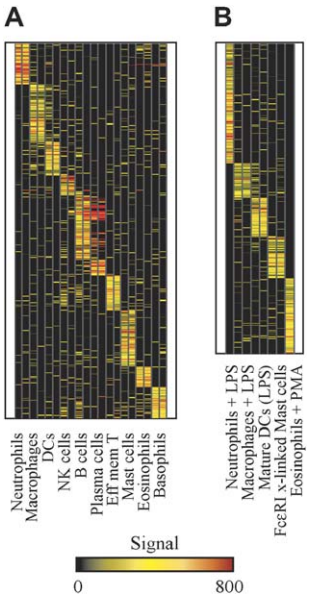

FIG 1. (A) Resting and **(B)** activated leukocyte-specific genes. Only genes not already represented in *Fig 2,* A, were included in activated leukocyte-specific genes in *Fig 2,* B. The color scale indicates transcript expression signal where *black* = absent, *yellow* = moderately expressed, and *red* = highly expressed. *Eff mem T,* Effector memory T cells.

leukocyte specific genes identified. Splenic B-cell and PC transcripts were not subtracted from the B-cell profile, and other T-cell subset transcripts were not subtracted from the effector memory T-cell profile. Transcripts found to be absent by MAS 5.0 software (Affymetrix) were reassigned a signal value of 0. Gene expression patterns were mapped using Spotfire DecisionSite (Somerville, Mass).

Nearest Neighbor Analysis of TLRs

Pearson correlation coefficients were calculated for each pair of TLR genes, and hierarchical clustering was performed using the Complete Linkage (maximum) method (Spotfire software) with correlation function. This analysis was performed using normalized leukocyte datasets. To further identify components in the TLR pathway, we performed Neighborhood Analysis on TLRs by taking each of the genes represented on the Affymetrix GeneChips and ranking their expression in neighborhood radii of 100, 250, and 500, with 100 the closest neighbors. These neighborhoods were then queried for the presence of genes coexpressed with 4 or more TLR genes to generate the TLR neighborhood index.

RT-PCR

cDNA was synthesized from 200 ng RNA using Reverse-IT RTase Blend Kit (ABgene, Epsom, United Kingdom). PCR was conducted using primers designed from Genbank sequences using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) or as referenced (for primer sequences, see this article's Table E1 in the Online Repository at www.jacionline.org) and synthesized by Geneworks (Sydney, Australia). RT-PCR was conducted using 1 μL cDNA template, 2.5 mmol/L $MgCl_2$, 0.4 μmol/L primer mix, and 5 U Taq polymerase (Promega, Sydney, Australia) in a final volume of 25 μL. After denaturation at 95°C for 5 minutes, the reactions were run for 30 cycles at 95°C, 62°C, and 72°C for 30 seconds each.

RESULTS

A comprehensive gene expression dataset for all of the major human leukocyte subsets

By using Affymetrix HG-U133A and B arrays, we screened >44,000 probe sets representing ~39,000 transcripts to develop transcript profiles for all of the major immune cell types including neutrophils, macrophages, DCs, NK cells, B-cell and T-cell subsets, mast cells, eosinophils, and basophils (see this article's Table E2 in the Online Repository at www.jacionline.org). Samples from at least 2 individuals were processed, and each was analyzed by microarray. The complete set of microarray data for all leukocyte subsets is available at http://linkage.garvan.unsw.edu.au/public/microarrays/ and the Gene Expression Omnibus database (accession number GSE3982).

Identification of leukocyte subset signatures and leukocyte-specific genes

The integration of a comprehensive collection of gene profiles of the main leukocyte types enabled identification of those genes uniquely expressed by each leukocyte. Specific gene expression signatures could be distinguished for all of the major human leukocytes after the genes and individual arrays had been arranged using unsupervised clustering (see this article's Fig E1 in the Online Repository at www.jacionline.org). Our analysis identified large numbers of specific genes for neutrophils (144), macrophages (203), DCs (118), NK cells (71), B cells (226), PCs (57), effector memory T cells (121), mast cells (199), eosinophils (73), and basophils (108; Fig 1, *A*). Because activated leukocytes represent the cells most closely associated with many immunologic functions, we also developed leukocyte subset signatures from LPS-stimulated neutrophils, macrophages, and DCs; mast cells activated via the FcεRI; and PMA-stimulated eosinophils (Fig 1, *B*). The 10 most abundantly expressed resting and activated leukocyte subset–specific gene transcripts are listed in this article's Tables E3 and E4 in the Online Repository at www.jacionline.org. The complete lists of leukocyte-specific genes are available at http://linkage.garvan.unsw.edu.au/public/microarrays/.

As expected, genes restricted to specific leukocyte types encoded molecules that are well known and currently used to define each of these cell types (see this article's Table E3 in the Online Repository at www.jacionline.org). For example, definitive mast cell genes like the tryptases (α and β) were identified as uniquely expressed by mast cells. CD209/DC-specific ICAM3-grabbing nonintegrin has been used to identify DCs, and the expression of this receptor was restricted to DCs in our transcriptome profiles. B-cell–specific transcripts included molecules intimately related to B-cell function such as immunoglobulin subunits (heavy and light chains) as well as 2 receptors, B cell maturation factor (BCMA) and transmembrane activator and CAML interactor (TACI), for the B-cell survival factor B cell–activating factor (BAFF) (see this article's Table E5 in the Online Repository at www.jacionline.org). The transcripts for

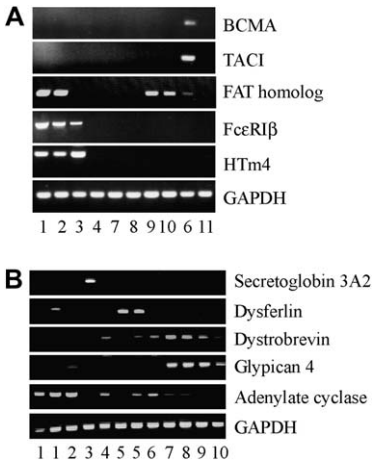

FIG 2. Confirmation of transcript expression discovered by microarray analysis. Transcript expression was assessed by RT-PCR using specific primers. Mast cells and neutrophils from more than 1 donor were used **(B)**. 1 = Mast cells, 2 = activated mast cells, 3 = basophils, 4 = eosinophils, 5 = neutrophils, 6 = B cells, 7 = immature DCs, 8 = mature DCs, 9 = macrophages, 10 = activated macrophages, 11 = NK cells. *BCMA*, B cell maturation factor; *TACI*, transmembrane activator and CAML interactor; *GAPDH*, glyceraldehyde-3 phosphate dehydrogenase.

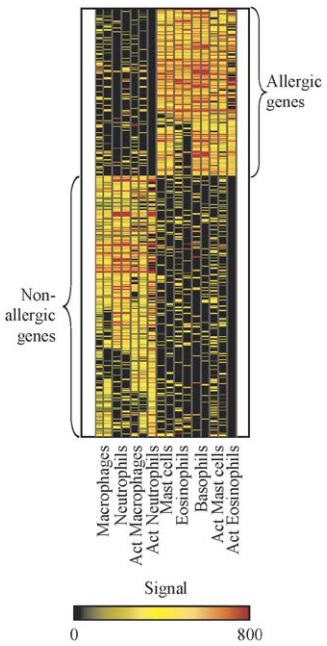

FIG 3. Gene expression patterns associated with functional leukocyte clusters. A total of 173 genes expressed in mast cells, eosinophils, and basophils and not macrophages and neutrophils were defined as allergic leukocytes. A total of 269 nonallergic leukocyte genes were defined as those that were present in macrophages and neutrophils and absent in mast cells, eosinophils, and basophils. The color scale indicates transcript signal where *black* = absent, *yellow* = moderately expressed, and *red* = highly expressed. *Act,* Activated.

these 2 BAFF receptors were expressed exclusively by B cells, as shown previously by mAb staining,[33] and confirmed by RT-PCR (Fig 2, *A*). Thus, the identification of these hallmark genes for the various leukocyte subsets validated our approach to identify subset-specific transcripts.

Remarkably, the majority of leukocyte subset-specific genes, even the most highly expressed transcripts, have not previously been associated with that particular subset. Examples of these novel leukocyte-specific genes included secretoglobin 3A2, which was restricted to basophils, dysferlin to neutrophils, and FAT tumor suppressor homolog to mast cells. By using RT-PCR, we confirmed that secretoglobin mRNA was indeed restricted to basophils (Fig 2, *B*) and that expression of dysferlin and FAT homolog was largely restricted to neutrophils and mast cells, respectively. Dystrobrevin α, glypican 4, and adenylate cyclase activating polypeptide 1 were also confirmed to be preferentially expressed by DCs, macrophages, and mast cells, respectively (Fig 2, *B*). Moreover, a large number of uncharacterized genes was identified by this analysis. Among novel genes discovered were membrane-associated molecules, identified using Gene Ontology cellular component annotations (NetAffx analysis tool, www.affymetrix.com, September 2005 update; see this article's Tables E5 and E6 in the Online Repository at www.jacionline.org).

Genes common to allergic effector cells

As a further example of the many possible applications for the comprehensive immune cell transcriptome, we assessed the gene expression profiles of certain effector cells critical for allergic inflammatory responses, mast cells, eosinophils, and basophils; this group of leukocyte subsets is hereafter referred to as *allergic leukocytes*. Allergic leukocyte genes were identified as those expressed in all 3 allergic leukocytes, either resting and/or activated, but not in resting or activated neutrophils and macrophages (Fig 3). By applying this contrived, but nevertheless relevant, type of analysis to the immune transcriptome, 173 genes were found to be preferentially expressed by allergic versus nonallergic leukocytes. Several of these genes have previously been associated with allergic inflammation, including chemoattractant receptor preferentially expressed on T_H2 cells and IL-5 receptor α.[23,34,35] There were 269 preferentially expressed probe sets for nonallergic leukocyte genes, arbitrarily defined as genes present in both neutrophils and macrophages, whether resting and/or activated, but absent in allergic leukocytes (Fig 3).

A subtraction procedure was used to identify genes restricted to allergic effector cells. After genes expressed by other leukocytes were subtracted, transcripts involved in vesicle transport, exocytosis, proteolysis, and gene transcription were among those identified as restricted to allergic effector cells. Examples of these transcripts are listed (Fig 4, *A*). There were 10 transcripts commonly expressed by the allergic leukocytes that were absent from all other leukocyte subsets profiled (Fig 4, *B*). The selective expression of 2 of these genes, FcεRIβ and hematopoietic cell 4–transmembrane protein on mast cells and basophils, was confirmed by RT-PCR (Fig 2, *A*). Further cross-validation

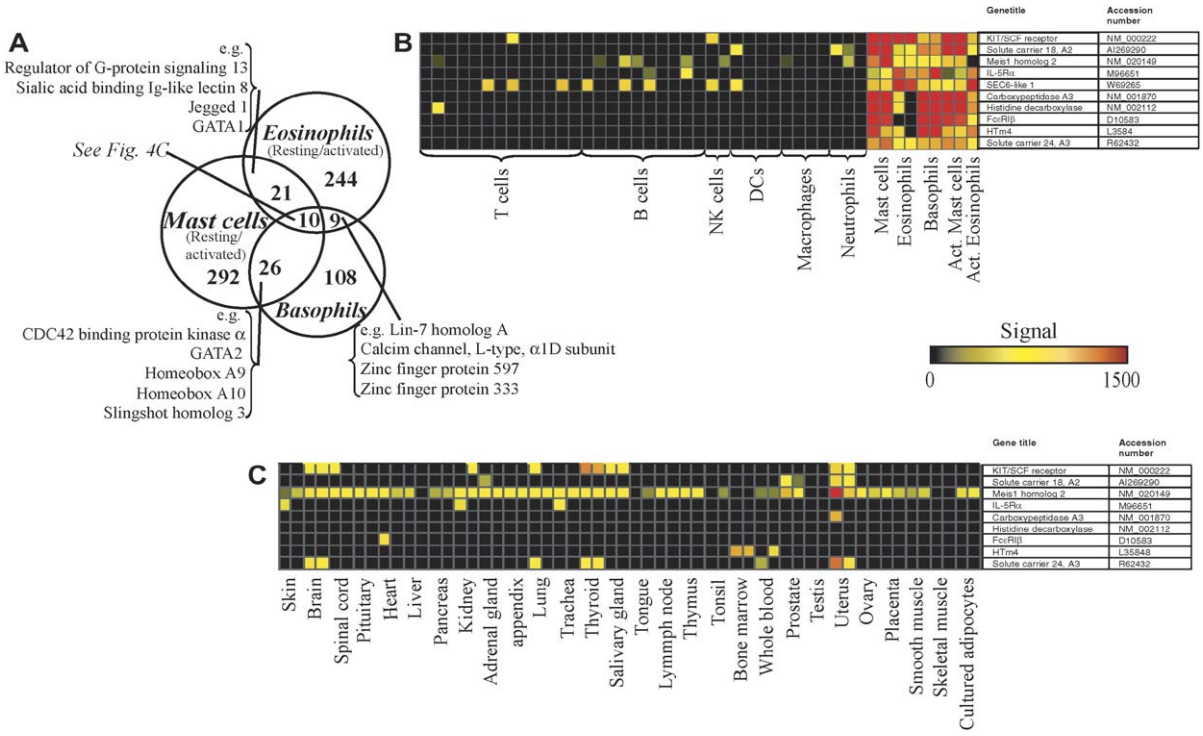

FIG 4. Specific allergic leukocyte genes. **A,** The number of genes expressed by allergic leukocytes not consistently detected in all of the other leukocytes we profiled. **B,** Specific allergic leukocyte genes were identified as those common to allergic leukocytes but consistently undetected in all of the other leukocytes we profiled. **C,** Expression pattern of specific allergic leukocyte genes *(B)* in different tissues.[36] *Act,* Activated; *SCF,* stem cell factor.

of our analyses using an independent database (http://symatlas.gnf.org/SymAtlas/)[36] revealed that almost all of the allergic leukocyte-specific transcripts were also largely undetectable in the major organs of the body (Fig 4, *C*). Conversely, the homeobox gene Meis 1 homolog 2 was widely distributed in many of the major organs, yet was absent in all nonallergic leukocytes profiled (Fig 4, *C*). Only 9 of the 10 specific allergic leukocyte genes were represented in this database.

A comprehensive immune cell GeneChip dataset to assess expression of molecular families *in silico*

Another powerful and logical approach to study gene regulation is to focus the analysis on molecular families. Such approaches have been used to generate individual gene expression profiles in many different cell types and tissues.[36,37] Our immune cell transcriptome dataset incorporates some of the important paradigms of immunology, such as T_H1/T_H2 differentiation, B-cell differentiation, effector cell activation, and DC maturation. We examined the expression of all of the known members of 2 important gene families: TLRs (Fig 5, *A*) and chemoattractant receptors (see this article's Fig E2 in the Online Repository at www.jacionline.org). TLRs are receptors for pathogen-associated molecular patterns that direct innate immune responses. Cell types associated with innate immune responses, such as neutrophils, macrophages, and DCs, expressed a wide range of TLRs (Fig 5, *A*). In contrast, T cells, which orchestrate adaptive immune responses, expressed few TLRs. Although B cells have a major role in adaptive immunity, they also possess antigen-presenting properties[38] and were found to cluster between cells of innate immunity and T-cell subsets (Fig 5, *A*). Expression analysis of chemoattractant receptors distinguished 2 broad categories of receptors: those associated with lymphoid tissue homing, and those more closely associated with inflammatory cells such as neutrophils, eosinophils, and activated macrophages (see this article's Fig E2 in the Online Repository at www.jacionline.org). A similar functional division of chemokine receptors was recently reviewed.[39]

In *Saccharomyces* and *Caenorhabditis elegans,* broad categories of functionally related genes are coexpressed, such as cell cycle regulators and components of organelles including mitochondria.[40] Recently, this functional clustering of coexpressed genes has been extended to vertebrates and exploited in 2 situations, to identify a gene associated with a genetically linked disease and a mechanism of action of cyclin D1 in cancer.[41,42] We performed Nearest Neighbor Analysis to identify genes coexpressed with each of the TLRs. There were 17 genes identified within the neighborhood of 4 or more TLR genes within a radius of 100 nearest neighbors, 137 within a radius of 250, and 376 within a radius of 500 (Fig 5, *B*). There

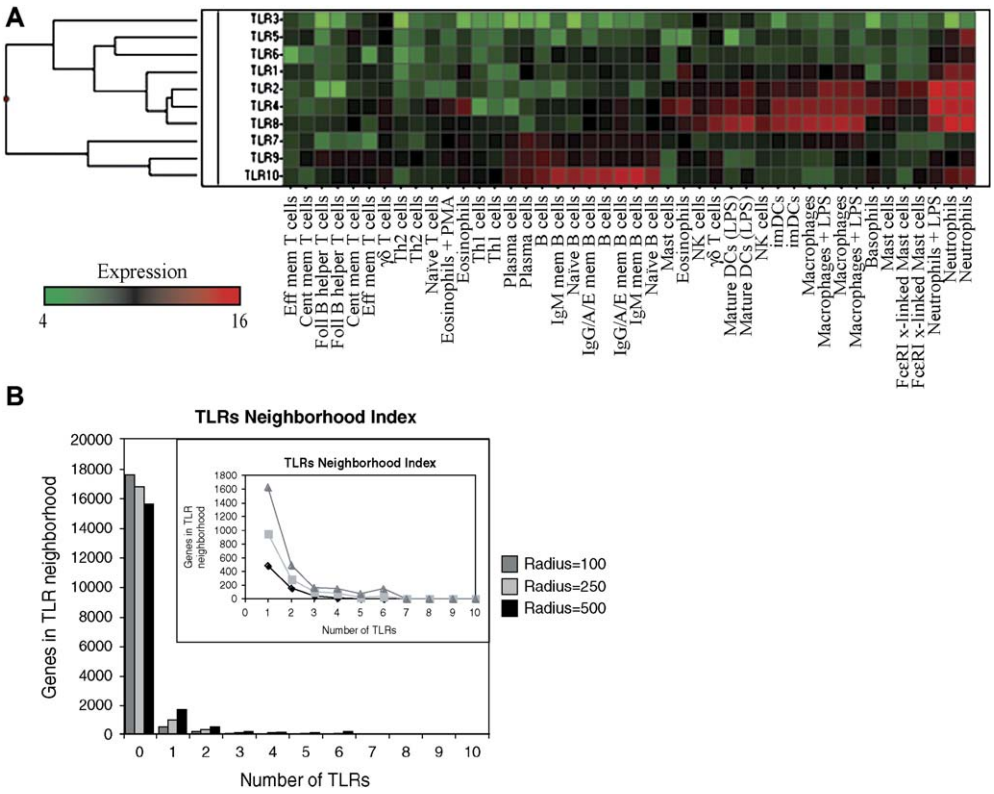

FIG 5. TLR expression pattern. **A,** Hierarchical clustering of TLR1-10 on leukocyte subsets with normalized expression levels represented by color. TLR1, 2, 4, 5, 6, 8 were grouped as an inner cluster. **B,** Neighborhood analysis of the 10 TLR genes. The number of genes within the neighborhood of TLR genes within a radius of 100, 250, or 500 is shown. The color scale indicates normalized expression; 0 = low, 1 = normal, and >1 = high. *EFF*, Effector; *cent*, central; *foll*, follicular; *mem*, memory; *imDC*, immature DC.

was a peak at the index of 6 TLRs in the neighborhood radius of 250, which may represent a closely coregulated subset. These 35 genes found in the neighborhood of 6 TLR genes within a radius of 250 (Table I) were found to be neighbors to the same 6 TLR genes: TLR1, 2, 4, 5, 6, and 8. We then calculated the Pearson correlation coefficients for each pair of TLR genes to assess the relative coexpression pattern of these 10 TLR genes. The 6 TLR members that were coexpressed with the 35 genes have a high coefficient to each other, whereas they all have low coefficients to the other 4 TLR genes (TLR3, 7, 9, 10). These 6 closely regulated TLR genes were also grouped together by hierarchical clustering (Fig 5, *A*).

Among the 35 genes coexpressed with 6 TLRs were genes associated with innate immune defense such as complement components, CXC chemokine receptor 2, and low-affinity IgG receptors (Table I). The cooperation of different innate factors, including TLRs, IgG-Fc receptors, and complement receptors to regulate inflammatory responses has been documented.[43] In addition, Nearest Neighbor Analysis identified coexpression of TLRs with novel genes not previously associated with TLR signaling, and the products of these genes should give deeper insight into mechanisms of innate immunity.

DISCUSSION

We sought to identify novel leukocyte subset–associated genes by developing gene expression profiles for all major human leukocytes using microarrays. This strategy revealed >2000 leukocyte subset–specific transcripts, many of which had not previously been associated with immune responses. The discovery of leukocyte subset–restricted genes (or their protein products) began in earnest with the emergence of mAbs in the 1970s. Most of these early markers were cell surface proteins, revealed by mAbs generated by immunizing with whole cells. Gene microarray datasets and methods of analysis presented here illustrate a powerful new approach for identifying leukocyte subset–restricted gene expression, particularly for intracellular proteins such as transcription factors and signaling molecules. We also assessed genes commonly expressed by allergic leukocytes (mast cells, basophils, and eosinophils) as an exercise to identify genes involved in allergic responses. The permutations for these types of analyses are numerous.

A promising new means of intervention in human disease is targeted elimination of pathogenic cells, such as cancer cells, using apoptosis-inducing mAbs or mAbs with lethal payloads. A good example is Rituximab (IDEC

TABLE I. Genes coregulated with 6 TLRs within 250 radius

Gene title	Accession number
Adiponectin receptor 1	NM_015999
Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A	U03891
B-cell CLL/lymphoma 6 (zinc finger protein 51)	NM_001706
Bone marrow stromal cell antigen 1	NM_004334
Chromosome 1 open reading frame 24	NM_022083
Complement component 1, q subcomponent, receptor 1	W72082
Complement component 5 receptor 1 (C5a ligand)	NM_001736
Endothelin converting enzyme 1	BF969352
Family with sequence similarity 53, member C	NM_016605
F-box and leucine-rich repeat protein 5	AF142481
Fc fragment of IgG, low-affinity IIa, receptor (CD32)	NM_021642/ U90939
Fc fragment of IgG, low-affinity IIIa, receptor (CD16a)/ Fc fragment of IgG, low-affinity IIIb, receptor (CD16b)	NM_000570
Fc fragment of IgG, low-affinity IIIb, receptor (CD16b)	J04162
FLJ00012 protein	AK024423
Hypothetical protein FLJ31978	AI041543
Putative lymphocyte G0/G1 switch gene	NM_015714
CXCR2	NM_001557
KIAA0329	AB002295
Kinesin family member 13A	NM_022113
Leukocyte Ig-like receptor, subfamily B (with TM and ITIM domains), member 2/member 3	AF009635
Hypothetical protein LOC284701	AL040396
Hypothetical protein LOC284701	AW006934
Maltase-glucoamylase (α-glucosidase)	NM_004668
Hypothetical protein MGC2752	NM_023939
Mitochondrial solute carrier protein	NM_018579
Pre-B-cell colony enhancing factor 1	BF575514
Prokineticin 2	AF182069
Pleckstrin homology, Sec7 and coiled-coil domains 4	AF125349
Retinol binding protein 7, cellular	AI733027
S100 calcium binding protein A9 (calgranulin B)	NM_002965
Secreted and transmembrane 1	BF939675
Stratifin	BC000329
TNF-α–induced protein 9	NM_024636
WD repeat and FYVE domain containing 3	BE348236
Zinc finger, CCHC domain containing 6	NM_024617

CLL, Chronic lymphocytic leukemia; *TM*, transmembrane; *ITIM*, immunoreceptor tyrosine-based inhibitory motif.

Pharmaceuticals, San Diego, Calif) (anti-CD20 mAb), an effective agent for B-cell lymphomas[44] and rheumatoid arthritis.[45] Mast cells are involved in the effector phase of both allergic and autoimmune diseases, and deletion of this cell type is an appealing therapeutic option. Eosinophils have long been incriminated in the pathogenesis of asthma and, expectedly, eosinophil-deficient mice are protected against airway hyperresponsiveness[46] and airway remodeling.[47] A critical link has also been established between basophils and chronic IgE-mediated allergic inflammation.[48] A conclusion from our study was the large number of novel molecules associated with individual leukocyte subsets. The relevance of these to the function of such subsets requires studies with gene-deficient mice or similar systems.

Transcript signal intensity was generally a powerful measure of the degree of gene expression on microarrays, yet caveats apply. Multiple probe sets for the 1 gene can be represented on the GeneChips, and it was noted that one probe set could yield different expression values compared with another; 1 explanation is the representation of different splice variants of the same gene, because splice variants for numerous genes have been represented on the U133 GeneChip set. For example, there are 6 different IL-5 receptor α variants that were represented by 4 probe sets on the Affymetrix GeneChips (data not shown). Only 1 of the 6 variants of IL-5 receptor α was found to be specifically expressed by allergic leukocytes (Fig 4, *B*). This does not mean that IL-5 receptor expression is completely restricted to allergic leukocytes, because other variants were expressed in the other leukocytes.

Despite the power of large microarray datasets, cautions need to be raised. Gene transcription does not always equate with protein translation or surface expression. For instance, BCMA is 1 of the best markers for B cells and PCs at the RNA level, but is only poorly expressed at the cell surface, and then on only a small proportion of these cells.[26,33] Moreover, despite our attempts to cover all of the leukocyte subsets, it is virtually impossible to include every stage of differentiation and activation of each subset. Further, nonhematopoietic cells and some minor subsets of leukocytes, such as T regulatory cells, NK T cells, or plasmacytoid DCs, have not yet been included in our analyses. As the collection of publicly available microarray datasets increases, there will be many opportunities to refine the lists of leukocyte subset–restricted genes presented here.

In conclusion, we identified >2000 leukocyte subset–restricted transcripts/molecules, and have chosen to emphasize the potential applications of our GeneChip dataset over the possible significance of individual genes that came from our analyses. The use of a comprehensive microarray dataset for the discovery of immunologically relevant molecules, through integrated analysis of gene expression, should find further interesting applications in immunology and other disciplines.

We thank L. Weininger, T. So, and B. Woehl for GeneChip analysis, and A. Quan, C. Xin, Z. Commandeur, and J. Thatcher for technical assistance.

REFERENCES

1. Feske S, Giltnane J, Dolmetsch R, Staudt LM, Rao A. Gene regulation mediated by calcium signals in T lymphocytes. Nat Immunol 2001;2:316-24.
2. Chtanova T, Kemp RA, Sutherland AP, Ronchese F, Mackay CR. Gene microarrays reveal extensive differential gene expression in both cd4(+) and cd8(+) type 1 and type 2 t cells. J Immunol 2001;167:3057-63.

3. Chtanova T, Tangye SG, Newton R, Frank N, Hodge MR, Rolph MS, et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. J Immunol 2004;173:68-78.
4. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.
5. Staudt LM. Molecular diagnosis of the hematologic cancers. N Engl J Med 2003;348:1777-85.
6. Underhill GH, George D, Bremer EG, Kansas GS. Gene expression profiling reveals a highly specialized genetic program of plasma cells. Blood 2003;101:4013-21.
7. Sayama K, Diehn M, Matsuda K, Lunderius C, Tsai M, Tam SY, et al. Transcriptional response of human mast cells stimulated via the FcεRI and identification of mast cells as a source of IL-11. BMC Immunol 2002;3:5.
8. Nakajima T, Inagaki N, Tanaka H, Tanaka A, Yoshikawa M, Tamari M, et al. Marked increase in CC chemokine gene expression in both human and mouse mast cell transcriptomes following Fcepsilon receptor I cross-linking: an interspecies comparison. Blood 2002;100:3861-8.
9. Nakajima T, Matsumoto K, Suto H, Tanaka K, Ebisawa M, Tomita H, et al. Gene expression screening of human mast cells and eosinophils using high-density oligonucleotide probe arrays: abundant expression of major basic protein in mast cells. Blood 2001;98:1127-34.
10. Oliveira SH, Taub DD, Nagel J, Smith R, Hogaboam CM, Berlin A, et al. Stem cell factor induces eosinophil activation and degranulation: mediator release and gene array analysis. Blood 2002;100: 4291-7.
11. Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, Young RA, et al. The plasticity of dendritic cell responses to pathogens and their components. Science 2001;294:870-5.
12. Nau GJ, Richmond JF, Schlesinger A, Jennings EG, Lander ES, Young RA. Human macrophage activation programs induced by bacterial pathogens. Proc Natl Acad Sci U S A 2002;99:1503-8.
13. Tureci O, Bian H, Nestle FO, Raddrizzani L, Rosinski JA, Tassis A, et al. Cascades of transcriptional induction during dendritic cell maturation revealed by genome-wide expression analysis. FASEB J 2003;17:836-47.
14. Koopman LA, Kopcow HD, Rybalov B, Boyson JE, Orange JS, Schatz F, et al. Human decidual natural killer cells are a unique NK cell subset with immunomodulatory potential. J Exp Med 2003;198:1201-12.
15. Rolph MS, Sisavanh M, Liu SM, Mackay CR. Clues to asthma pathogenesis from microarray expression studies. Pharmacol Ther 2006;109: 284-94.
16. Shaffer AL, Rosenwald A, Hurt EM, Giltnane JM, Lam LT, Pickeral OK, et al. Signatures of the immune response. Immunity 2001;15:375-85.
17. Zimmermann N, King NE, Laporte J, Yang M, Mishra A, Pope SM, et al. Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. J Clin Invest 2003;111:1863-74.
18. Fulkerson PC, Zimmermann N, Brandt EB, Muntel EE, Doepker MP, Kavanaugh JL, et al. Negative regulation of eosinophil recruitment to the lung by the chemokine monokine induced by IFN-gamma (Mig, CXCL9). Proc Natl Acad Sci U S A 2004;101:1987-92.
19. Manger ID, Relman DA. How the host "sees" pathogens: global gene expression responses to infection. Curr Opin Immunol 2000;12:215-8.
20. Sallusto F, Lanzavecchia A. Understanding dendritic cell and T-lymphocyte traffic through the analysis of chemokine receptor expression. Immunol Rev 2000;177:134-40.
21. Ohl L, Henning G, Krautwald S, Lipp M, Hardtke S, Bernhardt G, et al. Cooperating mechanisms of CXCR5 and CCR7 in development and organization of secondary lymphoid organs. J Exp Med 2003;197:1199-204.
22. Grunig G, Warnock M, Wakil AE, Venkayya R, Brombacher F, Rennick DM, et al. Requirement for IL-13 independently of IL-4 in experimental asthma. Science 1998;282:2261-3.
23. Nagata K, Hirai H, Tanaka K, Ogawa K, Aso T, Sugamura K, et al. CRTH2, an orphan receptor of T-helper-2-cells, is expressed on basophils and eosinophils and responds to mast cell-derived factor(s). FEBS Lett 1999;459: 195-9.
24. Sallusto F, Mackay CR, Lanzavecchia A. Selective expression of the eotaxin receptor CCR3 by human T helper 2 cells. Science 1997;277: 2005-7.
25. Tangye SG, Avery DT, Deenick EK, Hodgkin PD. Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. J Immunol 2003;170:686-94.
26. Avery DT, Kalled SL, Ellyard JI, Ambrose C, Bixler SA, Thien M, et al. BAFF selectively enhances the survival of plasmablasts generated from human memory B cells. J Clin Invest 2003;112:286-97.
27. Ellyard JI, Avery DT, Phan TG, Hare NJ, Hodgkin PD, Tangye SG. Antigen-selected, immunoglobulin-secreting cells persist in human spleen and bone marrow. Blood 2004;103:3805-12.
28. Aida Y, Pabst MJ. Priming of neutrophils by lipopolysaccharide for enhanced release of superoxide: requirement for plasma but not for tumor necrosis factor-alpha. J Immunol 1990;145:3017-25.
29. Hansel TT, Pound JD, Pilling D, Kitas GD, Salmon M, Gentle TA, et al. Purification of human blood eosinophils by negative selection using immunomagnetic beads. J Immunol Methods 1989;122:97-103.
30. Ochi H, Hirani WM, Yuan Q, Friend DS, Austen KF, Boyce JA. T helper cell type 2 cytokine-mediated comitogenic responses and CCR3 expression during differentiation of human mast cells in vitro. J Exp Med 1999; 190:267-80.
31. Jeffrey KL, Brummer T, Rolph MS, Liu SM, Callejas NA, Grumont RJ, et al. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. Nat Immunol 2006;7:274-83.
32. Chtanova T, Newton R, Liu SM, Weininger L, Young TR, Silva DG, et al. Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. J Immunol 2005;175:7837-47.
33. Ng LG, Sutherland AP, Newton R, Qian F, Cachero TG, Scott ML, et al. B cell-activating factor belonging to the TNF family (BAFF)-R is the principal BAFF receptor facilitating BAFF costimulation of circulating T and B cells. J Immunol 2004;173:807-17.
34. Spik I, Brenuchon C, Angeli V, Staumont D, Fleury S, Capron M, et al. Activation of the prostaglandin D2 receptor DP2/CRTH2 increases allergic inflammation in mouse. J Immunol 2005;174:3703-8.
35. Sehmi R, Wood LJ, Watson R, Foley R, Hamid Q, O'Byrne PM, et al. Allergen-induced increases in IL-5 receptor alpha-subunit expression on bone marrow-derived CD34+ cells from asthmatic subjects: a novel marker of progenitor cell commitment towards eosinophilic differentiation. J Clin Invest 1997;100:2466-75.
36. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004;101:6062-7.
37. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun 2005;6:319-31.
38. Lanzavecchia A. Antigen-specific interaction between T and B cells. Nature 1985;314:537-9.
39. Sallusto F, Mackay CR. Chemoattractants and their receptors in homeostasis and inflammation. Curr Opin Immunol 2004;16:724-31.
40. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997;278:680-6.
41. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc Natl Acad Sci U S A 2003;100:605-10.
42. Lamb J, Ramaswamy S, Ford HL, Contreras B, Martinez RV, Kittrell FS, et al. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. Cell 2003;114:323-34.
43. Benoist C, Mathis D. Mast cells in autoimmune disease. Nature 2002; 420:875-8.
44. Reff ME, Carner K, Chambers KS, Chinn PC, Leonard JE, Raab R, et al. Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. Blood 1994;83:435-45.
45. Edwards JC, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, Close DR, et al. Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N Engl J Med 2004;350: 2572-81.
46. Foster PS, Hogan SP, Ramsay AJ, Matthaei KI, Young IG. Interleukin 5 deficiency abolishes eosinophilia, airways hyperreactivity, and lung damage in a mouse asthma model. J Exp Med 1996;183:195-201.
47. Humbles AA, Lloyd CM, McMillan SJ, Friend DS, Xanthou G, McKenna EE, et al. A critical role for eosinophils in allergic airways remodeling. Science 2004;305:1776-9.
48. Mukai K, Matsuoka K, Taya C, Suzuki H, Yokozeki H, Nishioka K, et al. Basophils play a critical role in the development of IgE-mediated chronic allergic inflammation independently of T cells and mast cells. Immunity 2005;23:191-202.