Genomic-scale gene expression profiling of normal and malignant immune cells

Ash A Alizadeh* and Louis M Staudt†

Gene expression variation is critical for the normal development and physiology of immune cells. Using cDNA microarrays, a systematic, genomic-scale view of gene expression in immune cells at many stages of differentiation and activation can be obtained. From the high vantagepoint provided by this technology, the gene expression physiology of immune cells appears remarkably ordered and logical. Each stage of lymphocyte differentiation can be defined by a characteristic gene expression signature. Genes that are co-regulated over hundreds of experimental conditions often encode functionally related proteins. Gene expression profiles also provide unprecedented ability to define the molecular and functional relationships between normal and malignant lymphocyte cell populations.

Addresses
*Department of Biochemistry, M309 Genetics, Alway Building, Stanford University School of Medicine, Stanford, CA 94305, USA
†Metabolism Branch, Division of Clinical Sciences, Building 10, Room 4N114, National Cancer Institute, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA;
e-mail: lstaudt@box-l.nih.gov
Correspondence: Louis M Staudt

Current Opinion in Immunology 2000, **12**:219–225

0952-7915/00/$ – see front matter

Abbreviations

CLL	chronic lymphocytic leukemia
DLBCL	diffuse large-B-cell lymphoma
EST	expressed sequence tag
FL	follicular lymphoma
PCR	polymerase chain reaction

Introduction

The analysis of gene expression variation has been a cornerstone in the study of the immune system. The expression of cell surface markers has allowed the cellular components of the immune system to be subdivided into functionally homogeneous groups. Expression of cell surface receptors, cytokines and chemokines defines the ability of immune cell subpopulations to interact with one another and participate in the immune response. Although traditional methods of gene expression analysis have yielded a wealth of knowledge, they have only allowed a minute fraction of the estimated 100,000 genes in the mammalian genome to be studied. Until recently, a more comprehensive analysis of gene expression was hindered by the dearth of genome sequence and by the lack of methods for parallel gene-expression profiling.

Auspiciously, immunology is about to enter the post-genome era. Full genome sequences for several model organisms are already available and the completion of several more full genomes, including our own, is imminent. An important complement to whole genome sequencing has been the high-throughput, partial sequencing of cDNAs to generate expressed sequence tags (ESTs) [1]. Public databases now contain over two million EST sequences from cDNA libraries representing diverse cells and tissues. Bioinformatics algorithms such as Unigene assemble these ESTs into clusters that tentatively define distinct genes [2]. Currently 93,544 human Unigene clusters are defined (Unigene release 104) which still may only represent ~90% of all human genes. These rich cDNA resources have served as the foundation for recent advances in quantitative, parallel gene-expression analysis. The advent of DNA microarrays now makes analysis of gene expression on a genomic-scale routine and we have begun to view the immune system from this new perspective. Here we summarize work done jointly by the Staudt laboratory and the laboratory of Pat Brown that used DNA microarrays to investigate the responses of immune cell subsets to diverse stimuli and to discern the relationship between normal lymphocyte subpopulations and the human lymphoid malignancies.

DNA microarrays and expression profiling

Although there are currently several different technologies for the production of DNA microarrays, all produce an ordered array of thousands of immobilized nucleic acids on a solid substrate [3,4]. Each of these immobilized elements serves to assay the abundance of its cognate complement in a complex nucleic acid mixture, typically a fluorescently labeled cDNA probe prepared from mRNA. The hybridization of fluorescently labeled molecules to the microarray is captured in a digital image of the microarray using a scanning laser microscope. This image is then processed by various densitometric image-analysis algorithms to produce a table of measurements for each element on the array. Given this highly parallel architecture, each experiment generates abundant data very quickly, necessitating new visualization and analytical tools [5••].

Currently, the most commonly available DNA microarrays are of two main varieties: cDNA microarrays fabricated by robotic spotting of PCR products on glass slides; and *in-situ*-synthesized oligonucleotide arrays fabricated using photolithographic chemistry on silicon wafers. The data generated by these two methods are relatively concordant, as illustrated by the fact that parallel studies of the yeast cell cycle by these two methods yielded comparable results [6•,7••]. In this review, we summarize cDNA microarray experiments in which relative gene expression in two experimental mRNA samples is compared on a single microarray. A cDNA probe is prepared from the two mRNA samples, with each probe incorporating a distinct fluorophore. These two fluorescent probes are mixed and

simultaneously hybridized to a cDNA microarray under stringent conditions. Following washes, data are acquired on a laser scanning microscope for each of the two wavelengths emitted by the pair of fluorophores, allowing quantitation of the ratio of hybridization of the two probes to each cDNA spot on the microarray. These fluorescent ratios are robust measurements of relative gene expression and have been shown to agree well with more standard methods of gene expression analysis such as northern blot hybridization and quantitative RT (reverse transcriptase)-PCR [8,9]. mRNA specimens can be compared to one another using either a direct or an indirect experimental design. In time-course experiments, typically involving exposure of cells to mitogens and/or cytokines, each mRNA from stimulated cells is compared directly to the mRNA from unstimulated cells. For survey experiments requiring comparison of multiple unrelated samples, an indirect experimental design is used in which each mRNA sample is compared to a reference mRNA pool derived from related cell types. The use of a common reference mRNA pool on each cDNA microarray allows the relative gene expression in multiple samples to be determined [10•].

Specialized DNA microarrays for immunology

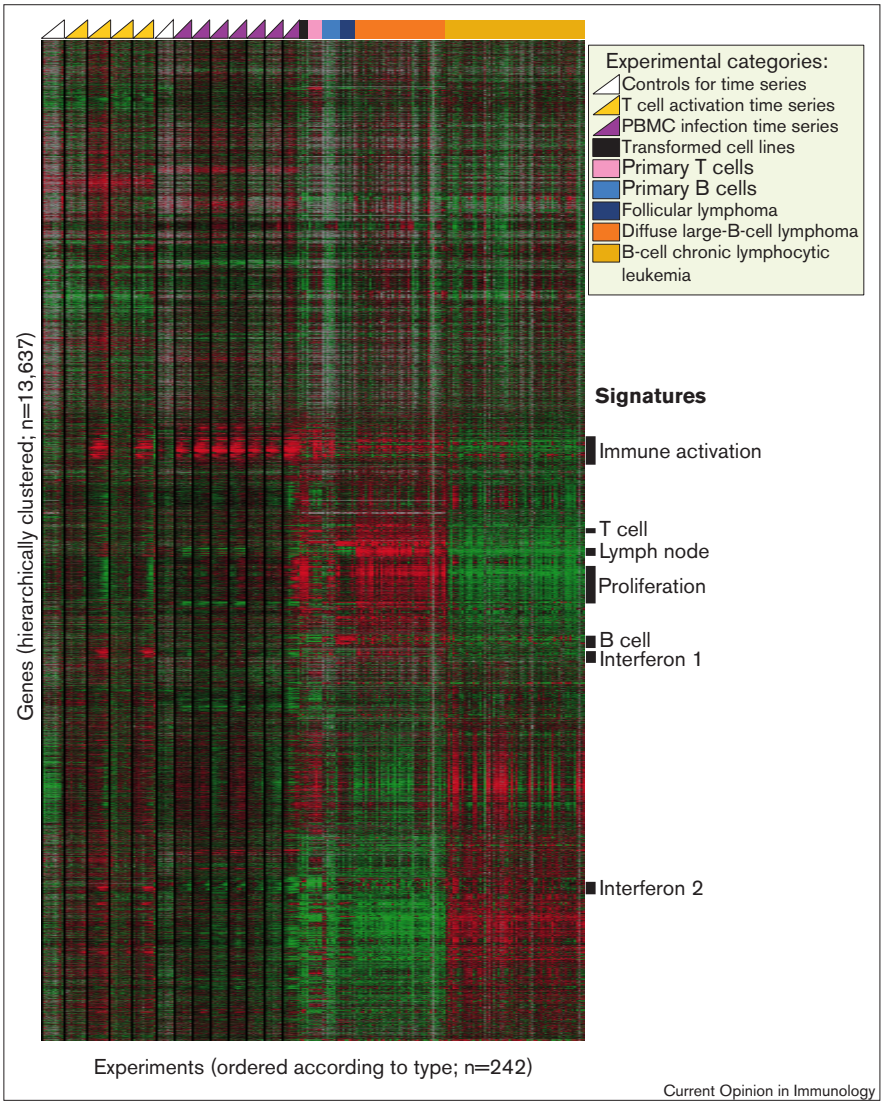
Since human microarrays with full genome coverage are currently unattainable, specialized microarrays tailored to the biological questions being asked are highly useful. We created one such specialized microarray, the 'Lymphochip', to investigate questions in normal and malignant lymphocyte biology by selecting genes that are preferentially expressed in lymphoid cells or that are of known immunological or oncological importance [10•]. One of our goals in constructing the Lymphochip was to survey gene expression in non-Hodgkin's lymphomas in order to classify these malignancies into molecularly defined and clinically useful categories. Because of the suspected importance of the germinal center B cell to the etiology of non-Hodgkin's lymphomas, we began by creating a cDNA library from germinal center B cells purified from human tonsils [10•]. Under the auspices of the Cancer Genome Anatomy project [11], high-throughput EST sequencing of 50,898 clones from this library revealed that roughly 18% of the ESTs from this library had not been previously observed. This rich source of novel genes formed the foundation of the Lymphochip and 12,069 out of 17,856 cDNA clones on the Lymphochip are derived from the germinal center B cell library. We also included an additional 2338 cDNA clones from libraries derived from several non-Hodgkin's lymphomas as well as chronic lymphocytic leukemia (CLL). Finally, we supplemented the array with clones representing a variety of genes that are induced or repressed during B and T lymphocyte activation by mitogens or cytokines [12] and a curated set of 3186 'named' genes of importance to lymphocyte and/or cancer biology. Different cDNA clones representing the same human gene were included for ~30% of the genes to provide controls for the reproducibility of gene expression quantitation. The complete annotated list of these cDNAs can be found at http://llmpp.nih.gov/lymphoma. The Lymphochip cDNA microarray thus provides two complementary genomic-scale views of normal and malignant lymphocytes: the 'named' gene component of the microarray generates insights and hypotheses related to known cellular pathways whereas the novel genes on the microarray can define gene expression profiles that relate or distinguish one lymphoid population from another.

The landscape of gene expression in normal and transformed immune cells

Using the Lymphochip, a database of several million separate measurements of gene expression in normal and malignant immune cells has been assembled which is the beginning of a systematic study of the genomic expression in various immune cells. Figure 1 depicts ~3.8 million measurements of gene expression made on ~13,000 genes using 242 Lymphochip microarrays. The data displayed in this figure have been organized using a hierarchical clustering algorithm and displayed as described in [5••]. In essence, this algorithm uses the standard Pierson correlation coefficient to define the similarity in expression of each pair of genes across all experimental samples. The algorithm then arranges the order of the genes to place genes with the most correlated expression near each other.

The organized data are graphically displayed in tabular format, with each horizontal row of colored boxes representing the variation in mRNA abundance for a given gene across all samples, and each vertical column representing the variation in mRNA abundance for a single sample across all genes. A color scale is used to represent variations in transcript abundance for each gene, where shades of red represent higher expression and shades of green represent lower expression; the intensity of the color corresponds to the magnitude of mRNA expression. The left half of Figure 1 depicts data from experiments studying the activation responses of Jurkat T cells to mitogens (5 experimental series with 10 time points each) and human peripheral blood mononuclear cells to various conditions including heat-killed bacteria and pharmacological stimulation (eight experimental series with eight time points each). Each time series is indicated in the legend above the figure by a triangle that reflects the progression of time during each experiment (from left to right). For these experiments, the gene expression color scale represents gene expression in each stimulated cell sample relative to unstimulated cells; red and green represent fold induction and repression, respectively. In contrast, the right half of Figure 1 represents data from a survey of gene expression variation in normal and malignant lymphocytes, with each cell type grouped together as indicated by the color code above the figure. For these experiments, the color scale reflects the relative expression level for a given gene across all samples surveyed; red and green represent fold overexpression and underexpression relative to the median expression of the gene in all samples, respectively.

Figure 1



A gene expression map of immune cells. Depicted are the results of over 3.8 million measurements of gene expression made on 13,637 genes using 242 Lymphochip microarrays. Each row represents one element on the microarray and each column represents one of 242 Lymphochip microarray experiments. The experiments are grouped by cell type as indicated by the color code and key at the top. Jurkat T cells were unstimulated (far left) or activated by mitogens (5 experiments with 10 time points each); peripheral blood mononuclear cells (PBMC) were unstimulated or activated (8 experiments with 8 time points each). Overexpressed or induced genes are colored red whereas underexpressed or repressed genes are colored green. Black indicates gene expression measurements close to the median or to that of the unstimulated sample in a time course series. Gray indicates undetermined values. Various gene expression signatures are indicated at the right. See text for more details.

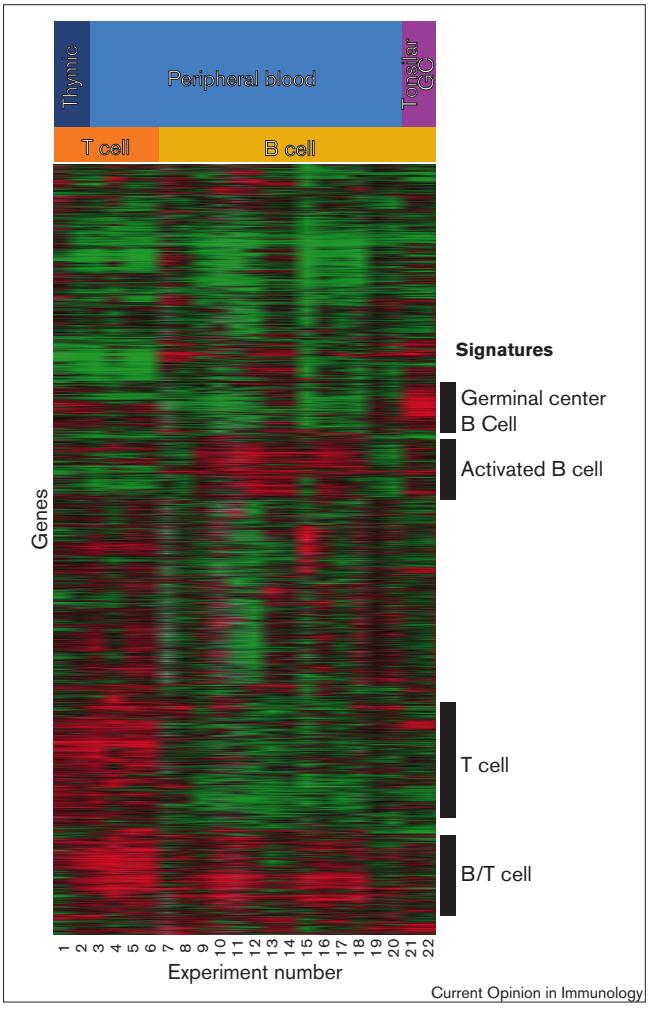
The hierarchical clustering algorithm groups genes according to the similarity in their expression and thus can be used to define 'clusters' of genes that are coordinately expressed [5••]. These gene expression clusters are often revealed by prominent contiguous patches of red or green color in graphical depictions of the data (Figure 1). If a gene cluster includes genes that are characteristically expressed in a given cell type, biological process or signaling pathway, the cluster can be operationally defined as a gene expression 'signature' (AA Alizadeh *et al.*, unpublished data; see Note added in proof). Several such gene expression signatures are described below and are indicated on the right-hand sides of Figures 1 and 2.

Gene expression signatures and function in the immune system

Genomic-scale gene expression studies have revealed, in many instances, a striking relationship between the expression pattern of a gene and its function [5••]. Several studies in the yeast *Saccharomyces cerevisiae* have shown that clusters of co-regulated genes often include genes that are subunits of the same protein complex or genes that are involved in the same biochemical pathway [5••,6•,7••,8,13,14] For example, all subunits of the proteosome in *S. cerevisiae* are co-regulated, as are genes involved in the diauxic shift and in sporulation. Analysis of promoters of such clustered genes has revealed common sequence motifs, in some cases reflecting regulation by known transcription factors [7••,8,13,14]. Because of the intimate relationship between gene expression and gene function, a novel gene can hypothetically be assigned a function if it is co-regulated with a set of genes that are involved in the same biochemical process.

The most easily interpreted gene expression signatures in Figure 1 involve genes that are expressed in distinct immune cell types at particular stages of differentiation. The T cell gene expression signature, for example,

Figure 2



Gene expression during lymphocyte differentiation and activation. For each gene, expression measurements were centered around the median expression measurement in this experimental series. Various gene expression signatures are indicated at the right. See the text and the legend to Figure 1 for details. Experiment numbers correspond to various treatments: 1, fetal $CD4^+$ thymic cells, I+P (ionomycin+PMA [phorbol myristate acetate]) stimulated; 2, fetal $CD4^+$ thymic T cells; 3, neonatal naive $CD4^+$ blood T cells; 4, adult naive $CD4^+$ blood T cells; 5, adult naive $CD4^+$ blood T cells, I+P stimulated; 6, neonatal naive cord blood T cells, I+P stimulated; 7, cord blood B cells; 8, blood B cells; 9, blood B cells, anti-IgM stimulated (24 hours); 10, blood B cells, anti-IgM stimulated (6 hours); 11, blood B cells, anti-IgM+CD40L stimulated (6 hours); 12, blood B cells, anti-IgM+CD40L stimulated (24 hours); 13, blood B cells, anti-IgM+CD40L stimulated (6 hours); 14, blood B cells, anti-IgM+CD40L stimulated (24 hours); 15, blood B cells, anti-IgM+CD40L+IL-4 stimulated (6 hours); 16, blood B cells, anti-IgM+CD40L+IL-4 stimulated (24 hours); 17, blood B cells, anti-IgM+IL-4 stimulated (24 hours); 18, blood B cells, anti-IgM+IL-4 stimulated (6 hours); 19, blood $CD27^+$ memory B cells; 20, blood $CD27^-$ naive B cells; 21, tonsilar GC (germinal center) B cells; 22, tonsilar GC centroblasts.

includes T cell differentiation genes (e.g. the gene encoding CD2), T cell receptor genes and genes whose products are involved in T cell receptor signaling (LAT and fyn). The B cell gene expression signature likewise includes B cell differentiation genes (encoding CD20 and immunoglobulin J chain) and genes encoding B cell receptor components (e.g. CD79B). Genes in the lymph node gene expression signature were generally found to be expressed in lymph node biopsies from patients with diffuse large-B-cell lymphoma (DLBCL) and reflect the complex biology of this secondary lymphoid organ. This gene expression signature includes monocyte differentiation genes (encoding CD14 and c-fms), natural killer (NK) cell differentiation genes (e.g. the gene encoding NK4), a variety of chemokine genes (encoding IP-10, MIG, PARC and RANTES) and genes whose products are involved in generating and remodeling the extracellular matrix (osteonectin and matrix metalloproteinase 9).

Other gene expression signatures in Figure 1 include genes involved in related biochemical pathways or processes. The 'proliferation gene expression signature' is comprised of hundreds of genes that function in one of the numerous biochemical processes that change during the cell cycle. As would be expected, the genes in the proliferation gene expression signature were highly expressed by the aggressive DLBCLs but not by the indolent lymphoid malignancies, follicular lymphoma (FL) and CLL. This gene expression signature includes genes regulating cell cycle progression (encoding cdk4 and cdc2; cyclins A, B1, E, F; and cdk kinase inhibitors p16 and p18), genes whose products are involved in DNA replication (replication factor C, MCM2 DNA replication licensing factor and dihydrofolate reductase) and checkpoint genes that monitor proper cell cycle progression (encoding BUB1 and CHK1).

The 'immune activation gene expression signature' includes hundreds of genes that were induced when B cells, T cells and peripheral blood mononuclear cells were activated by a variety of mitogens, cytokines and microorganisms (Figure 1). The organizing principle of this signature appears to involve the transcription factor NF-κB [15]. NF-κB factors are among the most important orchestrators of innate and acquired immunity and consist of homo- or hetero-dimers of proteins in the Rel family. Homologues of the mammalian NF-κB gene family are highly conserved in evolution. In insects, the NF-κB factors respond to a variety of environmental stresses. In the mammalian immune system, NF-κB factors can function in innate immunity by responding to components of microbial pathogens such as lipopolysaccharide and double-stranded RNA. In acquired immunity, NF-κB factors are activated by signals through the B cell and T cell antigen receptors.

A host of genes have already been identified as transcriptional targets of NF-κB factors [16]; many are present in the immune activation signature represented in Figure 1. Interestingly, genes encoding many of the known NF-κB/Rel proteins are in this signature (i.e. NF-κB 1, NF-κB 2, RelB, c-Rel and IκBα), reflecting the recognized autoregulatory nature of this intricately controlled signaling

system. The NF-κB cluster includes genes encoding a prominent group of secreted immune mediators including cytokines (IL-1α, IL-1β, IL-3, IL-6, IL-10, TNFα, TNFβ, SCF, G-CSF and GM-CSF) and chemokines (IL-8, MIP-1α, MIP-1β, MIP-2 β, MIP-3α). NF-κB is well known to block programmed cell death in many systems [15] and, correspondingly, the immune activation signature includes genes encoding A1, an anti-apoptotic bcl-2 family member, and c-IAP2, a homologue of the insect-virus-encoded inhibitor of apoptosis (IAP) protein. The immune activation gene expression signature additionally includes a variety of known and novel genes that may also be NF-κB target genes. However, the mRNA samples in which these genes were most highly expressed were those derived from Jurkat T cells or peripheral blood mononuclear cells that were mitogenically stimulated. Clearly, these stimuli engage multiple signal transduction pathways and, therefore, future experimentation will be required to determine which genes in this immune activation signature are NF-κB targets.

Two prominent gene expression signatures include targets of interferon signaling. The genes in both interferon signatures were induced during Jurkat T cell activation but the two signatures were separated by the hierarchical clustering algorithm because their component genes were differentially expressed in other cell types, notably CLL cells. The Interferon 1 gene expression signature includes the gene encoding the interferon-responsive transcription factor STAT1, MHC class I genes, genes involved in antigen presentation (TAP1 and the proteosome subunit LMP2) and a host of known interferon-responsive genes of unknown function (Figure 1). The Interferon 2 gene expression signature includes genes encoding several interferon-responsive transcription factor subunits (STAT2, IRF-2 and ISGF3γ), MHC class II genes, the influenza virus resistance gene MxB and several known interferon-responsive genes of unknown function (Figure 1).

Each of the gene expression signatures highlighted above contains a large number of novel genes of unknown function. Given the functional relatedness of the genes in these signatures, such gene expression data provide clues to the function of these novel genes, even without any knowledge of their predicted gene products.

Discovering etiological links between normal and malignant lymphocytes

One of our primary goals in creating the Lymphochip cDNA microarray was to develop a systematic understanding of gene expression changes during B cell differentiation. The motivation behind much of this work was to define in molecular terms the differences among the wide variety of B lymphoid malignancies. We suspected that the biology and clinical behavior of a B cell malignancy would be strongly influenced by the biology of its normal counterpart in B cell differentiation. Until now, hypotheses regarding the cellular origins of B cell lymphomas and leukemias have relied on the expression of a handful of surface markers and on the status of the immunoglobulin loci in the malignancies. Most of the non-Hodgkin's B cell lymphomas appear to be derived from mature B cells as they have rearrangements of the immunoglobulin heavy and light chain genes. Further insight into the origin of these malignancies has come from analysis of their immunoglobulin variable-region sequences. The process of somatic hypermutation of immunoglobulin genes is relatively restricted to the germinal center microenvironment [17]. Many categories of non-Hodgkin's B-cell lymphoma have mutated immunoglobulin variable genes, suggesting that the B cells that gave rise to these lymphomas traversed the germinal center microenvironment [18]. However, the presence of immunoglobulin gene somatic mutations in a lymphoma cannot be used to decide whether the lymphoma was derived from a germinal center B cell per se or from a post-germinal-center B cell.

We therefore sought to systematically define the gene expression programs of normal lymphocytes and to determine the extent to which these programs are inherited by human lymphoid malignancies. Figure 2 presents gene-expression measurements in a variety of lymphocyte subpopulations that were isolated by magnetic and fluorescent cell sorting techniques from human peripheral blood, thymus or tonsil. The lymphocyte subpopulations were assayed for gene expression either in an unstimulated state or after stimulation with a variety of mitogens and/or cytokines. Several gene-expression signatures are evident from this display. A large number of lineage-restricted genes distinguish B and T lymphocytes but these two lineages also share the expression of many genes. Within the B cell lineage, a striking signature of activated B cells is evident. These genes are much more highly expressed in activated B cells than activated T cells and are also not as highly expressed in germinal center B cells. Interestingly, two of these genes encode CXCR5 [19] and CCR7 [20], chemokine receptors involved in lymphocyte homing. CCR7 is responsible for retention of B cells in the periarteriolar lymphoid sheath whereas CXCR5 mediates homing of B cells towards follicular dendritic cells, leading to germinal center formation. Thus, these chemokine receptors have opposing effects on lymphocyte migration and an activated B cell apparently integrates the signals delivered by the binding of these receptors to their respective ligands [20]. BCL-2 is another interesting B cell activation gene product that is not upregulated in mitogenically stimulated T cells and is profoundly downregulated in germinal center B cells. BCL-2 may therefore play a restricted role in preventing activation-induced cell death in activated B cells that are destined to differentiate into plasma cells in the periarteriolar lymphoid sheath.

The Germinal center B cell signature consists of hundreds of genes that are not expressed in resting B cells, activated

B cells or T cells. The Lymphochip microarray was, of course, especially suited to discovering novel genes that contribute to this signature since this microarray contains over 12,000 cDNAs derived from a germinal center B cell library. The discovery of this germinal center B cell signature demonstrates that germinal center B cells represent a distinct stage of B cell development. Indeed it has not been possible, to date, to generate germinal center B cells *in vitro* by activation of resting B lymphocytes. It is likely, therefore, that specific *in vivo* stimuli, perhaps delivered by T cells and/or follicular dendritic cells, are required to establish the germinal center B cell gene expression program. Once established, the germinal center B cell signature can be a stable phenotype, as evidenced by the fact that some B cell lymphoma cell lines faithfully maintain the expression of these signature genes during *in vitro* culture (AA Alizadeh *et al.*, unpublished data; see Note added in proof). Although most of the genes in the germinal center signature are novel genes of unknown function, the few named genes in this signature are noteworthy. The gene for BCL-6 encodes a critical transcriptional repressor protein that is required for the normal development of germinal center B cells [21–23]. The BCL-6 gene is also the most commonly translocated gene in DLBCLs, the most common subtype of non-Hodgkin's lymphoma [24]. Thus, there is an intimate relationship between the expression of BCL-6 and its function as a regulator of germinal center formation and B cell malignancy. The gene encoding A-myb is another transcription factor gene in the germinal center B cell signature that can cause follicular hyperplasia when overexpressed in transgenic mice [25]. Both of these transcription factors may therefore contribute to the stable maintenance of the germinal center B cell program. BCL-6 is downregulated in plasma cells, suggesting that the loss of its function may allow the distinctive gene expression program of plasma cells to proceed [24].

This map of gene expression changes during lymphocyte differentiation provides a rich framework in which to interpret the gene expression in lymphoid malignancies. We have surveyed gene expression in three of the most common human B cell malignancies: DLBCL, FL and CLL (AA Alizadeh *et al.*, unpublished data; see Note added in proof). It has been suggested that both DLBCL and FL derive from a germinal center B cell or a B cell at a later differentiation stage, based on the fact that the immunoglobulin genes in these malignancies have somatic mutations. The gene expression profile of FLs strongly suggests that this malignancy is derived from the germinal center B cell itself. Virtually all of the genes in the germinal center B cell signature are expressed in FLs. On the basis of this shared gene expression signature, FLs would be expected to resemble normal germinal center B cells functionally. Indeed, it has been shown that FLs continue the process of somatic hypermutation of immunoglobulin genes after they have been transformed [26]. By contrast, CLLs did not express most of the genes in the germinal center B cell gene expression signature. DLBCLs were heterogeneous in their expression of the germinal center B cell signature. About half of the DLBCL cases strikingly resembled germinal center B cells in gene expression. The other DLBCL cases, by contrast, lacked expression of most of the genes in the germinal center B cell signature. Instead, these DLBCL cases had their own characteristic gene expression signature that overlapped significantly, but not completely, with the activated B cell signature described above (AA Alizadeh *et al.*, unpublished data; see Note added in proof). Thus, gene expression profiling was capable of subdividing a single diagnostic category of lymphoma into two apparently distinct diseases. One DLBCL subtype seems clearly to be derived from the germinal center B cell whereas the origin of the other subtype is presently enigmatic. We are currently extending this analysis to include all subtypes of lymphoid malignancy in a broadly collaborative project termed the Lymphoma/Leukemia Molecular Profiling Project (http://llmpp.nih.gov/lymphoma). We anticipate that our understanding of gene expression in normal lymphocytes will continue to inform our understanding of the pathogenesis of human lymphoid malignancies.

Conclusions

Very soon, DNA microarray analysis of gene expression will change from being an 'emerging technology' to a 'classic technique'. However, it will take much longer for the flood of data from this technology to be synthesized into basic principles of gene expression 'physiology'. Genomic-scale gene expression data should ultimately lead to an integrated view of cellular differentiation. In this arena, the immune system offers distinct advantages since cells at different stages of lymphocyte differentiation can be readily isolated and manipulated *in vitro*. A genomic description of differentiation should reveal orchestrated fluxes in molecules involved in cell proliferation, cell death, energy metabolism, intracellular signaling, intercellular communication and cell migration. A genomic description of diseases will provide molecular profiles of pathogenic cell types and reveal the signaling pathways that are perturbed in them. Autoimmune and immune deficiency diseases that are currently defined by their clinical phenotypes will soon be divided into molecular subtypes based on functional genomics experiments. Likewise, all lymphoid and myeloid malignancies will be compared and contrasted with normal subsets of immune cells, as described above for DLBCL. Such experiments will require textbooks of medicine to be reformulated to emphasize molecular definitions of disease. Armed with a battery of enabling computer tools, scientists should be able to parlay functional genomic data into new, fundamental biological principles and into rational therapies that target pathogenic molecules in molecularly defined diseases.

Note added in proof

The work referred to in the text as (AA Alizadeh *et al.*, unpublished data) has now been accepted for publication [27].

Acknowledgements

The authors wish to thank the many members of the Staudt laboratory and of the laboratories of Pat Brown and David Botstein who made contributions to the studies reviewed here. Especially, we thank Chih Long Liu, Maximilian Diehn, Oliver J Rando, Gerald R Crabtree, David A Relman, Jennifer C Boldrick and Christopher E Belcher whose unpublished data were included in Figure 1. AAA was initially supported by the Howard Hughes Medical Institute Research Scholar Program while at the National Institutes of Health and then by the Medical Scientist Training Program at Stanford University.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC: **Sequence identification of 2,375 human brain genes.** *Nature* 1992, **355**:632-634.

2. Schuler GD: **Pieces of the puzzle: expressed sequence tags and the catalog of human genes.** *J Mol Med* 1997, **75**:694-698.

3. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.

4. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14**:1675-1680.

5. •• Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
This critical paper describes the application of agglomerative hierarchical clustering to organize gene expression data produced from cDNA microarrays. It also presents visualization tools for large gene-expression data sets.

6. • Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, Davis RW: **A genome-wide transcriptional analysis of the mitotic cell cycle.** *Mol Cell* 1998, **2**:65-73.
See annotation [7••].

7. •• Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297.
Papers [6•,7••] report on the use of cDNA microarrays in a systematic analysis of the yeast cell cycle, leading to the discovery of hundreds of cell-cycle-regulated genes.

8. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.

9. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J Jr, Boguski MS *et al.*: **The transcriptional program in the response of human fibroblasts to serum.** *Science* 1999, **283**:83-87.

10. • Alizadeh A, Eisen M, Davis RE, Ma C, Sabet H, Tran T, Powell J, Yang L, Marti G, Moore T *et al.*: **The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes.** *Cold Spring Harbor Symp Quant Biol* 2000, in press.
This paper describes the construction of a custom cDNA microarray (Lymphochip) for studies in normal and malignant human cells.

11. Strausberg RL, Dahl CA, Klausner RD: **New opportunities for uncovering the molecular basis of cancer.** *Nat Genet* 1997, **15**:415-416.

12. Alizadeh A, Eisen M, Botstein D, Brown PO, Staudt LM: **Probing lymphocyte biology by genomic-scale gene expression analysis.** *J Clin Immunol* 1998, **18**:373-379.

13. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I: **The transcriptional program of sporulation in budding yeast.** *Science* 1998, **282**:699-705.

14. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-285.

15. Ghosh S, May MJ, Kopp EB: **NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses.** *Annu Rev Immunol* 1998, **16**:225-260.

16. Siebenlist U, Franzoso G, Brown K: **Structure, regulation and function of NF-kappa B.** *Annu Rev Cell Biol* 1994, **10**:405-455.

17. Kelsoe G: **In situ studies of the germinal center reaction.** *Adv Immunol* 1995, **60**:267-288.

18. Klein U, Goossens T, Fischer M, Kanzler H, Braeuninger A, Rajewsky K, Kuppers R: **Somatic hypermutation in normal and transformed human B cells.** *Immunol Rev* 1998, **162**:261-280.

19. Forster R, Mattis AE, Kremmer E, Wolf E, Brem G, Lipp M: **A putative chemokine receptor, BLR1, directs B cell migration to defined lymphoid organs and specific anatomic compartments of the spleen.** *Cell* 1996, **87**:1037-1047.

20. Forster R, Schubel A, Breitfeld D, Kremmer E, Renner-Muller I, Wolf E, Lipp M: **CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs.** *Cell* 1999, **99**:23-33.

21. Dent AL, Shaffer AL, Yu X, Allman D, Staudt LM: **Control of inflammation, cytokine expression, and germinal center formation by BCL-6.** *Science* 1997, **276**:589-592.

22. Fukuda T, Yoshida T, Okada S, Hatano M, Miki T, Ishibashi K, Okabe S, Koseki H, Hirosawa S, Taniguchi M *et al.*: **Disruption of the Bcl6 gene results in an impaired germinal center formation.** *J Exp Med* 1997, **186**:439-448.

23. Ye BH, Cattoretti G, Shen Q, Zhang J, Hawe N, de Waard R, Leung C, Nouri-Shirazi M, Orazi A, Chaganti RS *et al.*: **The BCL-6 proto-oncogene controls germinal-centre formation and Th2- type inflammation.** *Nat Genet* 1997, **16**:161-170.

24. Staudt LM, Dent AL, Shaffer AL, Yu X: **Regulation of lymphocyte cell fate decisions and lymphomagenesis by BCL-6.** *Int J Immunol* 1999, **18**:381-403.

25. DeRocco SE, Iozzo R, Ma XP, Schwarting R, Peterson D, Calabretta B: **Ectopic expression of A-myb in transgenic mice causes follicular hyperplasia and enhanced B lymphocyte proliferation.** *Proc Natl Acad Sci USA* 1997, **94**:3240-3244.

26. Bahler DW, Levy R: **Clonal evolution of a follicular lymphoma: evidence for antigen selection.** *Proc Natl Acad Sci USA* 1992, **89**:6770-6774.

27. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X *et al.*: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**:503-511.