Tetsuya Yamagata
Christophe Benoist
Diane Mathis

A shared gene-expression signature in innate-like lymphocytes

Authors' address
Tetsuya Yamagata[1], *Christophe Benoist*[2], *Diane Mathis*[2]
[1]Department of Hematology, Dokkyo Medical School, Tochigi, Japan
[2]Department of Medicine, Section on Immunology and Immunogenetics, Joslin Diabetes Center, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence to:
Diane Mathis
Department of Medicine
Section on Immunology and Immunogenetics
Joslin Diabetes Center
Brigham and Women's Hospital
Harvard Medical School
One Joslin Place
Boston, MA 02215
USA
Tel.: +1 617 264 2745
Fax: +1 617 264 2744
E-mail: cbdm@joslin.harvard.edu

Acknowledgements
We thank Ananda Goldrath, Ann Herman, Reinhard Obst, and Linh Nguyen for sharing microarray data and members of the T-cell differentiation group for insightful discussion. This work was supported by funds from the NIH. Tetsuya Yamagata is supported by a Mary K. Iacocca Research Fellowship. Microarray data have been deposited at the GEO database (http://www.ncbi.nlm.nih.gov/geo/), under accession GSE3039.

Immunological Reviews 2006
Vol. 210: 52–66

Immunological Reviews
0105-2896

Summary: Innate and adaptive immunities are the two major arms of the immune system, which rely on distinct cell types. These cells can be distinguished not only by the source of diversity for non-self recognition, of germline or somatic origin, but also by their localization and the pattern and rates of response after encounter of antigenic triggers. In addition, subsets of lymphocytes exist whose receptors require rearrangement but result in semi-invariant structures with a high degree of self-specificity. We hypothesized that these innate-like lymphocytes might share a common gene transcription signature that relates them to classic members of the innate immune system. This relationship was first observed in agonist-induced CD8αα T cells in fetal/neonatal thymus. We then asked whether this notion could be extended to other innate-like lymphocytes, by comparison of gene expression profiles of innate-like lymphocytes and closely paired adaptive system counterparts (NKT versus CD4T, CD8ααT versus CD8αβT, and B1 versus B2). A statistically significant 'innate signature' indeed was distilled. Particularly intriguing was the high representation of interferon-inducible guanosine triphophatases crucial for resistance against intracellular pathogens and of small G proteins involved in intracellular vacuole maturation and trafficking. Overall, this combined expression pattern can be designated as an innate signature among lymphocytes.

Introduction

The innate/adaptive dichotomy is now a strong principle of immunology, even if it is also clear that the boundaries are blurred and that both systems considerably affect each other. After a long dominance of interest in adaptive immunity (1, 2), the discovery of somatically invariant receptors (e.g. the Toll-like receptors) that trigger cell activation upon recognition of obligate microbial structures propelled a reawakening of interest in the innate immune system (3). Employing germline-encoded antigen receptors, innate immunity enables quick cascades of cellular reactions against microbial invasion, while adaptive immunity delivers a delayed but more specific response, exploiting selective amplification of cell clones displaying rearranged antigen receptors to achieve more effective clearance of the pathogens. This categorization is underlined by comparative immunology: the adaptive immune system

does not exist in living animals directly descending from ancestors older than the cartilaginous fish (e.g. sharks); subsequently, there was an extraordinary outbreak in the diversity of the antigen receptor (sometimes referred to as the 'immunological big bang') (4, 5). Comparative immunology initially defined adaptive immunity by the presence of rearranged antigen receptor genes (6), but more recent analyses have revealed many potential similarities in the innate and adaptive immune systems in terms of the diversity in antigen receptors and adaptation to microbial challenges (7–9). It can be considered that immune receptors vary and adapt to pathogens on two different time scales: slowly, on an evolutionary clock of Darwinian selection for the germline-encoded receptors of macrophages, natural killer (NK), dendritic, or mast cells and, rapidly, on the individual's scale of clonal selection for the somatically rearranged and mutated receptors of T and B lymphocytes.

Constituents of the innate immune system, however, are not defined solely by their lack of rearranged antigen receptors. In fact, several characteristics of innate system cell types distinguish them from cell types of the adaptive system: the two have different tissue localization, and the former exhibit a quicker response, more robust effector function, and absence of memory (10–12). There is also an intriguing group of cells whose phenotypic features place them at the border of the innate and adaptive immune systems (reviewed in 13). These 'innate-like' lymphocytes include B1 cells, marginal zone B cells, γδ T cells, NKT cells, and intestinal epithelial lymphocytes (IELs), including CD8αα cells. These cell types use somatically rearranged receptors for antigen recognition, but these receptors tend to be invariant, utilizing germline-encoded elements to recognize simple molecular structures from microbes and pathogens as well as self-antigens, such as altered carbohydrates or lipids. Innate-like lymphocytes reside mostly in tissues rather than in the secondary lymphoid organs, constitutively express a memory phenotype ($CD45RO^+$), and unleash quick effector functions such as instantaneous cytolysis and rapid cytokine, chemokine, and antibody secretion (*Fig. 1*).

These are all characteristics of cells of the innate immune system (11, 13–15). One might surmise, then, that these common phenotypic characteristics might be underpinned by common genomic characteristics, shared signatures of gene expression. In this text, we review our recent analyses in this area.

Genetic and genomic analyses are complementary approaches to a comprehensive understanding of gene function in complex biological systems. The quasi-complete

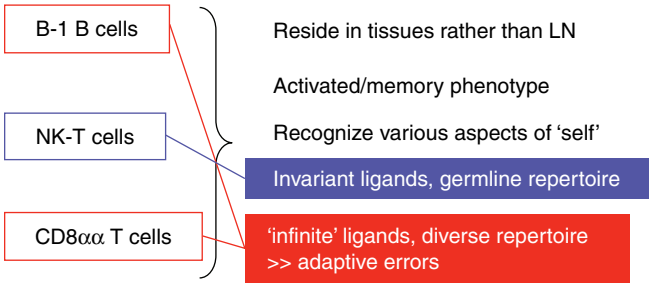


***Fig. 1.* Shared characteristics of innate-like lymphocytes.**

determination of genome sequences for certain species provides a longitudinal map of the receptor and effector structures that might operate in their immune systems, while the determination of gene expression patterns in individual cells or groups of related cells brings an orthogonal perspective on the extent to which this potential is actually utilized. The global perspective offered by microarray data also affords the possibility of analyzing the full complexity of gene regulatory networks and of identifying gene expression 'signatures' shared by functionally related cell types (16). The gene expression signature is a set of expressed genes that connects a given biological process or differentiation states of cells to a molecular reality represented at the mRNA expression level. Once determined, the signature can be used to search the gene expression profiles from different samples and identify characters associated with the signature in other contexts.

In this vein, comparing gene expression profiles between innate and adaptive immune system cell types might help to highlight fundamental features of innate and adaptive immunity. One might, for example, hypothesize that there are commonalities between macrophages, neutrophils, dendritic cells, and NK cells that set them apart from the classical, adaptive, B lymphocytes, and T lymphocytes. This question could be formulated as 'Is there a gene expression signature for innate immunity as a whole?' In the experiments reviewed here, we have analyzed gene expression signatures in a variety of cell types but, in particular in innate-like lymphocytes, in which an 'innate signature' might be easiest to distinguish, by comparison with closely paired adaptive lymphocytes.

Expression of genes typical of the innate immune system in agonist-induced CD8αα cells

Our first observations on the presence of a particular signature in innate-like cells arose during studies on T-cell differentiation induced in reaggregation thymic organ cultures (17). This experimental system started from 'blocked' double-positive cells, isolated from transgenic mice whose combination of

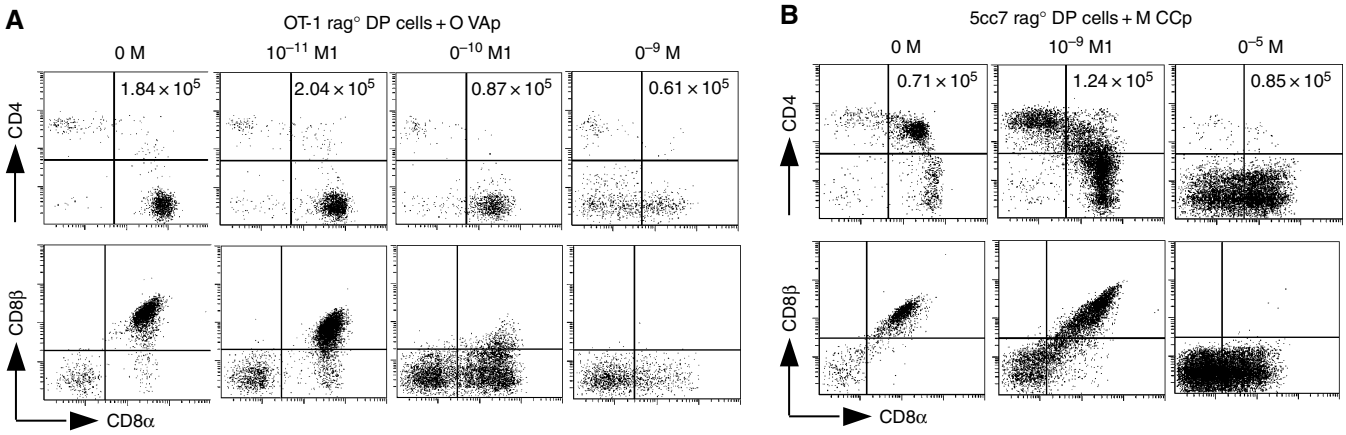


Fig. 2. Mature CD8αα cells of other TCR specificities are elicited by their cognate peptide. Unselected DPs from two different TCR-transgenic mice [(A) class I-restricted OT-1 and (B) class II-restricted 5CC7] were cultured in RTOC with graded doses of the relevant agonist peptide (OVAp or MCCp, respectively). The cultures were stained for CD4/CD8 or CD8α/CD8β after 4 days in culture.

T-cell receptor (TCR) and major histocompatibility complex (MHC) precludes any positive selection. Addition of these precursors to alymphoid thymic rudiments, in which the proper selecting molecules and peptides could be found, resulted in highly synchronous differentiation of the input cells. Most interestingly, when a strong agonist peptide was added in titrated doses, CD8αα T cells arose from these cultures, as illustrated in *Fig.* 2.

This induced differentiation was quite reminiscent of the influence of agonist peptide recognition in the presence of CD8αα cells in the male HY TCR-transgenic mouse (18–20). As shown in *Fig.* 2, the appearance of CD8αα cells was independent of the nominal MHC restriction of the transgenic TCR, as the same behavior could be elicited in OT-I or AND transgenic thymocytes. Global gene expression profiles were then determined to explore the characteristics of CD8αα T cells, comparing their microarray profiles to those of more conventional $CD8^+$ cells elicited by peptides normally loaded on the positively selecting MHC molecules or by much lower doses of agonist peptides. As illustrated in *Fig.* 3, many of the genes distinguishing these CD8αα T cells were genes typically expressed in NK cells.

Upon closer examination, it became apparent that it was not simply an NK signature that was showing up in these comparisons but that the differentially expressed genes also included transcripts more generally expressed in innate-like cells: genes encoding the transcription factor Id2 or the signal transducers DAP-12 and FcεRγ, whose broad role had been

CD8ααT versus CD8αβT	Gene	Affy ID	Entrez ID	Annotation
8.81	Car2	92642_at	12349	Carbonic anhydrase II
8.73	Cd244	95380_at	18106	Natural killer-cell receptor 2B4
7.44	Idb2	93013_at	15902	Inhibitor of DNA binding 2
7.33	Napsa	101972_at	16541	Aspartic peptidase
7.16	Klrb1a	94772_at	17057	Killer cell lectin-like receptor
6.71	FceR1g	162181_f_at	14127	ITAM signaling
5.82	Anxa2	100569_at	12306	Annexin A2
5.82	Mdfic	102644_at	16543	MyoD family inhibitor domain containing
5.61	Serpina3g	102860_at	20715	Serine/cysteine proteinase inhibitor
5.35	Wbscr5	161899_f_at	56743	Williams-Beuren syndrome
5.30	Tbl1xr1	103889_at	81004	Transducin (β)-like 1X-linked receptor 1
5.13	Ccl5	98406_at	20304	Chemokine (C-C motif) ligand 5
5.08	Ahnak	160255_at	66395	AHNAK nucleoprotein (desmoyokin)
4.84	EST	103090_at	56046	
4.69	Spock2	104375_at	94214	Calcium ion binding
4.60	Cxcr3	94173_at	12766	Chemokine (C-X-C motif) receptor 3
4.50	Xdh	97950_at	22436	Xanthine dehydrogenase
4.41	Kit	99956_at	16590	Transmembrane receptor tyrosine kinase
4.33	Klrb1c	93380_at	17059	Natural killer-cell-mediated cytotoxicity

Fig. 3. Genes most differentially expressed in CD8αα relative to CD8αβ T cells. Microarray analysis was used to compare genes expressed in mature CD8αα versus CD8αβ single-positive T cells, in reaggregation fetal thymic organ culture (essentially as in *Fig.* 2; see 17). Genes are sorted according to their differential expression in CD8αα T cells. Genes highlighted in orange are typically expressed in NK cells or in other cells of the innate immune system.

established in several knockout studies (21–25). Thus, a strong signal triggered at the TCR level elicits two fundamental (and yet perhaps not unrelated) changes. First, the shutoff of CD8β gene expression, in a single molecular switch, eliminates the CD8αβ coreceptor required for effective signaling downstream of classical TCR-MHC class I molecule engagements and brings out the CD8αα coreceptor, a TCR-independent ligand for the non-classical class I molecule TL (26). Second, a broader programmatic change activates a panel of genes typical of the innate immune system, ranging from cell-surface receptors to signaling molecules or transcription factors, several of which are integral to the differentiation and function of innate immune system cells: NK receptors (NCRs), DAP12, FcεRγ, c-kit, etc. Most provocative is the strong induction of Id2, a member of the inhibitor of DNA-binding family and a key factor for NK and dendritic cell development (24, 25). The Id2 overexpression in CD8αα cells is accompanied by a downregulation of E2A, which has exactly the opposite effect to Id2 on the NK-T cell balance (27). Might Id2 and E2A be the direct downstream targets of strong TCR engagement and control the other changes?

General differential gene expression profiles of innate-like lymphocytes

We then sought to broaden these observations in CD8αα cells, by asking whether transcripts typical of cells of the innate immune system would be found in other innate-like lymphocytes and whether one might identify transcriptional features shared between three innate-like cell types versus their paired adaptive system counterparts. The following cell pairs were then sorted: B1 lymphocytes ($IgM^{hi}IgD^{int}CD11b^{+}$) were sorted from peritoneal lavage cells, while conventional B2 lymphocytes ($IgM^{int}IgD^{hi}CD11b^{-}$) were isolated from spleen cell suspensions prepared from the same mice (*Fig. 4A, top panels*).

NKT cells were sorted from spleens as $CD3^{+}CD4^{+}NK1.1^{+}$ cells, together with $CD3^{+}CD4^{+}NK1.1^{-}$ conventional T cells (*Fig. 4A, middle panels*). Total RNA was extracted from these sorted populations and was used to probe microarrays (full data are available upon request). In these comparisons, we re-employed the data sets from conventional CD8αβT or agonist-induced CD8ααT cells induced in fetal thymic organ cultures. Typical cell isolations are reproduced in *Fig. 4A* (right panels) for comparison. In several instances, the initial microarray differences were verified on independent samples by quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) (see below).

The basic approach of the bioinformatic analyses was to first identify for each pair the differentially expressed genes and then to highlight elements conserved across the three differentials. This procedure aimed to extract a set of genes whose transcription was characterizing innate-like cell types (hereafter simplified to 'innate genes') as opposed to loci whose expression typified adaptive system cell types ('adaptive genes'). A preliminary hierarchical clustering of the 1936 differentially expressed genes (filtered on a fold-change >1.5 in any two-way comparison) served to substantiate our assumption that the innate (CD8ααT, NKT, and B1) versus adaptive (CD8αβT, CD4T, and B2) cell pairs are 'close but distinct' (*Fig. 4B*). The expression patterns of the two members of each pair were more similar to each other than to those of any other cell type, as evidenced by the dendogram on the top (e.g. the NKT profile was closest to that of $CD4^{+}$ T cells, those of B1 cells and B2 cells closest, etc.). Yet the expression profiles also showed differentially expressed genes within each pair. These differences are depicted on the fold-change/expression plots of each of the three innate versus adaptive pairs (*Fig. 4C*). Literature database searching revealed that a sizeable number of the differentially expressed genes corresponded to those expected to show divergent transcription in these cell pairs (28–31), validating the relevance and significance of these data (Given our stated purpose, to distill particular gene expression programs and signatures, we will not detail these individual differences any further here). For a broad view of the magnitude of differential gene expression between the paired innate and adaptive cell types, we plotted the number of loci whose transcription diverged between the two members of each pair using the 'Changes' algorithm (32) (*Fig. 4D*). These plots show that significantly more genes were overexpressed, as opposed to underexpressed, in innate-like lymphocytes (red and blue lines, respectively; compare with the black lines representing the number of changes in the randomized control data set). This observation suggests that innate lymphocytes actually have additional transcriptional potentialities but not at the expense of downregulating other expressions programs (13, 28, 33).

To address which categories of proteins (e.g. cell-surface receptors, signaling molecules, transcription factors) were overexpressed in innate-like lymphocytes, we filtered genes that were overexpressed greater than twofold in innate cells and sorted them into cell-component categories according to gene ontology (GO) database annotations (*Fig. 5A*).

While the overexpressed genes covered a broad spectrum of cellular components, nearly half of them fell into the 'integral to membrane' category. Frequent induction of integral to

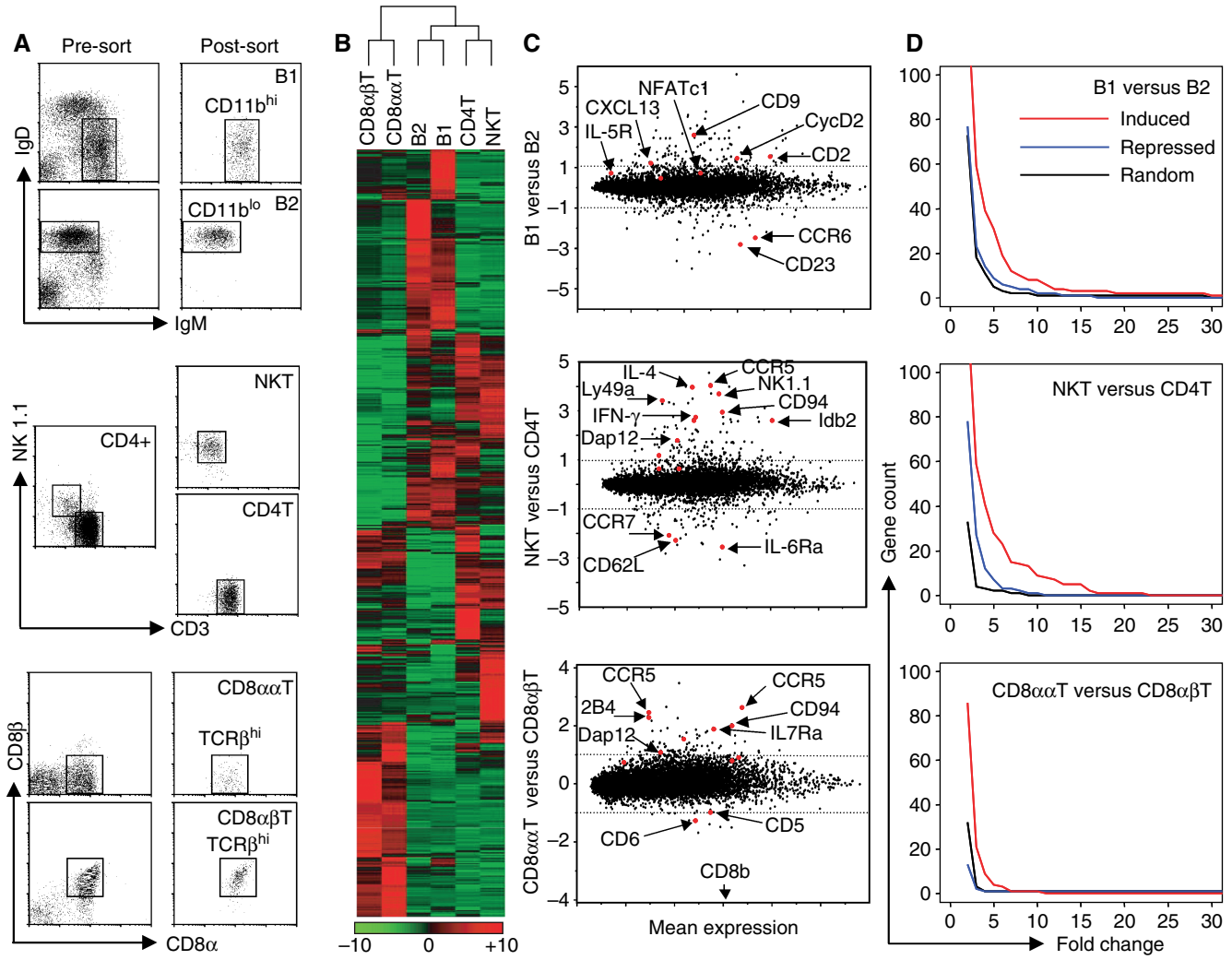


***Fig. 4.* Differential gene expression in innate-like lymphocytes relative to their adaptative counterparts.** (A) Flow cytometry profiles prior sorting and postsorting of the cells analyzed in this study. (B) Hierarchical clustering of the genes and cell populations, performed on 1936 differentially expressed genes. Red and green: relative expression superior or inferior to the gene-wise mean. (C) Fold-change (log2) versus expression plots for the three lymphocyte pairs. Each dot represents an individual gene. Red dots represent the genes whose differential expression was expected from the literature, or from our own quantitative RT-PCR analysis. Positive values indicate higher expression in innate cells. (D) Changes plot representing fold-change (*x*-axis) and number of genes for that fold-change (*y*-axis). More genes are overexpressed than underexpressed (red and blue, respectively) in innate lymphocytes compared with their adaptive counterparts. The black line represents the same Changes calculation performed after populationwise randomization of the data.

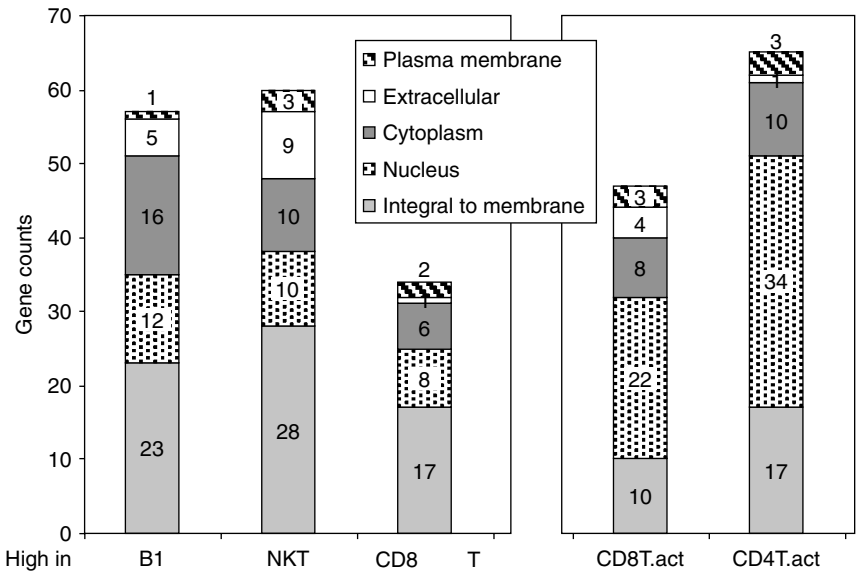
membrane molecules was statistically significant in all three innate : adaptive cell type pairs (Pearson χ^2 test), further supporting the significance of the observations (experimental noise would not be expected to be shared between data sets) (*Fig. 5B*).

Several reports have claimed that innate lymphocytes share several characteristics with activated lymphocytes (29, 34). Hence, the preferential expression of membrane-associated molecules could be thought to reflect merely a general state of activation rather than a specific feature of innate lymphocytes. Therefore, we took as additional comparators gene expression data sets comparing naive and activated mature T cells (both $CD4^+$ and $CD8^+$) (35, L. Nguyen and R. Obst, unpublished data). As illustrated in *Fig. 5C*, only a subset of the genes overexpressed in the innate-like lymphocytes relative to their adaptive counterparts corresponded to genes induced during cell activation/proliferation. In addition, changes occurring during lymphocyte activation were predominantly encoded nuclear proteins (cell-cycle control, transcription factors), differing from the membrane proteins that dominate the innate/adaptative comparison (*Fig. 5B*).

Coordinate view of the cell populations' expression profiles

We then attempted to utilize the global gene expression profiles of the six populations to discern their relative

A



B

	B1	NKT	CD8aaT	CD8T.act	CD4T.act
Plasma membrane	0.854	<0.001	0.205	0.267	0.058
Extracellular	0.785	0.018	0.462	0.878	0.107
Cytoplasm	0.829	0.282	0.236	0.056	0.039
Nucleus	0.564	0.297	0.630	0.104	<0.001
Integral to membrane	0.048	0.001	0.043	0.134	0.621

C

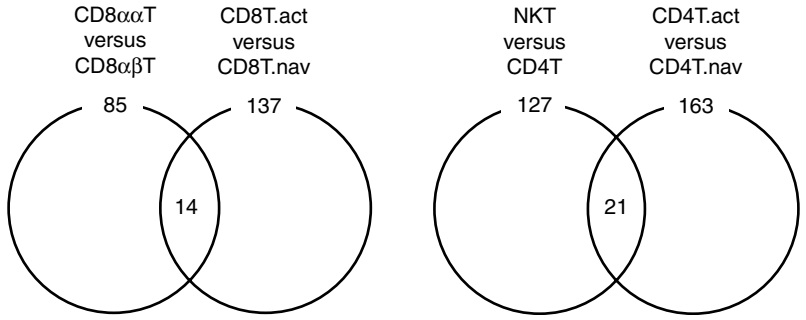


Fig. 5. Functional analysis of genes that distinguish innate-like lymphocytes. (A) Genes overexpressed (>twofold) in innate-like lymphocytes versus their adaptive counterparts or in activated versus naive CD8 and CD4 T cells were analyzed for their cellular localization using gene ontology annotations. The overexpressed genes were sorted into five major cellular components, as indicated. (B) *P* values for the distribution shown in (A). (C) Little overlap between the gene subsets that distinguish innate-like lymphocytes (relative to adaptive counterparts) and activated adaptive lymphocytes (relative to naive).

relationships. The first approach was to use principal component analysis (PCA), a mathematical dimensionality reduction tool that incorporates and summarizes the information from multiple categories (here cell types and genes) into a small number of principal components (PCs). PCA has been used in several instances to display microarray data (36, 37).

As illustrated in *Fig. 6A*, the first two components, which capture the two largest amounts of variability, reproduced the relationship seen in the hierarchical dendogram of *Fig. 4B*, with a tight positioning of the paired cell types. In contrast, for the third largest variability axis PC3 (*Fig. 6B*), all innate-like populations sorted together to the left of the graph and the adaptive populations to the right, indicating that some of the information in these data allows discrimination of innate-like versus adaptive lymphocytes as groups.

While PCA can be powerful in sorting variability in the data, it results in abstract values that are not intuitively connected to biological reality. Hence, we sought to complement this analysis by creating population-guided two-dimensional plots [reference population plots (RPPs)]. With this technique, a two-dimensional frame of reference is first created using the expression values of sets of genes that distinguish two reference populations in the data set (the *x* and *y* axes being defined by the values for the genes overexpressed in

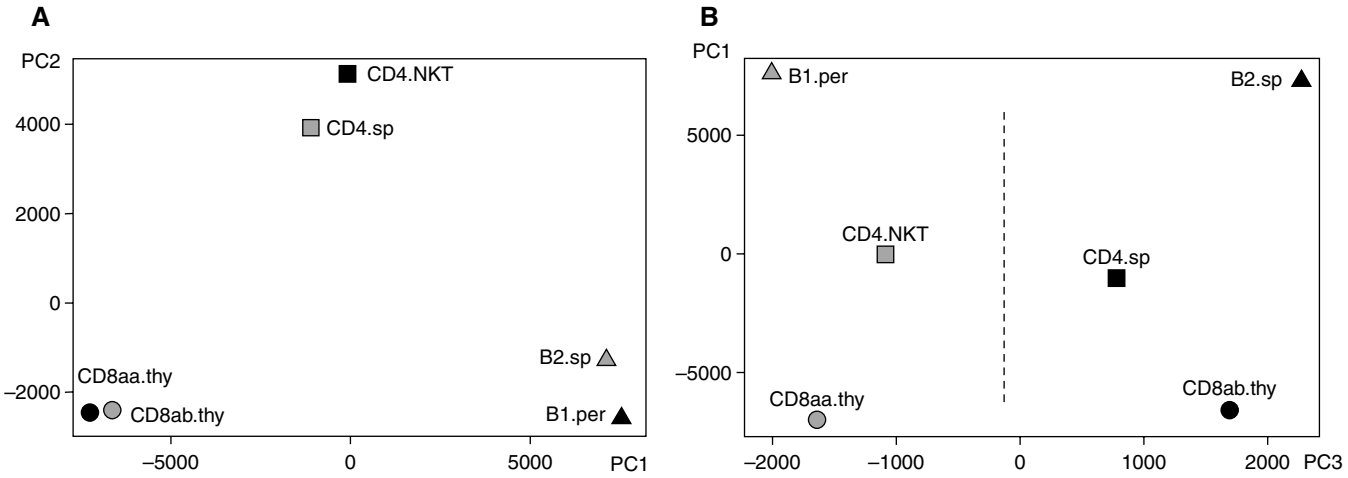
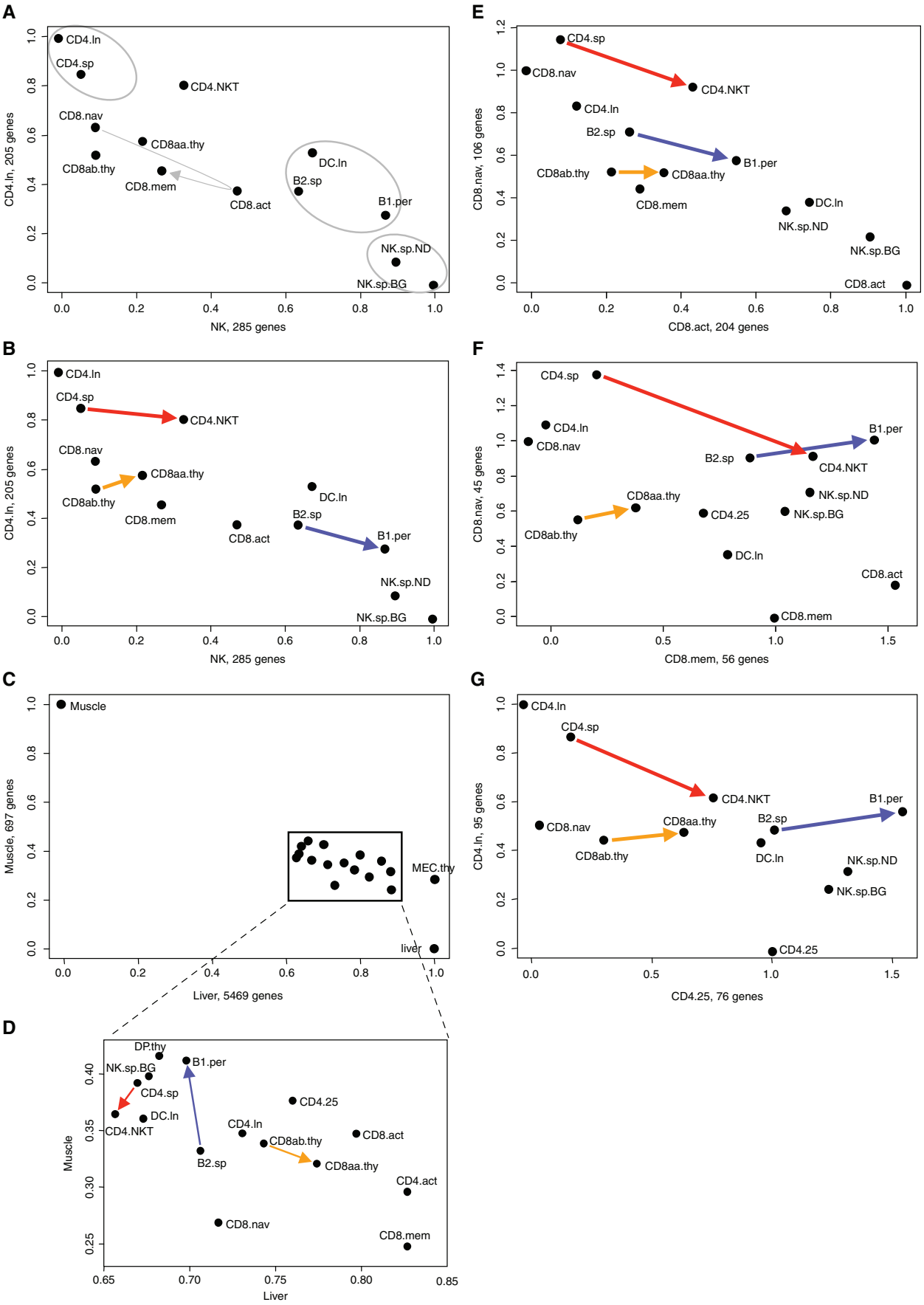


***Fig. 6.* Principal component analysis.** Principal component analysis performed on 1936 differentially expressed genes, plotting each population according to its score for components 1 and 2 (A), or components 1 and 3 (B).

one reference population relative to the other and *vice versa*). Other populations are then positioned in this space according to their aggregate expression of these distinguishing gene sets; in other words, if *x* and *y* are the populations of reference, other populations are positioned as a function of their '*x*-ness' and '*y*-ness'. Here, we took advantage of multiple data sets from various immunologically relevant populations available from previous work in our laboratory (Hyatt *et al.*, manuscript submitted; see also http://ImmGen.org). Reference genes were selected as those that best distinguish two reference populations, *x* and *y* (e.g. the two extremes of the fold-change distribution). Normalized expression values of these reference genes in the populations to be analyzed were then used to calculate the *x*- and *y*-coordinates for each population. As we were searching for commonalities between innate-like cells relative to their adaptive counterparts, this concept should be recapitulated in the two-dimensional space by an adaptive→innate vector that should be conserved across CD8αβT→CD8ααT, CD4T→NKT, and B2→B1 comparisons. 'Conserved' here refers to parallel or near-parallelism in the direction of vectors in the defined geometric space.

A first example of an RPP display can be found in *Fig. 7A*, where populations were positioned relative to gene expression values that distinguish splenic NK cells (NK.sp) and lymph node $CD4^+$ T cells (CD4.ln). As anticipated, the NK-cell populations from two different mouse strains, NK.sp.BG and NK.sp.ND, mapped close to each other, as did spleen and lymph node $CD4^+$ T cells (all circled in gray). In addition, we observed interesting positioning of $CD8^+$ naive, $CD8^+$ activated, and $CD8^+$ memory cells, consistent with the notion that activated effector $CD8^+$ T cells share characteristics with NK cells (38, 39), and the intermediate nature of the $CD8^+$ memory T cells (40). For the six populations of interest in the present study, the two members of each pair were connected by colored arrows in order to highlight the respective adaptive→innate vectors (*Fig. 7B*). The directions of the three vectors pointed from left to right, in the direction of maximum NK-ness. While one might have expected that the CD4.sp→CD4.NKT would display NK-ness, it was interesting that CD8αβT→CD8ααT and B2→B1 vectors pointed in the same direction, indicating that CD8ααT and B1 cells exhibit some degree of NK-ness relative to CD8αβT and B2 cells, respectively. As a control, the same RPP analysis was run using immunologically irrelevant cells as references, on the grounds that the liver-ness and muscle-ness of immune cell populations is probably meaningless (*Fig. 7C*). One immediately notes a two-dimensional squashing of the positions of the immunologically relevant cell populations (*Fig. 7C, black rectangle*) relative to those of liver and muscle cells. In the expanded rectangle (*Fig. 7D*), the locations of the immune cell populations were scattered, with the three adaptive→innate vectors devoid of any shared directionality. The relevance of the RPP profiles of *Fig. 7B* was also confirmed by RPP plots with randomized data sets; here again, the grouping of populations and the shared vector directionality were lost (data not shown).

Several studies have suggested that innate-like cells share some characteristics with the activated and memory states of conventional lymphocytes (13, 17, 41–43). We explored this concept with RPP by taking activated or memory versus naive $CD8^+$ T cells as RPP reference populations (*Fig. 7E,F*). All three adaptive→innate vectors pointed to the right, indicating that innate-like lymphocytes are more enriched in transcripts characteristic of activated and memory cells. Interestingly, true


A
CD4.ln
CD4.sp
CD4.NKT
CD8.nav
CD8aa.thy
CD8ab.thy
CD8.mem
CD8.act
DC.ln
B2.sp
B1.per
NK.sp.ND
NK.sp.BG
CD4.ln, 205 genes
NK, 285 genes
B
CD4.ln
CD4.sp
CD4.NKT
CD8.nav
CD8aa.thy
CD8ab.thy
CD8.mem
CD8.act
DC.ln
B2.sp
B1.per
NK.sp.ND
NK.sp.BG
CD4.ln, 205 genes
NK, 285 genes
C
Muscle
MEC.thy
liver
Muscle, 697 genes
Liver, 5469 genes
D
DP.thy
NK.sp.BG
B1.per
CD4.sp
CD4.25
DC.ln
CD4.NKT
CD4.ln
CD8ab.thy
CD8.act
B2.sp
CD8aa.thy
CD4.act
CD8.nav
CD8.mem
Muscle
Liver
E
CD4.sp
CD8.nav
CD4.NKT
CD4.ln
B2.sp
B1.per
CD8ab.thy
CD8aa.thy
CD8.mem
DC.ln
NK.sp.ND
NK.sp.BG
CD8.act
CD8.nav, 106 genes
CD8.act, 204 genes
F
CD4.sp
CD4.ln
CD8.nav
B1.per
B2.sp
CD4.NKT
CD8aa.thy
NK.sp.ND
CD8ab.thy
CD4.25
NK.sp.BG
DC.ln
CD8.act
CD8.mem
CD8.nav, 45 genes
CD8.mem, 56 genes
G
CD4.ln
CD4.sp
CD4.NKT
B1.per
CD8.nav
CD8aa.thy
B2.sp
CD8ab.thy
DC.ln
NK.sp.ND
NK.sp.BG
CD4.25
CD4.ln, 95 genes
CD4.25, 76 genes

innate cell types (NKs and DCs) also positioned close to activated or memory $CD8^+$ T cells in these comparisons.

The RPP plot can also be used to reveal unknown characteristics of innate lymphocytes. We have plotted CD4.ln versus CD4.25 [the latter data set from $CD4^+CD25^+$ regulatory T cells (44)] to represent the 'regulatory-ness' in each cell type (*Fig. 7G*). Interestingly, innate lymphocytes expressed more of a regulatory signature than did their adaptive counterparts, suggesting some degree of immune regulatory capacity in these cells, as already known for NKT cells (45–47) and also recently reported for CD8ααT cells (48). Quite unexpectedly, B1 cells showed an even stronger regulatory signature than did $CD4^+CD25^+$ regulatory T cells. This brings to mind the reported regulatory capacity of certain B-cell subsets (49, 50). Thus, the RPPs emphasize the versatile nature of innate lymphocytes – their NK-, activated-, memory-, and regulatory-like nature – and suggest the existence of a common signature shared by innate immune system players.

A shared innate signature distinguishes innate-like from adaptive lymphocytes

The PCA and RPP analyses suggested the existence of gene expression common to the three innate-like cell types. In order to identify the primary genes underlying the commonality, we followed two different bioinformatic approaches. In the first strategy, we simply selected genes differentially expressed >1.5-fold in each innate-like/adaptative pair and represented those shared between two or all three of the cell types on Venn diagrams (*Fig. 8A*). This procedure yielded 19 and five genes overexpressed and underexpressed, respectively, in innate-like cell types. These numbers are greater than would have occurred by chance as verified by using a randomized data group generated by permutation of the results: from the number of genes present on each list, one could expect between nine and 15 of them to be shared between any two comparisons, and one or less between all three (*Fig. 8A, right*). CD8ααT and either B1 or NKT cells also showed numbers of shared overexpressed genes higher than background levels (63 and 60, respectively). In contrast, the B1/NKT overexpression and all of the innate underexpression pairwise comparisons failed to show significant commonality.

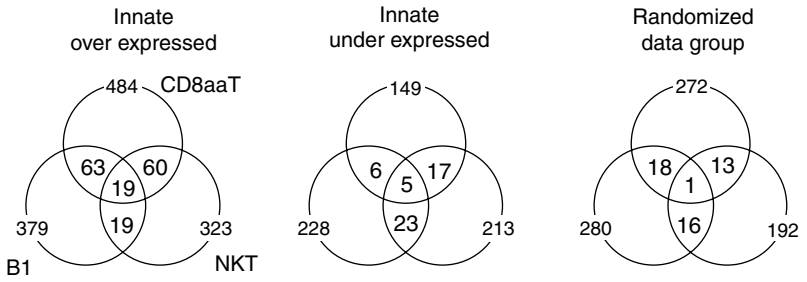
The identities of the genes overexpressed and underexpressed in all three innate-like cell types are listed in *Fig. 8B*, and some appear immediately relevant in the context of trafficking by or signaling in innate immune cells. It is important to note that these loci do not merely represent a collection of genes characteristic of cell activation: only a fraction of them were influenced by activation in conventional $CD4^+$ or $CD8^+$ T cells (*Fig. 8B, right*).

At this stage, we felt it important to validate the significance of the gene expression changes indicated by the microarray data. Independent pools of cells were sorted, RNA was prepared, and transcript levels of a number of the genes of the innate signature were measured by quantitative RT-PCR. The results, displayed in *Fig. 8C*, fully confirmed the microarray data, showing a higher level of expression for the innate-like member of the pair in each instance.

A complementary second approach to defining an innate signature was based on gene ontology annotations. We reasoned that proteins with a shared function would not necessarily be the same in all three innate-like cells but that a common or related molecular process might be performed by different members of the same family in different cells. Hence, restricting the analysis to the commonly overexpressed gene identities, as by filtering on common fold-change, might overlook significant genes or functions. Therefore, we sought the common molecular pathways overexpressed in the three innate cell types, using GO annotations attached to the microarrays. Genes that overexpressed >1.5-fold in any one of the

***Fig. 7.* Innate-like populations share global characteristics with innate NK cells.** Reference population plots (RPPs) were constructed. (A) Validation of the RPP taking NK-ness and CD4-ness on x- and y-axis, respectively. NK-ness is defined by genes overexpressed (fold-change >2) in spleen NK cells relative to $CD4^+$ lymph node T cells; CD4-ness is defined by genes overexpressed in $CD4^+$ cells in the same comparison. Thus, the lower right corner represents full NK-ness with minimum CD4-ness. The relative coordinates for each population are calculated for these two traits and plotted on the two-dimensional panel. Similar populations map close to one another, as indicated by the gray ellipses. Transition of $CD8^+$ T cells after activation is indicated by the looping arrow. (B) The three adaptive→innate-like pairings were connected by colored arrows in the same plot. (C and D) Control RPP with liver and muscle-specific genes; the tight grouping in (C) is expanded in (D). (E and F) Activation or memory character of innate lymphocytes was assessed using expression profiles from naive, activated, or memory $CD8^+$ T cells as references. (G) 'Regulatory' character of innate-like lymphocytes compared to their adaptive counterparts, using data sets from $CD4^+CD25^+$ regulatory T cells as reference. The populations shown are: CD4.sp, CD4 T cell from spleen; CD4.ln, CD4 T cells from lymph node; CD4.NKT, NKT cell from spleen; CD8.nav, naïve CD8 T cell; CD8.act, activated CD8 T cell; CD8.mem, memory CD8 T cell; CD8aa.thy, CD8αα thymocytes from FTOC; CD8ab.thy, CD8αβ thymocytes from FTOC; B1.per, B-1 B cell from peritoneal cavity; B2.sp., B-2 B cell from spleen; DC.ln, dendritic cell from lymph node; NK.sp.ND, NK cell from NOD mouse spleen; NK.sp.BG, NK cell from B6g7 mouse spleen, DP.thy, $CD4^+CD8^+$ double-positive thymocyte; CD4.25, $CD4^+CD25^+$ regulatory T cell.

A



B Innate overexpressed genes

Average FC	CD8aaT versus CD8abT	B-1 versus B-2	NKT versus CD4T	Affy ID	Entrez GeneI	Gene symbol	Gene alias	Gene description	Function/annotation	CD8T.act. versus CD8T.nav	CD4T.act versus CD4T.nav
8.01	5.47	1.99	16.56	161968_f_at	12774	Ccr5	Cmkbr5	Chemokine (C-C) receptor 5	Chemokine receptor	5.08	1.19
6.15	2.09	6.91	9.47	97949_at	14190	Fgl2	musfiblp	Fibrinogen-like protein 2	Prothrombinase, immunomodulatory activity	2.27	1.05
4.87	4.72	8.13	1.75	160255_at	66395	Ahnak	desmoyokin	RIKEN cDNA 2310047C17 gene	700-kDa phosphoprotein	5.81	3.04
3.80	1.72	7.66	2.01	92356_at	19260	Ptpn22	PEP, Ptpn8, 70zpep	Tyrosine phosphatase, non-receptor type 22	Regulation of effector/memory T cells	0.91	1.36
3.17	2.18	4.08	3.25	103560_at	70082	EST		RIKEN cDNA 2210402C18 gene	LysM domain bacterial cell wall degradation	0.91	0.77
2.88	1.98	4.34	2.31	98025_at	14017	Evi2	Evi-2	Ecotropic viral integration site 2	Leucine-zipper transmembrane protein	1.65	3.49
2.71	2.10	2.73	3.31	100397_at	22177	Tyrobp	Dap12, KARAP	TYRO protein tyrosine kinase-binding protein	ITAM-signalling adaptor protein	1.89	1.01
2.37	3.09	1.79	2.23	97885_at	65963	Lr8	Clast1	RIKEN cDNA 1810009M01 gene	Putative transmembrane protein	0.99	0.94
2.22	1.67	2.98	2.02	98026_g_at	14017	Evi2	Evi-2	Ecotropic viral integration site 2	Leucine-zipper transmembrane protein	1.60	2.74
2.22	2.59	1.68	2.41	104252_at	101757	EST	AU020206	Expressed sequence AU020206		1.87	1.05
2.21	2.98	1.67	1.99	93043_at	20320	Sdfr1	SDR1	Stromal cell-derived factor receptor 1	Immunoglobulin superfamily protein	1.72	1.69
1.94	2.23	1.54	2.06	96605_at	66058	Keg2	KEG2	RIKEN cDNA 0610011I04 gene	Putative transmembrane protein	1.00	0.97
1.90	1.57	1.55	2.58	96862_at	104725	EST		RIKEN cDNA 1110002B05 gene	Integral to membrane	1.29	1.08
1.88	2.19	1.65	1.81	93078_at	110454	Ly6a	Sca-1, TAP	Lymphocyte antigen 6 complex, locus A	Hematopoietic stem-cell antigen	4.65	2.03
1.79	2.04	1.72	1.60	160205_f_at	29864	Rnf11		Ring finger protein 11	Ubiquitin-mediated proteolysis	1.07	1.18
1.70	1.52	1.86	1.73	101060_at	14827	Grp58	PLCa, Erp, ERp57	Glucose-regulated protein, 58 kDa	Thiol oxidoreductase	1.48	2.01
1.68	1.65	1.88	1.52	98600_at	20195	S100a11	Calizzarin, Emap1	S100 calcium-binding protein A11	Ca(2+)-induced growth inhibition	2.21	1.18
7.50	1.78	18.59	2.15	92780_f_at	NA	NA		NA		0.77	0.36
5.49	1.54	13.13	1.79	92779_f_at	NA	NA		NA		0.91	0.62

Innate under-expressed genes

Average FC	CD8aaT versus CD8abT	B-1 versus B-2	NKT versus CD4T	Affy ID	Entrez GeneID	Gene symbol	Gene alias	Gene description	Function/annotation	CD8T.act versus CD8T.nav	CD4T.act versus CD4T.nav
0.43	0.54	0.53	0.20	102838_at	20343	Sell	CD62L, Ly-22	Selectin, lymphocyte	Lymph node homing	0.60	1.30
0.45	0.54	0.58	0.23	104443_at	12775	Ccr7	Cmkbr7	Chemokine (C-C) receptor 7	Second lymphoid organ homing	0.34	0.78
0.45	0.65	0.46	0.24	160150_f_at	71994	Cnn3		Calponin 3, acidic	Actin /calmodulin binding	0.59	0.39
0.49	0.65	0.48	0.35	103596_at	13139	Dgka	Dagk1	Diacylglycerol kinase alpha (80 kDa)	Attenuates Ras signaling	0.27	0.59
0.36	0.56	0.10	0.43	96791_at	76566	EST		RIKEN cDNA 1500005K14 gene		0.20	0.38

C

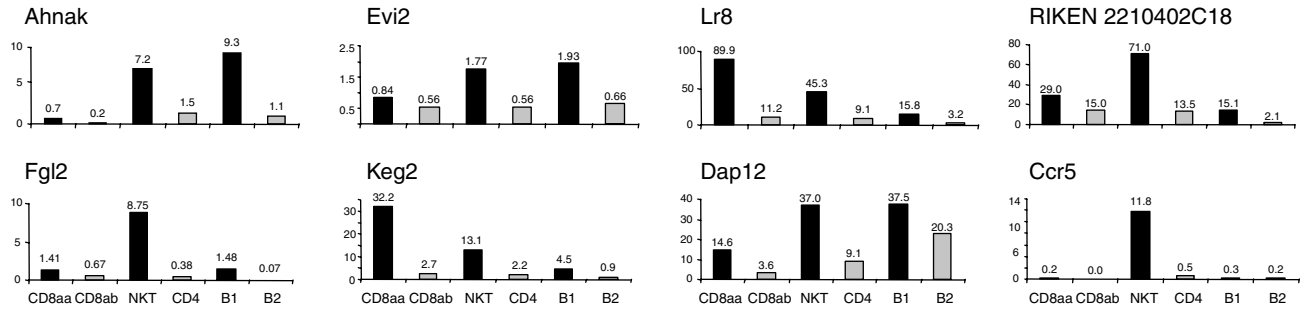


***Fig. 8.* A signature shared by innate-like lymphocytes identified by expression fold-change.** (A) Sets of genes overexpressed or underexpressed (fold-change >1.5-fold) in innate-like lymphocytes relative to their adaptive counterparts were selected and analyzed for overlap. The numbers of shared genes are shown in the Venn diagram. (B) Identities of 19 and five genes overexpressed and underexpressed, respectively, in three innate lymphocytes are shown. (C) Validation of the microarray data by quantitative RT-PCR, performed on selected genes from the list. The expression of *HPRT* gene is used as reference.

innate cell types were selected, and the corresponding GO 'molecular function' annotations were extracted. The union of these GO terms, processed with GO-Surfer (51), was displayed as an ontology tree (*Fig. 9A*), each dot representing a functional GO term with its connections to other terms. Within this framework, the significance of these occurrences for each innate-like/adaptive combination was calculated (relating the frequency with which it was relative to the frequency of the GO attribute for the whole array, Fisher's exact test); we then identified those GO terms that were significantly overrepresented in two or three of the innate-like cells (using Fisher's *P*-value <0.05 as a cutoff). These commonalities are numbered and color coded on the tree of *Fig. 9A*. Several of the terms directly relate to immunological functions such as antigen binding, toxin activity, and antimicrobial activity. The genes uncovered in this GO-based

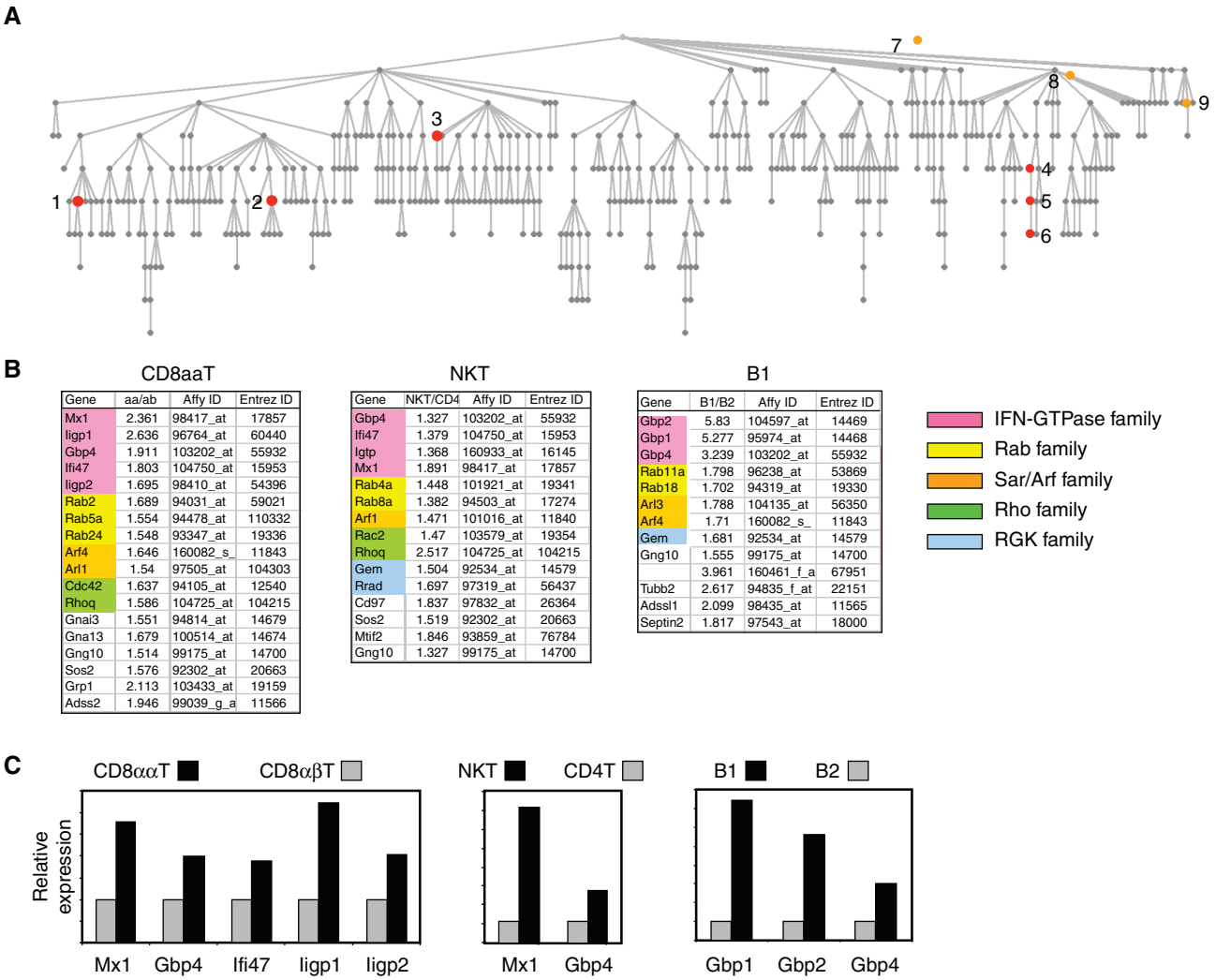


***Fig. 9.* A signature shared by innate-like lymphocytes identified by gene ontology analysis.** (A) Genes overexpressed (fold-change >1.5-fold) in any of the three innate lymphocytes were combined, and the corresponding GO 'molecular function' annotations were displayed as hierarchical tree. Of 1380 GO terms existing for the whole microarray, 449 terms appear in the tree. Terms that are statistically overrepresented ($P < 0.05$, no correction for multiple sampling) in two (orange) or three (red) innate cell types were numbered. Nodes #1, 4, 5, and 6 represent 'GTP-binding' and subclassified functions. (B) Lists of genes with function corresponding to nodes 4, 5, and 6 that are overexpressed in each innate cell type. The genes are color coded according to their subfamilies: IFN-inducible GTPase (purple); Rab family (yellow); Sar1/Arf family (orange); Rho family (green); RGK family (blue). (C) Confirmation by quantitative RT-PCR for the IFN-inducible GTPase family genes in the three innate-like versus adaptative lymphocyte pairs.

search did not overlap with those detected in the expression-based filtering (*Fig.* 8B), indicating that the two approaches are orthogonal and complementary.

Interestingly, three of these overrepresented GO terms are hierarchically related in the ontology. Numbers 4, 5, and 6 correspond to 'guanyl nucleotide binding' function, and number 1 also has a related function. The corresponding genes are summarized in *Fig.* 9B. They contain a surprisingly high number of loci encoding interferon-inducible guanosine triphosphatase (IFN-GTPase) family members, an emerging family of proteins involved in newly recognized forms of pathogen clearance (52, 53). The occurrence of this gene family is significant across all three pairwise comparisons (P values: 0.0008, 0.015, and 0.014 for CD8ααT, NKT, and B1, respectively, Fisher's exact test), and the differential expression was confirmed by quantitative RT-PCR (*Fig.* 9C). However, the relevance of the cellular process emerges more from the expression profiles of a broader set of genes than those encoding IFN-GTPases. Genes specifying members of the Rab, Arf, and Rho families of small G proteins, which are structurally and functionally related to IFN-GTPase, were also overexpressed in innate-like lymphocytes (*Fig.* 9B). These small G proteins are involved in intracellular vesicle/vacuole transport, as are several IFN-GTPases (54). It is also worth

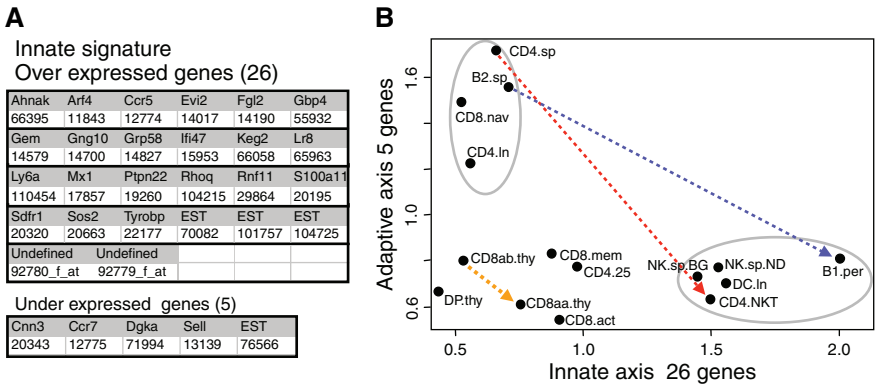


***Fig. 10.* General validity of the innate versus adaptive signature.** (A) Genes identified by expression differential and GO analyses were combined to create the 'innate signature' list. The gene names and Entrez IDs are given. For Undefineds, affyIDs are provided. (B) Reference gene plot was created using the 26 and five signature genes from the list to define 'innateness' and 'adaptiveness'. Coordinates for various cell types were calculated as in *Fig.* 7. Innate-like and adaptive lymphocytes partition on the lower right and upper left corners, respectively, as indicated by the ellipses. Population descriptions were as in *Fig.* 7.

pointing out that no genes encoding members of the Ras or Ran family of the small G proteins appeared on the list, further bolstering the specificity of induction of the small G proteins (note that *Gem* and *Rrad* belong to the RGK family rather than to the Ras family).

Broad significance of the innate signature

We then integrated the genes identified by the two bioinformatic approaches into a joint innate signature; genes identified by expression fold-change in *Fig.* 8B and those shared in at least two innate cell types in the ontology-based screen of *Fig.* 9B were combined (*Fig.* 10A). The validity of this list was tested in a population plot display, where the gene sets distilled above were used to define 'innate-ness' and 'adaptive-ness' on the *x*- and *y*-axes, that is, the ordinate and abscissa values derived from the expression of genes over-expressed and underexpressed in innate-like lymphocytes, respectively (*Fig.* 10B). As expected because they formed the 'training set' for this discrimination, innate-like lymphocytes partitioned at the bottom right of the plot, distanced from their adaptative counterparts on the top left. This segregation extended to other conventional T- and B-cell populations, which clustered in the upper left corner of the chart, while true innate cells, e.g. NK cells and dendritic cells, clustered in the opposite corner. Thus, the innate/adaptive distinction, derived from the analysis of innate-like lymphocyte particularities, extends more broadly to non-lymphoid innate cells.

The distinction between CD8αα and CD8αβ thymocytes was far less marked than between the B1 and B2 or NKT and CD4T pairs (*Fig.* 10B). The former cell types corresponded to very recently matured populations isolated from fetal thymic organ cultures, while the latter were purified from peripheral lymphoid organs. The difference might thus reflect an accentuation of the innate signature once cells colonize the peripheral lymphoid organs. This notion was confirmed, in preliminary analyses, by showing that the differences were more marked when comparing gut CD8αα IELs with CD8αβ lymph node cells (data not shown). Similarly, $CD4^+CD8^+$ double positive thymocytes (DP.thy) mapped to the 'undifferentiated' lower left corner of *Fig.* 10B. This position is consistent with the view that innate-like and adaptive T lymphocytes mature from the same pool of DP thymocytes, which one might expect to exhibit 'unbifurcated' characteristics. Finally, activated and memory $CD8^+$ lymphocytes mapped to an intermediate position, as did $CD4^+CD25^+$ T regulatory cells (CD4.25), consistent with the acquisition of innate-like characteristics by activated lymphocytes already mentioned above.

The innate signature

Here, we sought to exploit data sets of gene transcript profiles to identify gene expression signatures that were common to innate-like lymphocytes and to distinguish them from their close cousins of the adaptative immune system. Indeed, statistically significant gene sets were revealed by complementary bioinformatic analyses, and these were shared by more stereotyped members of the innate immune system, such as NK or dendritic cells. The identification of the innate signature relied on two complementary bioinformatic tools, searching directly for shared expression differences and also for functional families identified through GO annotations. Each approach is imperfect: with expression filtering, experimental variation will eliminate genes that did not meet the

fold-change threshold set between the three population pairs, and functions that rely on close homologs in a multigene family will also be missed. GO-based identification relies on statistically overrepresented gene families and will eliminate single genes, those only identified by GO terms that occur frequently overall, or those that fall within the gaps of the incomplete GO annotation.

Several conclusions emerge from the existence of a signature common to innate-like lymphocytes. Firstly, these observations confirm the notion that innate immune system cells are distinguished by more than the simple absence of recombination activating gene (RAG)-mediated rearrangement of antigen receptor genes (11, 13). Indeed, it is interesting that the characteristic signature entails primarily genes overexpressed in innate-like cells. Thus, these cells have supplementary capabilities and are not just 'poor relatives' of adaptive cells. (Admittedly, the differentiation stages analyzed here concerned essentially mature cells and therefore missed the transcripts directly linked to DNA rearrangement in immature precursors, such as RAG or TdT). Secondly, this signature provides a new annotation for genomic analyses of microarray data sets. Such information stands to be particularly informative in genomic analyses of hematopoietic tissues, where microarray data have proven instrumental in classifying disease subsets and predicting prognosis (55, 56). Lastly, the innate signature reported here is an interesting source of candidate genes to explore in dissecting fundamental functions or characteristics unique to innate immune system cell types. As detailed below, mutation studies of some of these genes have demonstrated their direct roles in the homeostasis and function of innate cell types.

A commonly upregulated molecular pathway in innate immune cells

An intriguing gene family was highlighted in this study. The IFN-GTPase family is the newly discovered GTPase family, whose impact on mammalian host defense was only very recently demonstrated, particularly in the context of intracellular bacteria and protozoa (57–59). This family comprises four subgroups: the p47 family, p65 guanylate-binding protein (GBP) family, Mx family, and the very large inducible GTPases. Each subgroup has a distinct scope in its antipathogenic effect. The p47 family defends against vacuolarized bacterial, viral, and protozoan pathogens, whereas the GBP and Mx families show mainly antiviral activities (52, 53).

These GTPases are normally expressed at low levels, but after infection, they accumulate to high levels in nearly all hematopoietic and non-hematopoietic tissues (58, 60, 61) in an IFN-dependent manner (60, 62). Here, we documented overexpression of IFN-GTPases in all innate-like lymphocytes. Moreover, we observed higher expression of GBP and Mx family genes in other innate immune system population such as dendritic cells and NK cells. One might argue that the higher levels of IFN-GTPase expression in innate cells merely reflects higher IFN concentrations in the tissue(s) they derive from. However, this explanation seems unlikely, because CD8$\alpha\alpha$T/CD8$\alpha\beta$T and NKT/CD4T cells were prepared from the same organs (fetal thymic organ cultures for one pair, spleens for the other). There was also some cell-type specificity to the expression of the IFN-GTPases; for instance, the GBP family seemed restricted to B1 cells, a finding not readily compatible with a generalized triggering of the IFN pathway.

Knockout studies of the IFN-GTPase have demonstrated the involvement of these genes in innate immune-mediated clearance of intracellular bacteria and protozoa (57–59). However, the mechanism of action of the IFN-GTPases in the resistance to such pathogens is not yet clearly understood. GTPases of the p47 family are known to be transported to pathogen-containing phagosomes/vacuoles after infection, where they modulate the maturation of these subcellular structures (52, 53); targeting of Gbp2 to intracellular vesicle-like structures has also been reported (63). In this context, the parallel induction of Rho/Rab/Arf family members as part of the innate signature is quite congruent, as these small G proteins are involved in intracellular vacuole transportations and cytoskeletal reorganizations (54), and there is precedence for such GTPases to have a proactive role in bactericidal processes, in particular, the Rab proteins (52, 64–66). The significance of these molecular pathways in innate-like lymphocytes remains to be elucidated, but one might speculate that they represent a new type of intracellular recognition system, detecting 'altered self' or 'foreign non-self' in the contents of phagocytic vesicles (53, 67). The endocytic and intracellular trafficking machinery is one of the most ancient cellular systems, extending back to amoebae, and one that would require aggressive monitoring against pathogen invasion. It is perhaps the evolutionary remnants of such a primitive system that we are detecting as an innate signature in rodents of today.

References

1. Ehrlich P. The Croonian lecture: on immunity. Proc R Soc London 1900;**B 66**:424–448.
2. Hozumi N, Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc Natl Acad Sci USA 1976;**73**:3628–3632.
3. Medzhitov R, Janeway CA Jr. Innate immunity: the virtues of a nonclonal system of recognition. Cell 1997;**91**:295–298.
4. Kasahara M, Suzuki T, Pasquier LD. On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol 2004;**25**:105–111.
5. Janeway C, ed. Immunobiology, 5th edn. New York: Garland Publishing, 2001.
6. Bernstein RM, Schluter SF, Bernstein H, Marchalonis JJ. Primordial emergence of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates homology to microbial integrases. Proc Natl Acad Sci USA 1996;**93**:9454–9459.
7. van den Berg TK, Yoder JA, Litman GW. On the origins of adaptive immunity: innate immune receptors join the tale. Trends Immunol 2004;**25**:11–16.
8. Cannon JP, Haire RN, Rast JP, Litman GW. The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms. Immunol Rev 2004;**200**:12–22.
9. Flajnik MF, Du PL. Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol 2004;**25**:640–644.
10. Janeway CA Jr, Medzhitov R. Innate immune recognition. Annu Rev Immunol 2002;**20**: 197–216.
11. Dempsey PW, Vaidya SA, Cheng G. The art of war: innate and adaptive immune responses. Cell Mol Life Sci 2003;**60**:2604–2621.
12. William PE, ed. Fundemental Immunology, 5th edn. Philadelphia: Lippincott Williams & Wilkins, 2003.
13. Bendelac A, Bonneville M, Kearney JF. Autoreactivity by design: innate B and T lymphocytes. Nat Rev Immunol 2001;**1**: 177–186.
14. Guy-Grand D, Vassalli P. Immunology. Tracing an orphan's genealogy. Science 2004;**305**:185–187.
15. Kearney JF. Innate-like B cells. Springer Semin Immunopathol 2005;**26**:377–383.
16. Shaffer AL, et al. Signatures of the immune response. Immunity 2001;**15**:375–385.
17. Yamagata T, Mathis D, Benoist C. Self-reactivity in thymic double-positive cells commits cells to a CD8 alpha alpha lineage with characteristics of innate immune cells. Nat Immunol 2004;**5**:597–605.
18. Rocha B, von Boehmer H, Guy-Grand D. Selection of intraepithelial lymphocytes with CD8 alpha/alpha co-receptors by self-antigen in the murine gut. Proc Natl Acad Sci USA 1992;**89**:5336–5340.
19. Cruz D, Sydora BC, Hetzel K, Yakoub G, Kronenberg M, Cheroutre H. An opposite pattern of selection of a single T cell antigen receptor in the thymus and among intra-epithelial lymphocytes. J Exp Med 1998;**188**:255–265.
20. Leishman AJ, et al. Precursors of functional MHC class I- or class II-restricted CD8alphaalpha(+) T cells are positively selected in the thymus by agonist self-peptides. Immunity 2002;**16**:355–364.
21. Moretta A, et al. Activating receptors and coreceptors involved in human natural killer cell-mediated cytolysis. Annu Rev Immunol 2001;**19**:197–223.
22. Luster AD. The role of chemokines in linking innate and adaptive immunity. Curr Opin Immunol 2002;**14**:129–135.
23. Vorbach C, Harrison R, Capecchi MR. Xanthine oxidoreductase is central to the evolution and function of the innate immune system. Trends Immunol 2003;**24**:512–517.
24. Hacker C, et al. Transcriptional profiling identifies Id2 function in dendritic cell development. Nat Immunol 2003;**4**: 380–386.
25. Yokota Y, et al. Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. Nature 1999;**397**:702–706.
26. Gao GF, Rao Z, Bell JI. Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. Trends Immunol 2002;**23**:408–413.
27. Bain G, et al. E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. Mol Cell Biol 1997;**17**: 4782–4791.
28. Berland R, Wortis HH. Origins and functions of B-1 cells with notes on the role of CD5. Annu Rev Immunol 2002;**20**:253–300.
29. Martin F, Kearney JF. B1 cells: similarities and differences with other B cell subsets. Curr Opin Immunol 2001;**13**:195–201.
30. Wortis HH, Berland R. Cutting edge commentary: origins of B-1 cells. J Immunol 2001;**166**:2163–2166.
31. Wilson SB, Byrne MC. Gene expression in NKT cells: defining a functionally distinct CD1d-restricted T cell subset. Curr Opin Immunol 2001;**13**:555–561.
32. Anderson MS, et al. Projection of an immunological self-shadow within the thymus by the aire protein. Science 2002;**298**: 1395–1401.
33. Taniguchi M, Harada M, Kojo S, Nakayama T, Wakao H. The regulatory role of Valpha14 NKT cells in innate and acquired immune response. Annu Rev Immunol 2003;**21**: 483–513.
34. Park SH, Benlagha K, Lee D, Balish E, Bendelac A. Unaltered phenotype, tissue distribution and function of Valpha14(+) NKT cells in germ-free mice. Eur J Immunol 2000;**30**:620–625.
35. Goldrath AW, Luckey CJ, Park R, Benoist C, Mathis D. The molecular program induced in T cells undergoing homeostatic proliferation. Proc Natl Acad Sci USA 2004;**101**:16885–16890.
36. Ricciardi-Castagnoli P, Granucci F. Opinion: interpretation of the complexity of innate immune responses by functional genomics. Nat Rev Immunol 2002;**2**:881–889.
37. Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics 2001;**17**:763–774.
38. Kaech SM, Hemby S, Kersh E, Ahmed R. Molecular and functional profiling of memory CD8 T cell differentiation. Cell 2002;**111**:837–851.
39. McMahon CW, Raulet DH. Expression and function of NK cell receptors in CD8+ T cells. Curr Opin Immunol 2001;**13**:465–470.
40. Holmes S, He M, Xu T, Lee PP. Memory T cells have gene expression patterns intermediate between naive and effector. Proc Natl Acad Sci USA 2005;**102**:5519–5523.
41. Hayday A, Theodoridis E, Ramsburg E, Shires J. Intraepithelial lymphocytes: exploring the Third Way in immunology. Nat Immunol 2001;**2**:997–1003.
42. Martin F, Kearney JF. B-cell subsets and the mature preimmune repertoire. Marginal zone and B1 B cells as part of a 'natural immune memory'. Immunol Rev 2000;**175**:70–79.
43. Brigl M, Bry L, Kent SC, Gumperz JE, Brenner MB. Mechanism of CD1d-restricted natural killer T cell activation during microbial infection. Nat Immunol 2003;**4**:1230–1237.
44. Herman AE, Freeman GJ, Mathis D, Benoist C. CD4+CD25+ T regulatory cells dependent on ICOS promote regulation of effector cells in the prediabetic lesion. J Exp Med 2004;**199**:1479–1489.
45. Singh AK, et al. Natural killer T cell activation protects mice against experimental autoimmune encephalomyelitis. J Exp Med 2001; **194**: 1801–1811.
46. Hong S, et al. The natural killer T-cell ligand alpha-galactosylceramide prevents autoimmune diabetes in non-obese diabetic mice. Nat Med 2001;**7**:1052–1056.
47. Sharif S, et al. Activation of natural killer T cells by alpha-galactosylceramide treatment prevents the onset and recurrence of autoimmune Type 1 diabetes. Nat Med 2001;**7**:1057–1062.
48. Poussier P, Ning T, Banerjee D, Julius M. A unique subset of self-specific intraintestinal T cells maintains gut integrity. J Exp Med 2002;**195**:1491–1497.

49. Mizoguchi A, Mizoguchi E, Takedatsu H, Blumberg RS, Bhan AK. Chronic intestinal inflammatory condition generates IL-10-producing regulatory B cell subset characterized by CD1d upregulation. Immunity 2002;**16**:219–230.
50. Mauri C, Gray D, Mushtaq N, Londei M. Prevention of arthritis by interleukin 10-producing B cells. J Exp Med 2003;**197**: 489–501.
51. Zhong S, Storch KF, Lipan O, Kao MC, Weitz CJ, Wong WH. GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in gene ontologytrade mark space. Appl Bioinformatics 2004;**3**: 261–264.
52. Taylor GA, Feng CG, Sher A. p47 GTPases: regulators of immunity to intracellular pathogens. Nat Rev Immunol 2004;**4**: 100–109.
53. MacMicking JD. IFN-inducible GTPases and immunity to intracellular pathogens. Trends Immunol 2004;**25**:601–609.
54. Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. Physiol Rev 2001;**81**: 153–208.
55. Bullinger L, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;**350**:1605–1616.
56. Valk PJ, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;**350**:1617–1628.
57. MacMicking JD, Taylor GA, McKinney JD. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science 2003;**302**:654–659.
58. Taylor GA, et al. Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IGTP. Proc Natl Acad Sci USA 2000;**97**:751–755.
59. Collazo CM, et al. Inactivation of LRG-47 and IRG-47 reveals a family of interferon gamma-inducible genes with essential, pathogen-specific roles in resistance to infection. J Exp Med 2001;**194**:181–188.
60. Boehm U, et al. Two families of GTPases dominate the complex cellular response to IFN-gamma. J Immunol 1998;**161**:6715–6723.
61. Zerrahn J, Schaible UE, Brinkmann V, Guhlich U, Kaufmann SH. The IFN-inducible Golgi- and endoplasmic reticulum- associated 47-kDa GTPase IIGP is transiently expressed during listeriosis. J Immunol 2002;**168**:3428–3436.
62. Haller O, Kochs G. Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. Traffic 2002;**3**:710–717.
63. Gorbacheva VY, Lindner D, Sen GC, Vestal DJ. The interferon (IFN)-induced GTPase, mGBP-2. Role in IFN-gamma-induced murine fibroblast proliferation. J Biol Chem 2002;**277**:6080–6087.
64. Meresse S, Steele-Mortimer O, Finlay BB, Gorvel JP. The rab7 GTPase controls the maturation of Salmonella typhimurium-containing vacuoles in HeLa cells. EMBO J 1999;**18**:4394–4403.
65. Alvarez-Dominguez C, Barbieri AM, Beron W, Wandinger-Ness A, Stahl PD. Phagocytosed live Listeria monocytogenes influences Rab5-regulated in vitro phagosome-endosome fusion. J Biol Chem 1996;**271**:13834–13843.
66. Scianimanico S, Desrosiers M, Dermine JF, Meresse S, Descoteaux A, Desjardins M. Impaired recruitment of the small GTPase rab7 correlates with the inhibition of phagosome maturation by Leishmania donovani promastigotes. Cell Microbiol 1999;**1**:19–32.
67. Halonen SK, Taylor GA, Weiss LM. Gamma interferon-induced inhibition of Toxoplasma gondii in astrocytes is mediated by IGTP. Infect Immun 2001;**69**:5573–5576.
68. You S, et al. Detection and characterization of T cells specific for BDC2.5 T cell-stimulating peptides. J Immunol 2003;**170**:4011–4020.