

FULL PAPER

Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data

AR Abbas¹, D Baldwin¹, Y Ma¹, W Ouyang², A Gurney^{2,3}, F Martin², S Fong², M van Lookeren Campagne², P Godowski², PM Williams³, AC Chan² and HF Clark^{1,2}

¹Department of Bioinformatics, Genentech, Inc., South San Francisco, CA, USA; ²Department of Immunology, Genentech, Inc., South San Francisco, CA, USA; ³Department of Molecular Biology, Genentech, Inc., South San Francisco, CA, USA

Immune cell-specific expression is one indication of the importance of a gene's role in the immune response. We have compiled a compendium of microarray expression data for virtually all human genes from six key immune cell types and their activated and differentiated states. Immune Response In Silico (IRIS) is a collection of genes that have been selected for specific expression in immune cells. The expression pattern of IRIS genes recapitulates the phylogeny of immune cells in terms of the lineages of their differentiation. Gene Ontology assignments for IRIS genes reveal significant involvement in inflammation and immunity. Genes encoding CD antigens, cytokines, integrins and many other gene families playing key roles in the immune response are highly represented. IRIS also includes proteins of unknown function and expressed sequence tags that may not represent genes. The predicted cellular localization of IRIS proteins is evenly distributed between cell surface and intracellular compartments, indicating that immune specificity is important at many points in the signaling pathways of the immune response. IRIS provides a resource for further investigation into the function of the immune system and immune diseases.

Genes and Immunity (2005) 6, 319–331. doi:10.1038/sj.gene.6364173

Published online 24 March 2005

Keywords: *microarray expression; specific expression; immune response*

Introduction

The immune system in higher eukaryotes has evolved into a complex network involving many specialized cell types with multiple points of regulation. This enables the immune response to attack foreign invaders while recognizing and tolerating self-antigens. Two distinct lineages of immune cells have evolved with cell types that specialize in each of the many roles that are required for this immune response. The myeloid lineage of monocytes, macrophages, dendritic cells and neutrophils carries out the innate immune response, recognizing microbial pathogens typically by carbohydrates found only in bacterial proteins. The lymphoid lineage of T cells, B cells and natural killer (NK) cells enables adaptive immunity by distinguishing self from non-self antigens and also providing a memory of foreign proteins seen before. Other immune cells, such as eosinophils, basophils and mast cells, also participate in the immune response. Although a large repertoire of genes that play key roles in the differentiation, function and regulation of these immune cells is already well-

described, this has not yet led to a complete understanding of immune diseases.

Genome-wide microarray expression profiling of immune cells provides opportunities for identifying other genes that may function in the immune response. Microarray experiments have been carried out on various immune cell subsets by a number of different investigators in order to understand gene expression differences during differentiation and activation.^{1–17} These studies have provided invaluable insight into comprehensive gene expression profiles that define many different immune cell subsets and states of differentiation and activation. In particular, a recent study detailed elegantly the genes expressed specifically in many subsets of the T-cell lineage.¹⁸ Comparison of gene expression profiles across a broad range of immune cell types and nonimmune tissues is necessary, however, to fully appreciate which gene expression differences are unique to each immune cell subset, which are found in multiple subsets or across immune cell lineages, and which are found more widely across the many cell types in the human body and thus may be involved in more general cellular processes. Differences among microarray platforms and experimental protocols used by different investigators often confound such comparison. Here, a compendium of microarray expression data from a broad representation of isolated immune cell subsets has been

Correspondence: Dr HF Clark, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA. E-mail: hclark@gene.com

Received 29 July 2004; revised 1 November 2004; accepted 8 December 2004; published online 24 March 2005

generated on the same microarray platform and analysis of gene expression profile across the major cell types of the immune system is made possible. Moreover, gene expression in all other major tissues allows determination of immune-specific expression. Traditional methods of determining expression are performed on a gene-by-gene basis, and single microarray experiments show differential expression between just a few immune cell subsets. Here, genes fitting a complex expression criterion are identified on a genomic scale.

Immune cell-specific expression is one indication of the importance of a gene's role in the function of the immune system. 'Cluster of differentiation' (CD) antigens are used experimentally as specific markers for immune cell subsets and they often play a key role in the function of that cell. For example, the T-cell receptor is expressed only on T lymphocytes and is responsible for self-antigen recognition, a primary function of this immune cell.¹⁹ Here, immune cell specifically expressed genes are identified by a survey of expression profiles across all the immune cells and major non-immune tissues by a method termed *Immune Response In Silico* (IRIS). Immune cell specificity is determined generally by higher expression in any immune cell than expression in any nonimmune cell tissue. Finer characterization of specificity within an immune cell type, such as T cells, and an immune cell lineage, such as lymphoid cells, is also determined. Furthermore, expression profiles within subsets of an immune cell, such as classes of T cells expressing CD4 or CD8 antigens, memory and helper T cells and resting *vs* activated cells, further refine the specificity of immune-specific genes. IRIS has identified both well-characterized immune genes and a number of highly immune-specific genes with unknown function.

Results

IRIS identifies genes more highly expressed in immune cells than in any of the major organs of the body. Gene expression is surveyed across a compendium of samples, including immune cell subsets (Table 1) and a comprehensive range of normal tissues. The immune cells have been isolated from normal human blood by purifying each cell type via its specific cell surface markers. Activated and differentiated subsets were developed by *in vitro* stimulations. RNA samples were labeled and run on microarray chips that include probesets for virtually every human gene. Genes are often represented by more than one probeset and the expression levels vary, although the expression profiles are usually consistent. Genes are selected for IRIS based on cutoff values for expression levels in immune *vs* nonimmune samples, as shown in Supplementary Figure 1. Prior to the determination of these cutoffs, genomewide expression levels were evaluated by surveying the range of expression of all genes and the expression levels of a few families of genes already known to play important roles in the immune system. Gene expression profiles for each IRIS gene can be viewed from our website (<http://share.gene.com/share/clark.iris.2004/iris/iris.html>).

Expression consistent with immune cell lineages

The expression pattern of IRIS genes recapitulates the phylogeny of immune cells, as shown in Figure 1. The

highest mean expression level of any immune cell subset within each cell type as defined in Table 1 is calculated for each IRIS gene. Hierarchical clustering, a statistical method for grouping genes by the similarity of their expression profiles, reveals patterns of gene expression that are distinct among immune cell lineages. The dendrogram of the relationships of these genes mimics the evolutionary relationships of the lineages of immune cell types. The heat map showing all IRIS genes also illustrates the complexity of expression profiles, with most genes expressed at some level in more than one immune cell type.

IRIS categories

While clustering is a useful method for identifying patterns of gene expression, it fails to clearly define the parameters of these patterns. Therefore, cutoff values of gene expression signatures in various immune cell types were approximated and used to define the IRIS categories that were observed in the clustering (Table 1) and these parameters are defined in the *Cell & Lineage Assignment* section of Materials and methods. The probesets are categorized by their profiles according to the degree of specificity within immune cells. Profiles specific to one cell type are assigned to the categories T cell, NK cell, B cell, monocyte, dendritic or neutrophil. Profiles specific to more than one cell type within a lineage are assigned to lymphoid or myeloid categories. Profiles specific to cell types across lineages, as well as probesets with expression levels equivalent in all immune cells, are assigned to the multiple category. Genes represented by several probesets may appear in more than one category. A gene is best described as specific to the most exclusive category to which it has been assigned, that is, the highest expressing probeset may be in the T-cell category and more weakly expressing probesets for that gene may fall into the lymphoid or multiple categories because of the stringency of cutoff levels used.

Patterns of expression profiles within each lineage

Cell types within a lineage share some immune-specific genes, suggesting that they confer common functions among those cells. The expression profiles are very diverse and complex, but some general patterns emerge. As described above, the IRIS categories attempt to group genes within a cell type, or, if specific to more than one cell type, within a lineage. Here, these categories are assessed for general patterns of expression profiles. K-means clustering reveals groups of expression profiles within an IRIS category, as shown in Figure 2. Within the lymphoid category clustered in Figure 2a, T cells and NK cells are shown to share a number of immune-specific genes (clusters 1, 4 and 6), whereas B cells express more distinct genes (clusters 3 and 5). Often, the genes shared between T and NK cells are specifically expressed in the activated state of both cells (cluster 6). The myeloid category shown in Figure 2b reveals a similar relationship between genes transiently expressed in differentiating macrophages and also in LPS-induced dendritic cells (cluster 8). Profile clusters of genes highly specific to single-cell types are shown in Supplementary Figure 2. T cells have several general patterns of specificity among subsets of T cells, as shown in Supplementary Figure 2a. CD8 cells have a number of unique genes (cluster 1), and

Table 1 IRIS categories, the immune cell subset expression used to determine them and the number of genes in each category

<i>IRIS category</i>	<i>Immune cell subsets</i>	<i>Replicates</i>	<i>Genes</i>
Lymphoid			241
T cell	CD8 T cell (naïve and resting)	4	77
	CD4 T cell (naïve and resting)	3	
	Helper Th1—12 h (of differentiation)	2	
	Helper Th1—48 h (of differentiation)	2	
	Helper Th2—12 h (of differentiation)	3	
	Helper Th2—48 h (of differentiation)	3	
	Memory T—resting (naïve)	3	
	Memory T—activated	3	
NK cell	NK cell (resting)	2	16
	NK—IL2 (activated)	5	
	NK—IL15 (activated)	6	
B cell	B cell (naïve)	7	93
	Memory B—IgG/IgA	4	
	Memory B—IgM	4	
	Plasma B—PBMC	3	
	Plasma B—bone marrow	4	
Myeloid			343
Monocyte	Monocyte	12	84
	Macrophage—1 day (of differentiation)	12	
	Macrophage—7 days (of differentiation)	12	
Dendritic cell	Dendritic cell (resting)	6	74
	Dendritic—LPS (activated)	6	
Neutrophil	Neutrophil	5	45
Multiple			821
Total IRIS			1688

Immune cell subsets were isolated and either differentiated or activated as described in Materials and Methods. Replicates indicates the number of samples for each cell subset run on the microarray; the mean of their intensity values is used in the selection of IRIS, the standard deviations are shown in Figure 3, Supplementary Figure 3 and Supplementary Table 1. The immune cell subset labels are used in subsequent figures and tables, and further descriptors are in parentheses. A gene may be counted in multiple categories, but the total number of IRIS genes is non-redundant.

also share some genes with CD4 cells (cluster 2). T-helper cells share some specific genes (clusters 3 and 6) and a number of genes appear only in activated memory T cells (cluster 5). B cells also show distinct profiles, as shown in Supplementary Figure 2b. Some genes are expressed specifically in both naïve and memory B cells (cluster 2), others only in plasma cells (clusters 3 and 4), while there are no specific B-cell genes expressed distinctly in naïve or memory cells alone. Likewise, monocytes and macrophages shown in Supplementary Figure 2c reveal a class of genes specific to monocytes (cluster 1) and another transiently expressed in differentiating macrophages (cluster 2), but none specific to fully differentiated macrophages. However, Figure 2b shows that fully differentiated macrophages share profiles with neutrophils (cluster 2) and dendritic cells (cluster 5). All of these observations provide insights into the differentiation of immune cells as they have evolved to play specific roles in the immune response.

Statistical significance of IRIS

The statistical significance of specificity is assessed for the genes within each IRIS cell type or lineage, as shown in Supplementary Figure 4. The *F* statistic measures both

immune cell variation and sample variation simultaneously, so a low *F* value can either indicate similar expression of a gene across immune cells, or gene expression variability among samples, but does not distinguish between the two. Conversely, a gene highly specific to one cell type has a high *F* value, particularly if the replicate samples all have similar expression of that gene. Sample variability occurs because of both incomplete reproducibility of microarray conditions and biological variation between blood donors due to unknown causes. The multiple category has the highest proportion of genes with low *F* values, suggesting that this category includes most of the genes falsely assigned as immune-specific. While the microarray technology used here is a well-established experimental method for detecting differential gene expression, independent confirmation by real-time quantitative PCR (RT-QPCR) is carried out before pursuing further functional studies on genes of interest.²⁰

Gene ontology and gene families

Genes of known function are assigned a term from Gene Ontology, a structured vocabulary for describing biological process, molecular function and cellular

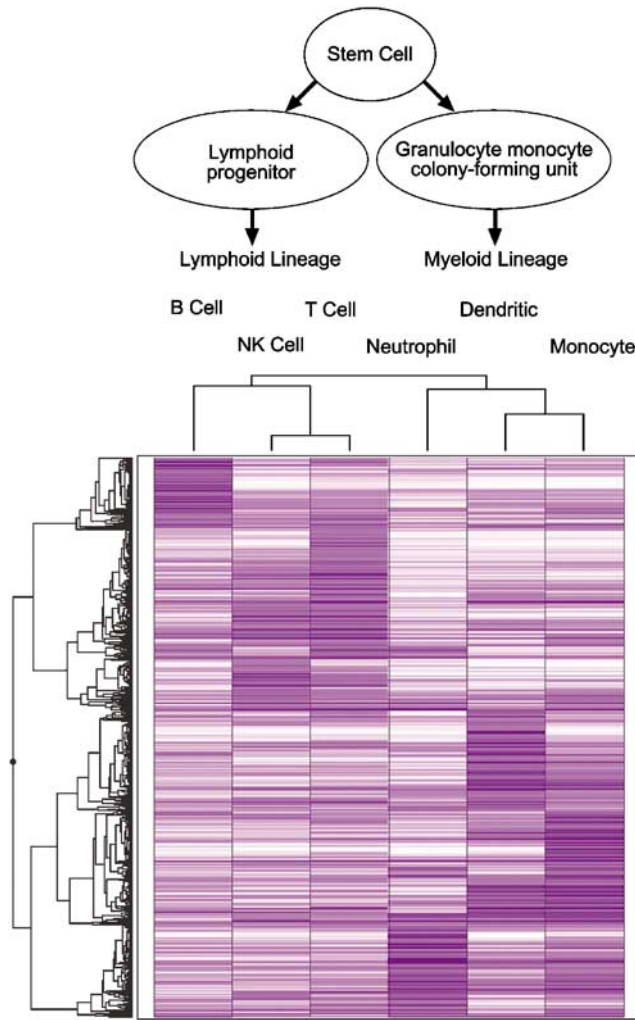


Figure 1 IRIS genes are clustered based on expression for each immune cell type and shown with a schematic of their lineages. The highest mean expression of any cell subset for each probeset represents the expression of that cell type. The dendrograms shows the similarity of expression profiles for genes in the rows and cell types in the columns. Intensity values less than 1000 have the lightest shading and intensity values greater than 10,000 have the darkest shading.

component.²¹ Table 2 shows the Biological Processes most highly specific to IRIS, meaning that these ontologies are seen in IRIS in higher numbers than in a random set of genes. Many specific responses to different foreign invaders, chemotaxis of immune cells to sites of inflammation and other aspects of the immune response are represented. In summary, 57% of the well-characterized IRIS genes have these immune functions. Similar results are seen with the molecular function ontologies (data not shown), with the most highly represented being antigen binding and the activities of chemokines, cytokines and their receptors. Table 3 shows the gene families represented in IRIS that are known to have many members with key roles in the differentiation, function and regulation of the immune system, including those with the molecular functions mentioned. The Protcomp algorithm (Softberry, Inc.) predicts for the 1589 IRIS genes with ORFs that 24% of the encoded proteins are in

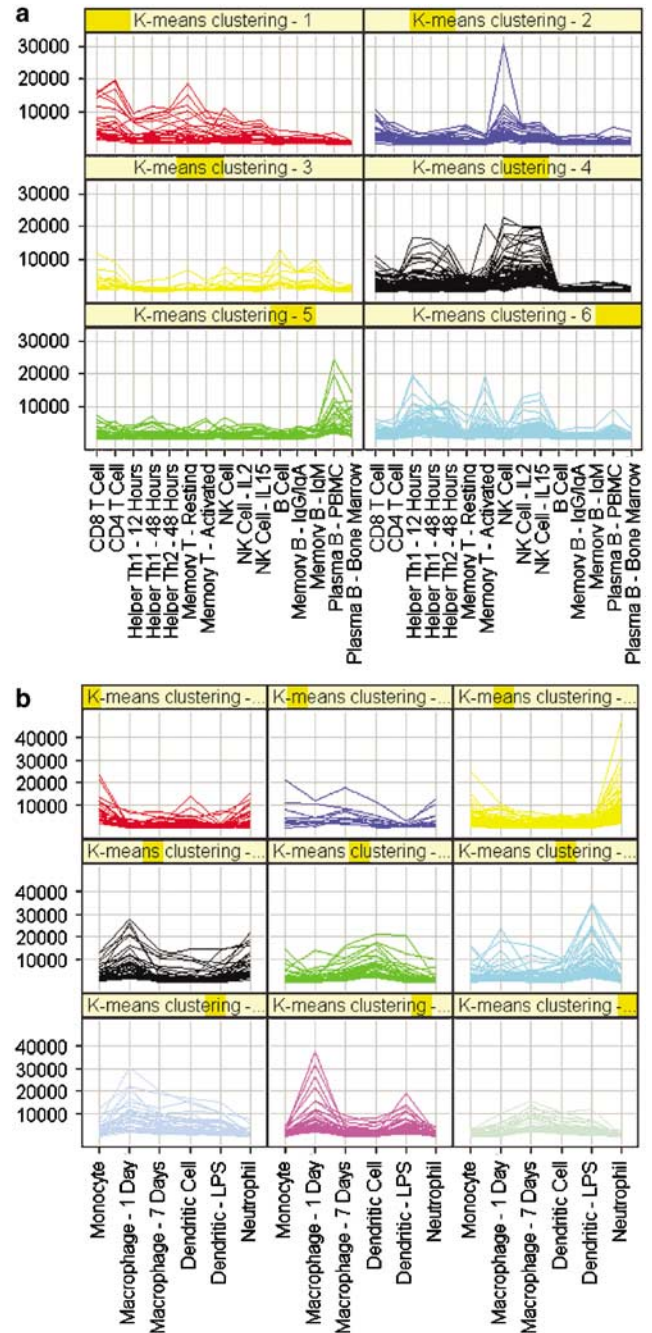


Figure 2 K-means clustering of the mean expression level in each cell subset of the genes in the Lymphoid (a) and Myeloid (b) category is shown. The y-axis is intensity level.

the plasma membrane, 13% are secreted, 24% are nuclear and the remaining 39% are in other intracellular compartments. These localization assessments were made for comparison to previously reported observations, but the conclusions were deemed insignificant.²²

Highly specific genes with known function

Genes with highly specific expression in each immune cell type are likely to confer that cell's distinct properties. An expression differential of 10-fold between the highest nonimmune tissue level and highest immune cell level

Table 2 Gene Ontology (GO) biological process assignments with highest degree of statistical significance measured by the P -value of the χ^2 test

<i>P</i> -value	IRIS genes	All genes	GO ID	GO term
1.35E-101	216	582	9607	Response to biotic stimulus
1.84E-101	204	530	6952	Defense response
2.84E-93	187	483	6955	Immune response
3.73E-72	141	355	9613	Response to pest/pathogen/parasite
9.73E-64	231	833	9605	Response to external stimulus
1.18E-55	89	196	9611	Response to wounding
2.07E-49	242	1004	50896	Response to stimulus
1.76E-38	154	571	6950	Response to stress
5.54E-36	56	121	6954	Inflammatory response
5.54E-36	56	121	45087	Innate immune response
2.36E-34	41	74	6968	Cellular defense response
1.54E-25	43	98	19730	Antimicrobial humoral response
1.16E-24	207	1043	50874	Organismal physiological process
3.68E-24	53	142	6959	Humoral immune response
2.50E-20	36	85	6935	Chemotaxis
2.50E-20	36	85	42330	Taxis
1.76E-13	36	109	42221	Response to chemical substance
9.29E-10	242	1625	7165	Signal transduction

Ontologies may appear multiple times in a hierarchical tree, so a gene may be assigned to multiple GO terms. 367 IRIS genes are assigned to these 18 ontologies 2270 times. 644 IRIS genes have GO Biological Process assignments.

Table 3 Families of genes encoding proteins with key immune functions that are highly represented in IRIS are listed with the number of IRIS members

Gene Family	Number of genes
CD 'cluster of differentiation' antigens	171
Immunoglobulin domain-containing proteins	157
Immunoglobulin and transmembrane-containing proteins	136
SH2/SH3 domain proteins	66
Immunotyrosine activation/inhibition proteins	53
G-protein-coupled receptors	45
Interleukins and receptors	43
Chemokine ligands and receptors	40
C-type lectins	36
Interferon, receptors, regulatory and induced proteins	28
Tumor necrosis factors, receptors and induced proteins	27
Solute carriers	26
Fc receptors	18
Killer cell immunoglobulin-like receptors	15
Killer cell lectin-like receptors	14
Leukocyte immunoglobulin-like receptors	13
Caspases	13
Signaling lymphocytic activation molecules	12
Translocation-related lymphoma genes	11
Disintegrins and metalloproteinases	11
Integrins	9
Toll-like receptors	8
Colony stimulating factors and receptors	8
Major histocompatibility complex	6

selects 122 genes from the single-cell-type categories. The expression profiles of the highest expressing of these genes are shown in Figure 3 and the rest are shown in Supplementary Figure 3. Reported literature on many of these genes confirms that specific expression of transcripts often results in specific protein expression and that they have key roles in the immune response. For each cell type, a few genes of interest are described below.

T cells

Genes with specificity in T cells have expression profiles in a wide variety of patterns as shown in Figure 3a and Supplementary Figure 3a. The T-cell receptor alpha locus, TCRA, as well as CD3 subunits D, E & G are expressed on most T-cell subsets. CD8A and B1 are heterodimers of the established protein marker for CD8+ T cells which stabilizes interactions between the T-cell receptor and an MHC class I-peptide complex,²³ and are expressed only in CD8 cells. LEF1 is expressed specifically in both CD8 and CD4 cells. LEF1 is a transcription factor that has been implicated as critical in a differentiation defect that leads to T-cell lymphomas.²⁴ A number of genes exhibit exquisite selectivity for memory *vs* naive T cells and most of these show enhanced expression in Th1 or Th2 subsets. Cytokine IL9 is specific to both Th2 and activated memory cells, whereas CTLA4 and ICOS are highly expressed in activated memory cells but also seen at lower levels in other T-cell subsets and are both cell-surface regulatory receptors.²⁵ GZMK is specific to CD8, Th2 and resting memory cells and is a serine protease that is proposed to prevent damage of bystander cells at sites of inflammation.²⁶ LAG3 is found primarily on Th1 cells, but also on stimulated CD8+ T cells (data not shown), and it stimulates the maturation of dendritic cells.²⁷

NK cells

The highest expressing specific gene in NK cells is KLRF1, a C-type lectin with immunoreceptor tyrosine-based inhibitory motifs (ITIM) motifs that stimulates cytotoxicity²⁸ as shown in Figure 3b. Its expression is highest in resting cells, and decreases upon activation. The KIR gene cluster is also represented, although the probesets probably cross-hybridize with these closely related genes that have probably arisen from gene duplication. Some members are established protein markers for this cell type and all are known to play a

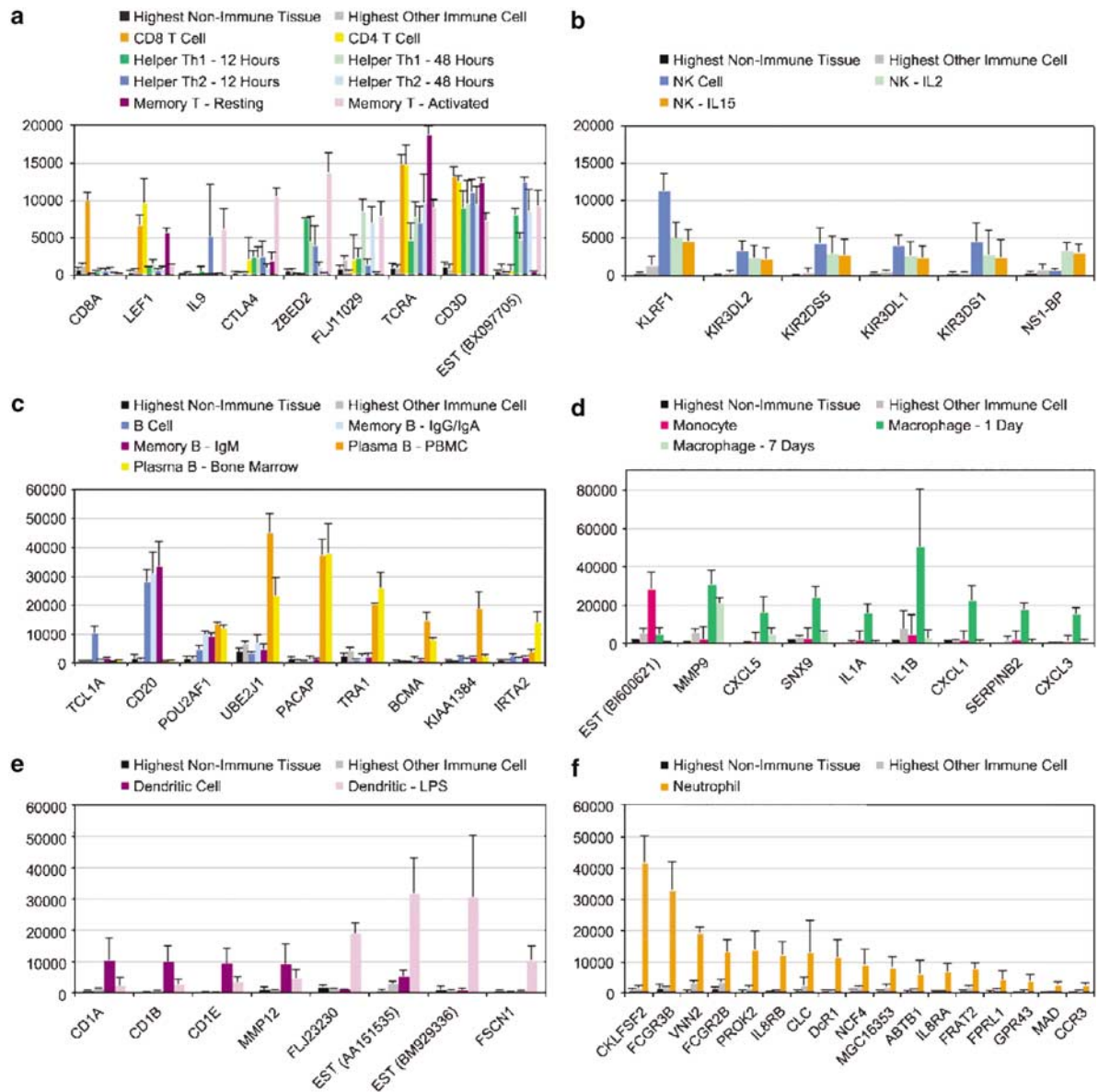


Figure 3 Highly specific genes for each IRIS category are shown with their expression profiles in immune cell subsets for that category as compared to the level in non-immune tissues and in other immune cells: a) T cell, b) NK cell, c) B cell, d) monocyte, e) dendritic cell, and f) neutrophil. Highly specific genes with lower expression levels are shown in Supplementary Figure 3.

key role in recognition of HLA class I ligands.²⁹ NS1-BP is specifically expressed in activated NK cells and binds the influenza A virus NS1 protein.³⁰

B cells

B cells are represented by a diversity of specific expression profiles. TCL1A is expressed solely in naïve B cells and is an intracellular enzyme that regulates an early stage of B- and T-cell differentiation,^{31,32} as shown with the other B-cell genes in Figure 3c and Supplementary Figure 3b. CD20 and a number of other genes have equivalent expression in both naïve B cells and memory cells expressing either IgG/IgA or IgM. An antibody therapeutic targeting CD20, Rituxan, is extremely effective in specific ablation of B cells.³³ CD19 is an established protein marker for B cells. BAFFR is the

principal receptor for the signaling pathway that maintains mature B-cell survival.³⁴ BANK1 protein has also been shown to be specific to B cells and may play a role in foreign antigen-induced immune response.³⁵ PAX5 is a transcription factor that plays an essential role in commitment to the B cell lineage.³⁶ CD79A and B are two components of the B-cell antigen receptor³⁷ and are expressed at lower levels in plasma cells. FCRH1, FCRH2 and IRTA2 are closely related in sequence to the Fc receptor which functions in antibody binding and regulation of the immune response,³⁸ but they have not been functionally characterized.^{39–42} FCRH1 and FCRH2 are found primarily on naïve and memory B cells, while IRTA2 is specific to plasma cells from bone marrow. A number of other genes are specific to plasma cells from blood, from bone marrow or from both sources. BCMA is

a member of the TNF receptor family that is hypothesized to transduce signals for B-cell survival and proliferation.⁴³

Monocytes and macrophages

Several distinct patterns of specific monocyte and macrophage expression are seen in Figure 3d and Supplementary Figure 3c. Expressed only in monocytes is PRAM1, an adaptor protein that appears to be involved in the differentiation of monocytes.⁴⁴ The second and most common pattern is transient expression in monocytes differentiated after 1 day, and lost after 7 days. Chemokines CCL24, CXCL1 and CXCL3, and chemokine receptor CCRL2,⁴⁵ are expressed in the same pattern. GPR84 is a G-protein-coupled receptor, possibly a chemokine receptor, with protein expression detected on neutrophils and eosinophils,⁴⁶ with this transient macrophage expression not yet reported. IL1B has the highest level of mean expression in this category, although there is a wide range of expression levels in individual blood donors. IL1B protein is also expressed by other cell types under certain conditions.^{47,48} IL1A, the other chain in the IL1 heterodimer, is also expressed in this pattern, although it has been shown that IL1A and IL1B are regulated independently.⁴⁹ Three other cytokines, IL1F9, IL19 and IL24, are expressed in a pattern similar to that of IL1. IL1F9 appears to play a modulatory role in IL1 receptor signaling.⁵⁰ IL19 and IL24 proteins have been shown to be preferentially expressed on monocytes.⁵¹ MMP9 has the highest expression in both day 1 and fully differentiated macrophages and is a matrix metalloprotease well studied in tumor immunosuppression.⁵²

Dendritic cells

CD1A, CD1B and CD1E are specific to resting dendritic cells and are established markers for this cell type,⁵³ while CD1D is also expressed in B cells⁵⁴ and is not found to be specific only to dendritic cells here (Figure 3e and Supplementary Figure 3d). CLECSF13 is a lectin that was identified in macrophages,⁵⁵ although here it is seen to have significantly higher expression in dendritic cells. ALOX15 protein expression can be induced in monocytes,⁵⁶ but this study shows dendritic cell specificity. Specific here to both resting and induced dendritic cells are two chemokines, CCL13 and CCL17, that have been shown to be expressed in different cell types under different conditions.^{57,58} FSCN1 is also specifically induced with LPS here and is an actin-bundling cytoplasmic protein that has been shown to be expressed in dendritic cells, as well as other cell types that develop membrane protrusions.⁵⁹ PDL2 protein has been shown to be expressed on dendritic cells and to play a role in inhibiting T-cell activation.⁶⁰

Neutrophils

Many genes are seen to be specific to neutrophils (Figure 3f). FCGR3 is an established protein marker for neutrophils.⁶¹ NCF4 mediates intracellular killing following ingestion of microorganisms, one of the main functions of neutrophils.⁶² VNN2 mediates neutrophil adherence migration.⁶³ IL8RA and IL8RB, two receptors of the cytokine IL8, are seen here to be expressed specifically in neutrophils, but the protein has also been found to be induced on a number of other immune cell

types.⁶⁴ DcR1 is a receptor antagonist of TRAIL/Apo-2L and has been identified in a number of other cell types^{65,66} but here the neutrophil expression is most significant.

Discussion

Immune-specific expression is an indication of the role of a gene in a function specific to the immune system. These functions include cell-type-specific differentiation, chemotaxis and migration of infiltrating immune cells, cell surface receptors that identify a specific cell, signaling molecules that communicate between the cell types through these receptors, immune functions such as antigen recognition and self-recognition to maintain tolerance, modulation of the immune response, and apoptosis. IRIS includes a significant representation of genes known to have key roles in these functions. Furthermore, surveying complex gene expression profiles across many immune cell subsets provides an insight into the distinct role played by each different immune cell type. In fact, overall patterns of immune specific expression recapitulate the lineages of immune cell type differentiation. Commonalities and differences between different immune cells within a lineage are revealed by the expression profiles of IRIS genes.

Intriguingly, many IRIS genes have no known function. Some are represented only by an expressed sequence tag (EST), with no evidence to date that they are transcribed from a protein-encoding gene. This phenomenon has also been described recently for noncoding RNAs regulated in response to retinoic acid.⁶⁷ Highly specific genes (Figure 3) that have no known function are highlighted here. Figure 3a shows ZBED2, FLJ11029 and an EST (BX097705) to be highly expressed in activated memory cells but also seen at lower levels in other T-cell subsets. In Supplementary Figure 3a, HGFL is seen to be specifically expressed in both CD8 and CD4T cells and appears to be a membrane protein with proteolytic activity implied by its predicted Kringle domain (PFAM PF00051). B-cell-specific gene NALP7 (Supplementary Figure 3b) is a member of a family of pyrin domain proteins with some members implicated in inflammatory signalling.⁶⁸ LOC220213 is in only bone marrow plasma cells and FLJ23235 is also expressed in plasma cells from blood. KIAA1384 is specifically and highly expressed only in blood plasma cells (Figure 3c). The highest specific expression in only monocytes is seen for an EST (BI600621), as shown in Figure 3d. RPEL1, ZFYVE16 and C21orf42 have transient expression in monocytes differentiated after one day (Supplementary Figure 3c). HOZFP appears to be a transcription factor, but has no known function and its dendritic cell specific expression (Supplementary Figure 3d) has not been previously described.⁶⁹ Most highly and specifically induced with LPS are RhoGAP2, FLJ23230 and two ESTs (Figure 3e). CKLFSF2, the most highly and specifically expressed gene in neutrophils (Figure 3f), appears to be a tetraspanning cell membrane molecule with a MARVEL domain (PFAM PF01284) that is implicated in lipid recognition.⁷⁰ MGC16353 also has no known function.

Immune-specific expression is only one criterion by which to identify immune genes. Table 2 shows that while many IRIS genes have roles in the immune

response, many more are not included. Cutoffs used to select IRIS genes and other technical reasons account for this in some cases. However, genes also can be expressed outside of immune cells and still have a function in the immune response.

Immune-specific genes are highly variable in expression both within and between different immune cell types. This demonstrates the specialization of immune cells as they have evolved from the innate immunity found in all eukaryotes to the adaptive immunity found only in vertebrates. Similarities between different immune cell subsets within the myeloid and lymphoid lineages are revealed by patterns of shared expression profiles, such as activated T and NK cells which share cytolytic functions, and day 1 macrophages and LPS-induced dendritic cells that both differentiate from monocytes. IRIS genes are predicted to encode proteins similarly represented in all cellular compartments, with equal proportions on the cell surface and in the nucleus, as well as a significant proportion of secreted signaling molecules and a large proportion of intracellular proteins.

IRIS provides a resource for further investigation into immune cell differentiation, regulation and disease. The data provided here demonstrate that immune specificity of a gene may indicate immune function. Many IRIS genes have no known function and the expression profiles provided here suggest roles for the encoded proteins that can be investigated experimentally. Early in the development of IRIS, an EST was noticed because of its specific expression in B and T cells. Identification of neighboring ESTs and predicted exons in the human and syntenic mouse genome sequences enabled the construction of composite orthologous sequences encoding an Ig domain and intracellular ITIM and ITSM domains. PCR was successfully used to isolate a cDNA for each species that is now being pursued in functional studies. Soon after, a manuscript reporting a mouse knockout of the same gene with an immune response phenotype was submitted by another group that had discovered the gene independently.⁷¹

IRIS genes can also be used as cell markers in the analysis of microarray experiments. For example, tumor *vs* normal tissue analysis may benefit from recognition of genes involved in the immune response as opposed to carcinogenesis. We have also run microarray experiments with various immune disease biopsies and blood samples with the appropriate normal controls and then viewed the expression of genes in IRIS categories to provide a qualitative assessment of the preponderance of immune cell types, or change in the expression of immune-specific genes, in the diseased tissue *vs* the unaffected tissue (data not shown). Immune cell infiltration into tissue can clearly be identified in this way. An example of analysis of blood from systemic lupus erythematosus and rheumatoid arthritis patients is shown in Supplementary Figure 5. The genes most highly expressed in all the samples are in the myeloid lineage, with some of these genes much more highly expressed in disease. A few myeloid genes are dramatically up regulated only in some patients with either disease. Included are IFIT1, MX1 and cig5, which are considered to show the interferon-inducible gene expression signature in patients with severe lupus.⁷²⁻⁷⁴ While this kind of IRIS analysis provides a qualitative

view of cellular composition, it is not possible in this way to quantitatively predict the relative proportions of immune cell types in diseased tissues. Therefore, further development and validation of an algorithm to accomplish this challenging goal is underway.

IRIS is also being used to highlight candidate genes in genetic studies of immune disease where genetic linkage mapping has identified a locus that includes tens or hundreds of genes. Orthology mapping is used to cross-reference IRIS genes to their mouse counterpart to aid in both microarray analysis of mouse tissues from disease models and candidate identification in mouse genetic screens. Not surprisingly, we have observed that IRIS genes have a lower rate of orthologous pairing between human and mouse than do other genes (data not shown), apparently due to well-known and possibly unknown differences in the immune systems of the two species.

In conclusion, IRIS provides focus on genes known to conduct the immune response, insight into genes of unknown function that may play an immunological role and also provides a resource for future discovery into both the normal and diseased immune paradigms.

Materials and methods

Microarray sample preparation

Immune cells were isolated from leukopacks from normal human donors with RosetteSep™ kits (StemCell Technologies) and MACS[®] MicroBeads (Miltenyi Biotec), with a few exceptions, and cell subsets isolated as described below. RNA was extracted using Qiagen RNeasy mini-column with DNase treatment and prepared for the microarray using standard Affymetrix™ GeneChip™ protocol. Table 1 lists the cell subsets and the number of replicates for each subset used in the analysis of the expression data. The cell isolations and *in vitro* stimulations used to achieve each subset are described in the protocols following.

T-cell subset isolation

CD8 cells were isolated with RosetteSep™ CD8 + T-cell enrichment cocktail and cells were collected for the 'CD8 T-cell' subset. CD45RA + CD8 + cells were further purified by CD45RO MicroBeads through negative selection. Cell purity of greater than 90% was confirmed by FACS analysis after staining with anti-CD3-FITC, anti-CD8-PE and anti-CD45RA. CD4 + T cells were purified similarly with RosetteSep™ CD4 T-cell enrichment cocktail. CD45RA^{high} naïve CD4 T cells were further isolated by negative depletion through CD45RO MicroBeads, greater than 90% purity was confirmed by FACS analysis and cells were collected for the 'CD4 T-cell' subset. CD4 cells were then activated by plate-bound anti-CD3 (10 µg/ml) and anti-CD28 (5 µg/ml) in PBS at 4°C overnight. The CD4 cells were stimulated at 20 × 10⁶ cells per 6 ml DMEM plus 10% FCS for 48 h as follows: IL12 (1 ng/ml), IFN-δ (10 ng/ml) and anti-IL4 (1 µg/ml) were added to the media to differentiate Th1 cells, and anti-IL12 (5 µg/ml), anti-IFN-δ (5 µg/ml) and IL4 (1 ng/ml) to differentiate Th2 cells. A control condition using anti-IL12, anti-IFN-δ and anti-IL4 of the same concentrations was also included. Cells were expanded fourfold with fresh media at day 3, further cultured until day 7 and re-stimulated with anti-CD3 (10 µg/ml) and anti-

CD28 (5 µg/ml) for 12 or 48 h and collected for the 'Helper Th1' and 'Helper Th2' subsets at each time point. Th1 and Th2 differentiation was confirmed both by ELISA and intracellular cytokine staining of IFN- δ and IL4.

Memory T cells were isolated with ficoll[®] (ICN Biomedicals) by step gradient separation. Monocytes were depleted by adherence to the culture flask. FACS separated CD45RA and CD45RO highly expressing cells with additional gating on lymphocytes by forward and side scatter. Cells of intermediate expression of either CD45RA or CD45RO were not collected. CD45RO cells were cultured for 16 h in RPMI 1640, 10% heat inactivated FBS, 100 U/ml of Penicillin, 100 mg/ml of streptomycin, 2 mM L-glutamine and IL-2 (100 U/ml) for the 'Memory T-resting' subset, and with the addition of plate-bound anti-CD3 (10 µg/ml) and soluble anti-CD28 (10 µg/ml) for the 'Memory T - activated' subset. FACS confirmed CD69 and CD25 expression for the activated state.

NK cell subset isolation

NK cells were isolated with RosetteSep[™] NK-cell enrichment cocktail and further purified by CD2 microBeads. Cell purity was confirmed by staining with PE anti-CD56 for FACS analysis. Cells were cultured in RPMI 1640, 10% heat-inactivated FBS, 100 U/ml of penicillin, 100 mg/ml of streptomycin, 2 mM L-glutamine, and 5.5×10^{-5} beta-mercaptoethanol. Cells were isolated at time 0 for the 'NK cell' subset, and at 16 h after treatments with IL2 (10 nM) or IL15 (10 nM) for each of the respective activated subsets. FACS determined expression of CD56, defining NK cells, and CD56 and CD69, defining activated NK cells.

B-cell subset isolation

B cells were isolated with MACS[®] CD138 microbeads and CD19 microbeads. FACS sorting separated CD19 APC, CD27 PE and IgM FITC, or IgG and IgA FITC (BD Biosciences) stained cells into the subsets of 'B cell' (CD19+CD27- IgG/A-), 'memory B-IgG/A' (CD19+ CD27+ IgM-) and 'memory B-IgM' (CD19+ CD27+ IgG/A-). Plasma cells were isolated from leukopacks for the 'Plasma B-PBMC' subset or bone marrow for the 'Plasma B-bone marrow' subset by CD138 microbeads and further purified by FACS to collect cells positive for CD20 FITC, CD138 PE and CD19 APC.

Monocyte and macrophage subset isolation

Monocytes were isolated with MACS[®] CD14 Microbeads and cells collected for the 'monocyte' subset. Macrophages were differentiated from monocytes in DMEM 4.5 µg/ml glucose, Pen.-strep.-glutamine, 20% FBS, 10% human AB serum (Gemini, Cat # 100-512). Cells were seeded at 1.5×10^6 per well and grown at 37°C in 7% CO₂. Cells were collected after 1 day for the 'macrophage-1 day' subset, and the remaining cultured for 7 days for the 'macrophage-7 day' subset, with medium fed at about day 5. Macrophage identity (in the absence of a known surface marker) was confirmed by morphology and phagocytic capacity. The differentiated macrophages were highly adherent large cells with numerous filopodia and lamellipodia. Fluorescent latex beads were added to the cultures of monocytes and macrophages

and uptake over 30 min was monitored by fluorescent microscopy; macrophages ingested over 50-fold more beads than did monocytes.

Dendritic cell subset isolation

Monocytes isolated as above were cultured in RPMI with 10% heat-inactivated FBS, 1 × Pen/Strep, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, then differentiated into dendritic cells by 17 ng/ml IL4, and 67 ng/ml GM-CSF for 5 days at 5×10^6 cells/ml. Cells were then collected for the 'dendritic cell' subset. The remaining cells were induced with 1 µg/ml LPS at 2.5×10^6 cells/ml for 1 day for the 'dendritic-LPS' subset.

Neutrophil isolation

Heparanized blood (100 ml) was mixed with an equal volume of dextran/saline solution (3% dextran T-500 in 0.9% saline) in 4 × 50 ml conical tubes and allowed to settle in an upright position for ~20 min. The leukocyte-rich fraction above the red blood cells was layered onto Ficoll-Hypaque solution and centrifuged at 1400 rpm (400 g) for 30 min at 20°C to remove the mononuclear cells and pellet the neutrophils.

Immune-specific gene expression selection

The mean expression level for each probe on Affymetrix[®] HGU133A and HGU133B GeneChip[®] is determined from each of the immune cell subsets listed in Table 1 (immune level) and the following normal tissues (nonimmune level) in the BioExpress[™] System (Gene Logic Inc., Gaithersburg, MD, USA): breast, brain, spinal chord, colorectal, heart, kidney, liver, lung, ovary, pancreas, prostate, small intestine and stomach. The Affymetrix[®] Microarray Suite version 5.0 'signal' is used as the metric of expression level. The BioExpress[™] nonimmune levels are adjusted by a factor of 5 for comparison to the immune levels that had been normalized by a different method, according to the conclusions of chip-wide comparison of the same genes, with the same samples normalized by each method (data not shown). Initial evaluation of gene expression in all the normal tissues represented in BioExpress[™] revealed cases of genes expressed preferentially in immune cells but also in organs with a role in immune function, such as lymphoid tissues (spleen, lymph node, appendix and tonsil), thymus and skin, or other tissues such as endometrium, testis and thyroid (data not shown). Therefore, these tissues were not included in the analysis. Immune cells are also found in tissues such as liver and lung, but the mean expression level dilutes the specific expression of individual cell types among the many other cell types in those tissues, as compared to the expression of a gene in isolated immune cells. Many probesets were found to represent antibody sequences and these were also eliminated from this analysis. A gene (probeset) is defined to be immune-specific if the highest immune level is greater than 2000, the highest non-immune level is less than 4000, and the nonimmune level is less than 30% of the immune level.

Cell and lineage assignment

Patterns of immune cell lineage expression patterns are shown in Figures 1 and 2 and Supplementary Figure 2. This hierarchical clustering identifies patterns of gene

expression in a general way, but the methods used to define the IRIS categories listed in Table 1 are described here. Genes are categorized into immune cell types as defined by the highest level of a probeset's expression multiplied by an arbitrary factor of 0.1625, plus the highest nonimmune level, being greater than the highest level in any other cell type. Probesets with specific expression in more than one cell type are assigned to the lymphoid (T, B and NK cells), myeloid (monocytes, dendritic cells and neutrophils) or the multiple category, which indicates specificity to cell types across these two lineages. For example, if the highest level for a probeset is 10 000 in CD8 T cells but also 2000 in resting NK cells, and the highest nonimmune level is 1000 in kidney, the probeset has T-cell specificity. But if the level is greater than 2625 in NK cells, the probeset has lymphoid specificity. Probesets are assigned to genes as described in *Gene Annotation* and the number of genes within each category is shown in Table 1. A gene may be represented by multiple probesets with varying levels of expression and so may appear in multiple categories, usually due to lower expression that falls below the cutoffs described above.

Statistical assessment

The statistical significance of differential gene expression between the subsets of immune cells is assessed with the *F*-test.⁷⁵ The intensity value from the microarray experiment of each replicate is log-transformed and the variance of the replicates within a cell subset is compared to the variance between the means of the intensity values of all cell subsets. Thus, the *F*-test measures experimental reproducibility and biological differential expression simultaneously. The range of *F* statistics for each IRIS category is shown in Supplementary Figure 4. Table 1 lists the IRIS categories, the cell subsets in each category and number of replicates for each subset.

Gene annotation

Genes are annotated by use of publicly available resources supplemented by additional analysis. In general, IRIS probesets are mapped to genes by aligning both GenBank cDNA and Affymetrix™ target sequences to the NCBI Human Genome Assembly (release 34). Probesets are assigned to the nearest gene by a region defined as 10 kb up- and downstream of cDNAs mapped to the genome. This analysis was carried out independently of publicly available annotation to ensure quality gene assignments that are difficult on a large-scale basis, particularly for ESTs. The LocusLink identifier (<http://www.ncbi.nlm.nih.gov/LocusLink/>) or Genbank accession number is provided for each gene in Supplementary Table 1. Gene Ontology (<http://www.geneontology.org/>) terms associated with a probeset (Table 2) are determined by mapping the probeset to GO terms based on NetAffx™ (Affymetrix) probeset annotation (<http://www.affymetrix.com/analysis/index.affx>). This publicly available annotation was used in this case because of the value added by the GO terminology. Tallies of the incidence of terms associated with a set of probesets are assembled by condensing duplicate probesets representing the same gene and then tallying the terms from each probeset in the condensed list. Deviations of the expected distribution of term mapping of a set of probesets are determined by performing the χ^2 test (once for each

term) on the two variables: (1) genes related to the term *vs* those not related, and (2) genes in IRIS *vs* those not in IRIS. Gene family classifications shown in Table 3 are based on review of literature and web resources such as HUGO (<http://www.gene.ucl.ac.uk/nomenclature>), augmented by alignment with the PFAM (<http://pfam.wustl.edu/index.html>) and SMART (<http://smart.embl-heidelberg.de/>) collections of protein domains. While PFAM is very useful and comprehensive, additional analysis and use of the other resources mentioned is required to define a gene family. The Protcomp algorithm (Softberry, Inc.) is used to predict the cellular component of the proteins encoded by IRIS genes (Protcomp accuracy measurements at <http://www.softberry.com>). The nuclear designation here includes just nuclear (non-membrane-bound), plasma membrane includes plasma membrane and membrane-bound extracellular and secreted are just extracellular. Localization assignments were corrected for obvious mis-predictions of well-characterized genes.

Statistical clustering

Clustering and visualization are carried out using the DecisionSite™ software (Spotfire™). Hierarchical clustering (Figure 1) is performed by the UPGMA (unweighted average method), the correlation similarity measure and the average value-ordering function. The immune cell subset expression values are summarized into the highest mean expression in a subset for each cell type and all IRIS genes are included. K-means clustering (Figure 2 and Supplementary Figure 2) is also carried out with the correlation similarity measure and evenly spaced profiles are used for cluster initialization. Here, all the immune cell subset mean expression levels of the relevant cell type or lineage are included for just the genes within that category.

Highly specific genes

Genes shown in Figure 3 and described in Results and discussion were selected from the IRIS categories for just one cell type and have the highest immune level 10-fold higher than the highest nonimmune level. A more refined CD45RO memory T-cell experiment is found to show somewhat different expression levels for some genes, although this did not affect the selection of IRIS genes significantly (data not shown). However, the expression levels for this experiment, as well as an additional data point for the Th2 subset at 12 h, are shown in Figure 3a and Supplementary Figure 3a to best represent these genes, and are also provided in Supplementary Table 1. One probeset per gene was selected based on the highest level of expression in any immune cell subset. Protein domains were predicted from PFAM for genes of unknown function.^{76,77}

Web Sites

IRIS gene expression profiles

<http://share.gene.com/share/clark.iris.2004/iris/iris.html>

Gene Ontology Consortium

<http://www.geneontology.org/>

NetAffx™ GO assignments

<http://www.affymetrix.com/analysis/index.affx>

LocusLink

<http://www.ncbi.nlm.nih.gov/LocusLink/>

HUGO
<http://www.gene.ucl.ac.uk/nomenclature>
 PFAM
<http://pfam.wustl.edu/index.html>
 SMART
<http://smart.embl-heidelberg.de/>
 Protcomp
<http://www.softberry.com>

Acknowledgements

Kay Dennis, Henry Chiu, Sudeepta Aggarwal, Vandana Chauhan and Fay Wang for sample preparation, Karen Toy, Grazyna Fedorowicz and Rob Soriano for microarray experiments, Thomas Wu, Cynthia Honchell and Mike Ward for microarray analysis tools, Zemin Zhang, Yan Zhang and Shih-Ming Luoh for gene annotation tools, Michael Ostland for statistical analysis, Jill Schoenfeld for gene analysis, Dan Eaton and Bernd Wranik for identification of BTLA gene and Allison Bruce for graphics services.

References

- 1 Alizadeh A, Staudt LM. Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr Opin Immunol* 2000; **12**: 219–225.
- 2 Marrack P, Mitchell T, Hildeman D *et al*. Genomic-scale analysis of gene expression in resting and activated T cells. *Curr Opin Immunol* 2000; **12**: 206–209.
- 3 Calame K. Plasma cells: finding new light at the end of B cell development. *Nat Immunol* 2001; **2**: 1103–1108.
- 4 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 2T cells. *J Immunol* 2001; **167**: 3057–3063.
- 5 Ma C, Staudt LM. Molecular definition of the germinal centre stage of B-cell differentiation. *Philos Trans R Soc London Ser B Biol Sci* 2001; **356**: 83–89.
- 6 Wilson S, Byrne MC. Gene expression in NKT cells: defining a functionally distinct CD1d-restricted T cell subset. *Curr Opin Immunol* 2001; **13**: 555–561.
- 7 Hodge D, Schill WB, Wang JM *et al*. IL-2 and IL-12 alter NK cell responsiveness to IFN-gamma-inducible protein 10 by down-regulating CXCR3 expression. *J Immunol* 2002; **168**: 6090–6098.
- 8 Ollila J, Vihinen M. Microarray analysis of B-cell stimulation. *Vitamins Hormones* 2002; **64**: 77–99.
- 9 Re F, Belyanskaya SL, Riese RJ *et al*. Granulocyte-macrophage colony-stimulating factor induces an expression program in neonatal microglia that primes them for antigen presentation. *J Immunol* 2002; **169**: 2264–2273.
- 10 Richards J, Le Naour F, Hanash S, Beretta L. Integrated genomic and proteomic analysis of signaling pathways in dendritic cell differentiation and maturation. *Ann NY Acad Sci* 2002; **975**: 91–100.
- 11 Rogge L. A genomic view of helper T cell subsets. *Ann NY Acad Sci* 2002; **975**: 57–67.
- 12 Verdeil G, Puthier D, Nguyen C, Schmitt-Verhulst AM, Auphan-Anezin N. Gene profiling approach to establish the molecular bases for partial *vs* full activation of naive CD8T lymphocytes. *Ann NY Acad Sci* 2002; **975**: 68–76.
- 13 Vinuesa C, Cook MC, Cooke MP, MacLennan IC, Goodnow CC. Analysis of B cell memory formation using DNA microarrays. *Ann NY Acad Sci* 2002; **975**: 33–45.
- 14 Weng N, Liu K, Catalfamo M, Li Y, Henkart PA. IL-15 is a growth factor and an activator of CD8 memory T cells. *Ann NY Acad Sci* 2002; **975**: 46–56.

- 15 Koopman L, Kopcow HD, Rybalov B *et al*. Human decidua natural killer cells are a unique NK cell subset with immunomodulatory potential. *J Exp Med* 2003; **198**: 1201–1212.
- 16 Messmer D, Messmer B, Chiorazzi N. The global transcriptional maturation program and stimuli-specific gene expression profiles of human myeloid dendritic cells. *Int Immunol* 2003; **15**: 491–503.
- 17 Tsapogas P, Breslin T, Bilke S *et al*. RNA analysis of B cell lines arrested at defined stages of differentiation allows for an approximation of gene expression patterns during B cell development. *J Leukocyte Biol* 2003; **74**: 102–110.
- 18 Lee S, Hanspers K, Barker CS, Korn AP, McCune JM. Gene expression profiles during human CD4+ T cell differentiation. *Int Immunol* 2004; **16**: 1109–1124.
- 19 Mantovani S, Palermo B, Garbelli S *et al*. Dominant TCR-alpha requirements for a self antigen recognition in humans. *J Immunol* 2002; **169**: 6253–6260.
- 20 Schmidt K, Leung B, Kwong M *et al*. APC-Independent Activation of NK Cells by the Toll-Like Receptor 3 Agonist Double-Stranded RNA. *J Immunol* 2004; **172**: 138–143.
- 21 Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004; **32**: D258–D261.
- 22 Winter E, Goodstadt L, Ponting CP. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res* 2004; **14**: 54–61.
- 23 Schmitz JE, Forman MA, Lifton MA *et al*. Expression of the CD8-heterodimer on CD8+ T lymphocytes in peripheral blood lymphocytes of human immunodeficiency virus and human immunodeficiency virus+individuals. *Blood* 1998; **92**: 198–206.
- 24 Dorfman D, Greisman HA, Shahsafaei A. Loss of expression of the WNT/beta-catenin-signaling pathway transcription factors lymphoid enhancer factor-1 (LEF-1) and T cell factor-1 (TCF-1) in a subset of peripheral T cell lymphomas. *Am J Pathol* 2003; **162**: 1539–1544.
- 25 Hutloff A, Dittrich AM, Beier KC *et al*. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* 1999; **397**: 263–266.
- 26 Wilharm E, Parry MA, Friebel R *et al*. Generation of catalytically active granzyme K from *Escherichia coli* inclusion bodies and identification of efficient granzyme K inhibitors in human plasma. *J Biol Chem* 1999; **274**: 27331–27337.
- 27 Andreea S, Piras F, Burdin N, Triebel F. Maturation and activation of dendritic cells induced by lymphocyte activation gene-3 (CD223). *J Immunol* 2002; **168**: 3874–3880.
- 28 Vitale M, Falco M, Castriconi R *et al*. Identification of NKp80, a novel triggering molecule expressed by human NK cells. *Eur J Immunol* 2001; **31**: 233–242.
- 29 Winter C, Gumperz JE, Parham P, Long EO, Wagtmann N. Direct binding and functional transfer of NK cell inhibitory receptors reveal novel patterns of HLA-C allotype recognition. *J Immunol* 1998; **161**: 571–577.
- 30 Wolff T, O'Neill RE, Palese P. NS1-Binding protein (NS1-BP): a novel human protein that interacts with the influenza A virus nonstructural NS1 protein is relocalized in the nuclei of infected cells. *J Virol* 1998; **72**: 7170–7180.
- 31 Virgilio L, Lazzeri C, Bichi R *et al*. Deregulated expression of TCL1 causes T cell leukemia in mice. *Proc Natl Acad Sci USA* 1998; **95**: 3885–3889.
- 32 Laine J, Kunstle G, Obata T, Sha M, Noguchi M. The protooncogene TCL1 is an Akt kinase coactivator. *Mol Cell* 2000; **6**: 395–407.
- 33 Grillo-López A. AntiCD20 mAbs: modifying therapeutic strategies and outcomes in the treatment of lymphoma patients. *Expert Rev Anticancer Ther* 2002; **2**: 323–329.
- 34 Thompson J, Bixler SA, Qian F *et al*. BAFF-R, a newly identified TNF receptor that specifically interacts with BAFF. *Science* 2001; **293**: 2108–2111.

- 35 Yokoyama K, Su IH, Tezuka T *et al*. BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J* 2002; **21**: 83–92.
- 36 Nutt S, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* 1999; **401**: 556–562.
- 37 Brouns G, de Vries E, Borst J. Assembly and intracellular transport of the human B cell antigen receptor complex. *Int Immunol* 1995; **7**: 359–368.
- 38 Daeron M. Structural bases of Fc gamma R functions. *Int Rev Immunol* 1997; **16**: 1–27.
- 39 Davis RS, Wang YH, Kubagawa H, Cooper MD. Identification of a family of Fc receptor homologs with preferential B cell expression. *Proc Natl Acad Sci USA* 2001; **98**: 9772–9777.
- 40 Hatzivassiliou G, Miller I, Takizawa J *et al*. IRTA1 and IRTA2, novel immunoglobulin superfamily receptors expressed in B cells and involved in chromosome 1q21 abnormalities in B cell malignancy. *Immunity* 2001; **14**: 277–289.
- 41 Facchetti F, Cella M, Festa S, Fremont DH, Colonna M. An unusual Fc receptor-related protein expressed in human centroblasts. *Proc Natl Acad Sci USA* 2002; **99**: 3776–3781.
- 42 Miller I, Hatzivassiliou G, Cattoretti G, Mendelsohn C, Dallafavera R. IRTAs: a new family of immunoglobulinlike receptors differentially expressed in B cells. *Blood* 2002; **99**: 2662–2669.
- 43 Hatzoglou A, Roussel J, Bourgeade MF *et al*. TNF receptor family member BCMA (B cell maturation) associates with TNF receptor-associated factor (TRAF) 1, TRAF2, and TRAF3 and activates NF-kappa B, elk-1, c-Jun N-terminal kinase, and p38 mitogen-activated protein kinase. *J Immunol* 2000; **165**: 1322–1330.
- 44 Moog-Lutz C, Peterson EJ, Lutz PG *et al*. PRAM-1 is a novel adaptor protein regulated by retinoic acid (RA) and promyelocytic leukemia (PML)-RA receptor alpha in acute promyelocytic leukemia cells. *J Biol Chem* 2001; **276**: 22375–22381.
- 45 Borish LaSJ. Cytokines and chemokines. *J Allergy Clin Immunol* 2003; **111**: 460–475.
- 46 Yousefi S, Cooper PR, Potter SL, Mueck B, Jarai G. Cloning and expression analysis of a novel G-protein-coupled receptor selectively expressed on granulocytes. *J Leukoc Biol* 2001; **69**: 1045–1052.
- 47 Willmann K, Dunne JF. A flow cytometric immune function assay for human peripheral blood dendritic cells. *J Leukoc Biol* 2000; **67**: 536–544.
- 48 Hattar K, Fink L, Fietzner K *et al*. Cell density regulates neutrophil IL-8 synthesis: role of IL-1 receptor antagonist and soluble TNF receptors. *J Immunol* 2001; **166**: 6287–6293.
- 49 Bakouche O MJ, Lachman LB. Induction of IL-1: independent production of IL-1 alpha and IL-1 beta. *Cytokine* 1992; **4**: 106–113.
- 50 Debets R, Timans JC, Homey B *et al*. Two novel IL-1 family members, IL-1 delta and IL-1 epsilon, function as an antagonist and agonist of NF-kappa B activation through the orphan IL-1 receptor-related protein 2. *J Immunol* 2001; **167**: 1440–1446.
- 51 Wolk K, Kunz S, Asadullah K, Sabat R. Cutting edge: immune cells as sources and targets of the IL-10 family members? *J Immunol* 2002; **168**: 5397–5402.
- 52 Sheu B, Hsu SM, Ho HN, Lien HC, Huang SC, Lin RH. A novel role of metalloproteinase in cancer-mediated immunosuppression. *Cancer Res* 2001; **61**: 237–242.
- 53 Martin L, Calabi F, Lefebvre FA, Bilslund CA, Milstein C. Structure and expression of the human thymocyte antigens CD1a, CD1b, and CD1c. *Proc Natl Acad Sci USA* 1987; **84**: 9189–9193.
- 54 Blumberg R, Terhorst C, Bleicher P *et al*. Expression of a nonpolymorphic MHC class I-like molecule, CD1D, by human intestinal epithelial cells. *J Immunol* 1991; **147**: 2518–2524.
- 55 Suzuki N, Yamamoto K, Toyoshima S, Osawa T, Irimura T. Molecular cloning and expression of cDNA encoding human macrophage C-type lectin. Its unique carbohydrate binding specificity for Tn antigen. *J Immunol* 1996; **156**: 128–135.
- 56 Conrad D, Kuhn H, Mulkins M, Highland E, Sigal E. Specific inflammatory cytokines regulate the expression of human monocyte 15-lipoxygenase. *Proc Natl Acad Sci USA* 1992; **89**: 217–221.
- 57 Ying S, Meng Q, Zeibecoglou K *et al*. Eosinophil chemotactic chemokines (eotaxin, eotaxin-2, RANTES, monocyte chemoattractant protein-3 (MCP-3), and MCP-4), and C-C chemokine receptor 3 expression in bronchial biopsies from atopic and nonatopic (Intrinsic) asthmatics. *J Immunol* 1999; **163**: 6321–6329.
- 58 Penna G, Vulcano M, Roncari A, Facchetti F, Sozzani S, Adorini L. Cutting edge: differential chemokine production by myeloid and plasmacytoid dendritic cells. *J Immunol* 2002; **169**: 6673–6676.
- 59 Yamashiro S, Yamakita Y, Ono S, Matsumura F. Fascin, an actin-bundling protein, induces membrane protrusions and increases cell motility of epithelial cells. *Mol Biol Cell* 1998; **9**: 993–1006.
- 60 Brown J, Dorfman DM, Ma F-R *et al*. Blockade of programmed death-1 ligands on dendritic cells enhances T cell activation and cytokine production. *J Immunol* 2003; **170**: 1257–1266.
- 61 Bux J, Kissel K, Hofmann C, Santoso S. The use of allele-specific recombinant Fc gamma receptor IIIb antigens for the detection of granulocyte antibodies. *Blood* 1999; **93**: 357–362.
- 62 Hua J, Hasebe T, Someya A, Nakamura S, Sugimoto K, Nagaoka I. Evaluation of the expression of NADPH oxidase components during maturation of HL-60 cells to neutrophil lineage. *J Leukoc Biol* 2000; **68**: 216–224.
- 63 Yoshitake H, Takeda Y, Nitto T, Sendo F. Cross-linking of GPI-80, a possible regulatory molecule of cell adhesion, induces up-regulation of CD11b/CD18 expression on neutrophil surfaces and shedding of L-selectin. *J Leukoc Biol* 2002; **71**: 205–211.
- 64 Nagase H, Miyamasu M, Yamaguchi M *et al*. Cytokine-mediated regulation of CXCR4 expression in human neutrophils. *J Leukoc Biol* 2002; **71**: 711–717.
- 65 Sheridan JP, Marsters SA, Pitti RM *et al*. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 1997; **277**: 818–821.
- 66 Zhang X, Nguyen T, Thomas WD, Sanders JE, Hersey P. Mechanisms of resistance of normal cells to TRAIL induced apoptosis vary between different cell types. *FEBS Lett* 2000; **482**: 193–199.
- 67 Cawley S, Bekiranov S, Ng HH *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 2004; **116**: 499–509.
- 68 Grenier J, Wang L, Manji GA *et al*. Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-kappaB and caspase-1. *FEBS Lett* 2002; **530**: 73–78.
- 69 Strausberg R, Feingold EA, Grouse LH *et al*. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* 2002; **99**: 16899–16903.
- 70 Sanchez-Pulido L, Martin-Belmonte F, Valencia A, Alonso MA. MARVEL: a conserved domain involved in membrane apposition events. *Trends Biochem Sci* 2002; **27**: 599–601.
- 71 Watanabe N, Gavrieli M, Sedy JR *et al*. BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1. *Nat Immunol* 2003; **4**: 670–679.
- 72 Baechler E, Batliwalla FM, Karypis G *et al*. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 2003; **100**: 2610–2615.

- 73 Bennett L, Palucka AK, Arce E *et al*. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 2003; **197**: 711–723.
- 74 Han G, Chen SL, Shen N, Ye S, Bao CD, Gu YY. Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. *Genes Immunity* 2003; **4**: 177–186.
- 75 Kerr K, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *J Comput Biol* 2000; **7**: 819–837.
- 76 Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998; **14**: 755–763.
- 77 Bateman A, Birney E, Cerruti L *et al*. The Pfam protein families database. *Nucleic Acids Res.* 2002; **30**: 276–280.

Supplementary information accompanies the paper on Genes and Immunity website (<http://www.nature.com/gene>).