

Arthur L. Shaffer
George Wright
Liming Yang
John Powell
Vu Ngo
Laurence Lamy
Lloyd T. Lam
R. Eric Davis
Louis M. Staudt

A library of gene expression signatures to illuminate normal and pathological lymphoid biology

Authors' addresses

Arthur L. Shaffer¹, George Wright², Liming Yang³,
John Powell³, Vu Ngo¹, Laurence Lamy¹, Lloyd T.
Lam¹, R. Eric Davis¹, Louis M. Staudt¹

¹Metabolism Branch, Center for Cancer
Research, National Cancer Institute, Bethesda,
MD, USA.

²Biometric Research Branch, DCTD, National
Cancer Institute, Bethesda, MD, USA.

³Bioinformatics and Molecular Analysis Section,
Computational Bioscience and Engineering
Laboratory, CIT, National Institute of Health,
Bethesda, MD, USA.

Correspondence to:

Louis M. Staudt, MD, PhD
Metabolism Branch
CCR, NCI, Bldg. 10
Rm. 4 N114, NIH
Bethesda, MD 20892
USA

Tel.: +1 301 402 1892

Fax: +1 301 496 9956

E-mail: lstaudt@mail.nih.gov

Acknowledgements

We thank the members of the Staudt laboratory for their input and support. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Immunological Reviews 2006

Vol. 210: 67–85

Printed in Singapore. All rights reserved

© 2006 Blackwell Munksgaard

No claim to original US
government works

Immunological Reviews
0105-2896

Summary: Genomics has provided a lever to pry open lymphoid cells and examine their regulatory biology. The large body of available gene expression data has also allowed us to define the of coordinately expressed genes, termed gene expression signatures, which characterize the states of cellular physiology that reflect cellular differentiation, activation of signaling pathways, and the action of transcription factors. Gene expression signatures that reflect the action of individual transcription factors can be defined by perturbing transcription factor function using RNA interference (RNAi), small-molecule inhibition, and dominant-negative approaches. We have used this methodology to define gene expression signatures of various transcription factors controlling B-cell differentiation and activation, including BCL-6, B lymphocyte-induced maturation protein-1 (Blimp-1), X-box binding protein-1 (XBP1), nuclear factor- κ B (NF- κ B), and c-myc. We have also curated a wide variety of gene expression signatures from the literature and assembled these into a signature database. Statistical methods can define whether any signature in this database is differentially expressed in independent biological samples, an approach we have used to gain mechanistic insights into the origin and clinical behavior of B-cell lymphomas. We also discuss the use of genomic-scale RNAi libraries to identify genes and pathways that may serve as therapeutic targets in B-cell malignancies.

Introduction

The oft-quoted dictum 'chance favors only the prepared mind' (Louis Pasteur) was coined in the pre-genomic era. How can one possibly prepare one's mind for the rush of information provided by genomics technologies? One such technology, termed gene expression profiling, provides a comprehensive view of the transcriptome and can be used to measure the expression of all genes in hundreds of biological samples within a few weeks' time. Fortunately, there are many coordinated efforts to annotate genes with respect to the function of their protein products, and these have been widely applied to extract meaning from these large data sets. For many genes, however, the functional annotations in current databases present an average view, often culled from the analysis of the gene in many different cell types under

disparate conditions of activation. What may be equally important in interpreting the expression of a gene in a biological sample is to understand its regulatory biology, i.e. which transcription factors and signaling pathways regulate its expression.

One contextual approach that has aided us in the analysis of large gene expression data sets involves the identification of gene expression signatures (Fig. 1). Signatures are sets of coordinately expressed genes that define a cell's physiology (differentiation stage, activation state, proliferation status, cellular environment, etc.). Therefore, signatures provide a way to derive biological meaning from the expression of an individual gene in a biological sample (1). This kind of analysis, however, relies on access to a relatively large and diverse set of gene expression data. Fortunately, genomic researchers have adopted a 'share-the-wealth' ethos that leads to the public deposition of entire gene expression data sets upon publication.

In this review, we present the methodology that we have used to define gene expression signatures, with particular emphasis on signatures of B-cell transcription factor action. We describe the creation of SignatureDB, a compendium of gene expression signatures culled from our own work and the literature. We show how this database can be used to gain insights into the mechanisms underlying B-cell malignancies. Finally, we discuss our efforts to use genomic-scale RNA interference (RNAi) expression libraries to understand the functional anatomy of B cells and to identify new targets for therapeutic intervention in mature B-cell malignancies.

Regulatory biology of individual B-cell transcription factors

Gene expression profiling provides a comprehensive phenotype of cells that can be used to monitor changes in cellular

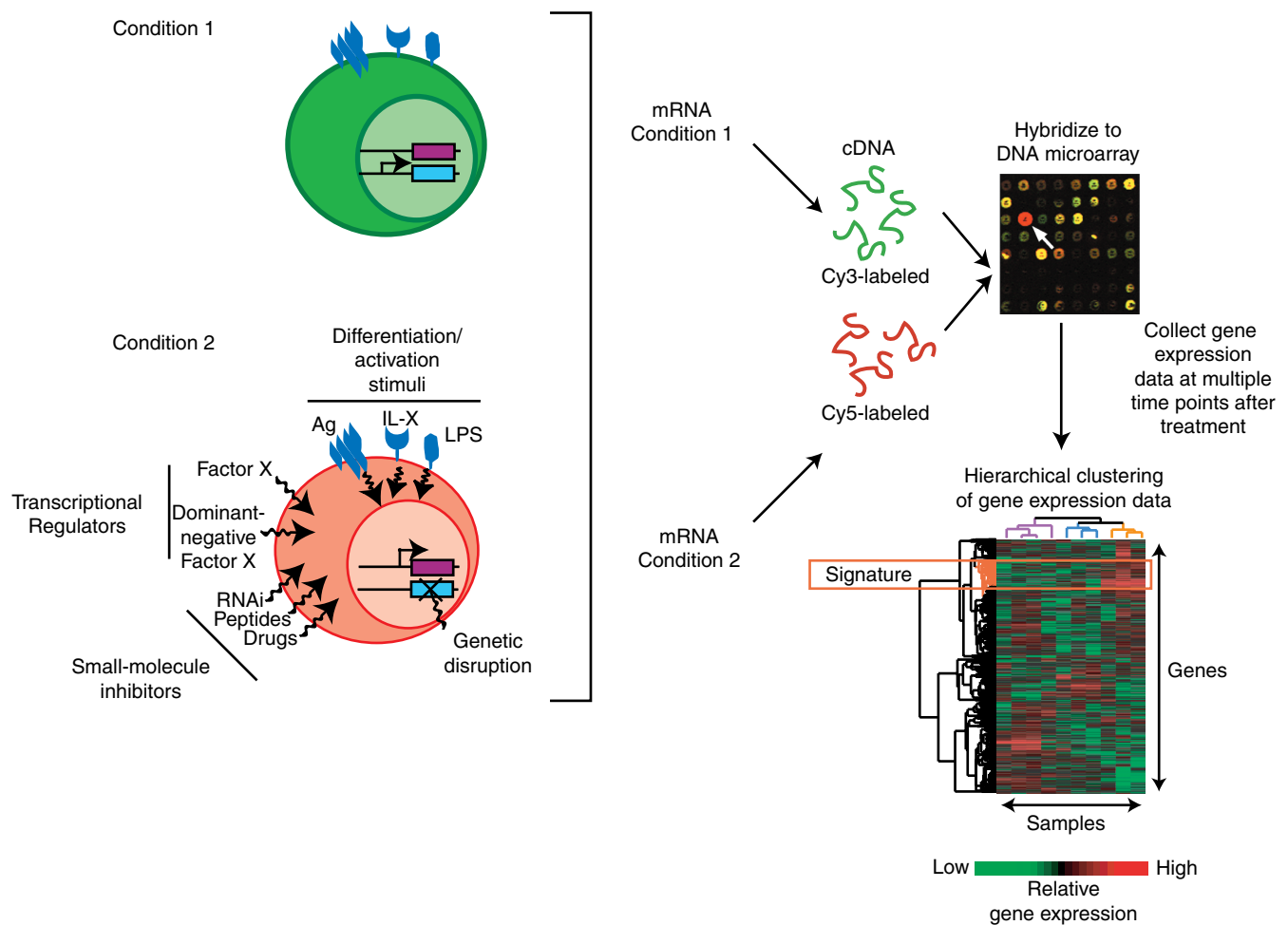


Fig. 1. Uncovering the systems biology of gene expression using DNA microarrays. Illustrated are several experimental approaches that can be used to identify genes under the control of a particular regulatory factor or signaling pathway including overexpression, blockade of expression or

activity, and genetic ablation. After DNA microarray hybridization and data collection, statistical or pattern recognition methods can identify a set of genes whose coordinate expression reflects a particular aspect of cellular physiology. Such a gene set is termed a gene expression signature (1).

biology that occur when a particular regulatory pathway is perturbed (Fig. 1). Regulatory pathways can be manipulated experimentally by the stimulation of cells through surface receptors, the ectopic expression of a regulatory factor, the knockdown of gene expression using RNAi, the use of small-molecule pathway inhibitors, and the genetic ablation of genes in the genome (Fig. 1). By collecting gene expression profiles after such experimental manipulations, it is possible to gain insight into the regulatory strategy of cells. This methodology now falls under the general rubric of 'systems biology'. Attempts are being made to reconstruct regulatory networks using computational methods, starting with the large sets of 'static' gene expression profiles obtained from tumor and normal cells (2). Such methodology will be dramatically strengthened by the analysis of gene expression changes that accompany the direct manipulation of regulatory pathways.

We have studied the regulatory biology of individual B cell transcription factors to understand their role in B-cell and disease. We have used gene expression profiling in conjunction with three complementary experimental approaches to identify the target genes of individual transcriptional factors. In one approach, a regulatory factor is ectopically introduced into cells that normally lack its expression. Subsequent gene expression analysis reveals those genes whose expression levels are increased or decreased by the direct or indirect action of the regulatory factor. This approach alone is insufficient to confidently identify transcription factor targets, given the complicated interactions among transcription factors (dimerization, interaction with co-activators/repressors, cooperative DNA binding, etc.) that may be disrupted by the overexpression of a regulatory factor. Consequently, gene expression data derived from this approach, in the absence of other data, can lead to inappropriate conclusions regarding the function of a regulatory factor.

A second important approach is to block the action of a regulatory factor, either by expressing dominant-negative forms of the factor or by knocking down the expression of the factor by RNAi. This approach in isolation can also be problematic. Indirect effects on gene expression may arise when the steady-state level of a targeted transcription factor is reduced (e.g. the release of a co-repressor or co-activator from a factor and subsequent recruitment to another transcriptional factor). The overexpression of dominant-negative forms of a transcription factor may titrate interacting proteins away from other cellular functions. Finally, the overexpression of double-stranded interfering RNAs can lead to non-specific changes in gene expression in certain cell types under certain transfection conditions. This last caveat appears

not to be a concern using short hairpin RNA (shRNA) approaches in the human B-cell lines that we have studied thus far (see below).

A third approach is to make use of a dynamic cellular system in which a regulatory factor is normally induced or repressed, such as when cells differentiate or respond to extracellular signals. Manipulation of a regulatory factor under these circumstances may reveal a different regulatory biology than can be gleaned from the analysis of static systems such as cell lines. In particular, one can study the differentiation or activation of cells collected from mice engineered to lack a particular regulatory factor in comparison with wild-type cells. For example, we have studied the gene expression profiles of B cells during *in vitro* differentiation to plasma cells (PCs) from mice lacking particular key regulators of plasmacytic differentiation (3).

Finally, an emerging and powerful genomics technology employs chromatin immunoprecipitation (ChIP) to identify the binding sites of individual regulatory factors in the genome (4). An important caveat to this approach, however, is that a transcription factor can be bound to promoters and enhancers in the genome at which it is functionally inactive. This inactivity is presumably either due to the lack of cooperating factors or due to the presence of dominant repressing factors in the vicinity of the transcription factor.

Given the virtues and flaws of each of these approaches, an integrated effort using each of these approaches is likely to provide the truest view of the regulatory biology of a transcription factor, as was evidenced in early work in yeast (4). As detailed below, we have employed many of these genomic approaches to understand the role of B-cell transcription factors in normal and malignant B-cell biology.

BCL-6

Malignancies of mature B cells are characterized by disease-specific translocations that impart a unique biology on each type of lymphoma or leukemia (5). In roughly 30% of diffuse large B-cell lymphomas (DLBCLs), translocation of the BCL-6 gene results in the replacement of the promoter region with a heterologous promoter provided by the translocation partner. BCL-6 is a POZ-domain-containing zinc-finger transcriptional repressor, whose normal protein expression in the B-cell lineage is largely restricted to germinal center B cells (GCBs) (6, 7). Mice lacking BCL-6 demonstrated that its expression is required, in a B cell-autonomous fashion, for germinal center formation (8–10). However, these studies did not shed light on the oncogenic potential of BCL-6 (8–10).

BCL-6 translocations found in DLBCL do not generally lead to its overexpression but rather dysregulate its expression. Normally, BCL-6 expression is extinguished in PCs, and the BCL-6 translocations in DLBCL prevent this physiological downregulation. Translocations involving BCL-6 never disrupt its coding sequence but instead replace its normal upstream regulatory regions with promoters from other genes expressed in GCBs (11, 12). The non-coding 5' end of the BCL-6 gene is also subject to somatic hypermutation that may affect transcriptional regulatory elements. Indeed, there appears to be a well-conserved set of BCL-6 binding sites in the 5' untranslated region (UTR) of BCL-6 that act as negative-feedback elements regulating its expression (13, 14). These regulatory sites can be lost during the translocation of BCL-6 or altered by somatic mutation, thereby contributing to the dysregulated expression of BCL-6 in malignant B cells (Fig. 2).

Although its frequent appearance as a translocation partner in DLBCL strongly suggests that BCL-6 is an oncogene, the proof of this role and insights into the mechanisms by which BCL-6 drives B-cell malignancy had remained elusive. Gene expression analysis gave the initial insights into the mechanisms by which BCL-6 could act as an oncogene. The first study to identify the targets of BCL-6 repression in B cells employed *in vitro* systems to manipulate BCL-6 expression followed by gene expression analysis DNA microarrays using. A consistent set of target genes was identified (15) repressed when BCL-6 was expressed in BCL-6-negative cells and induced when BCL-6 action was blocked using a dominant-negative form of the factor. As those genes that were.

The list of BCL-6 target genes suggests at least three ways in which BCL-6 could contribute to the development and maintenance of malignant B cells: by repressing inflammation

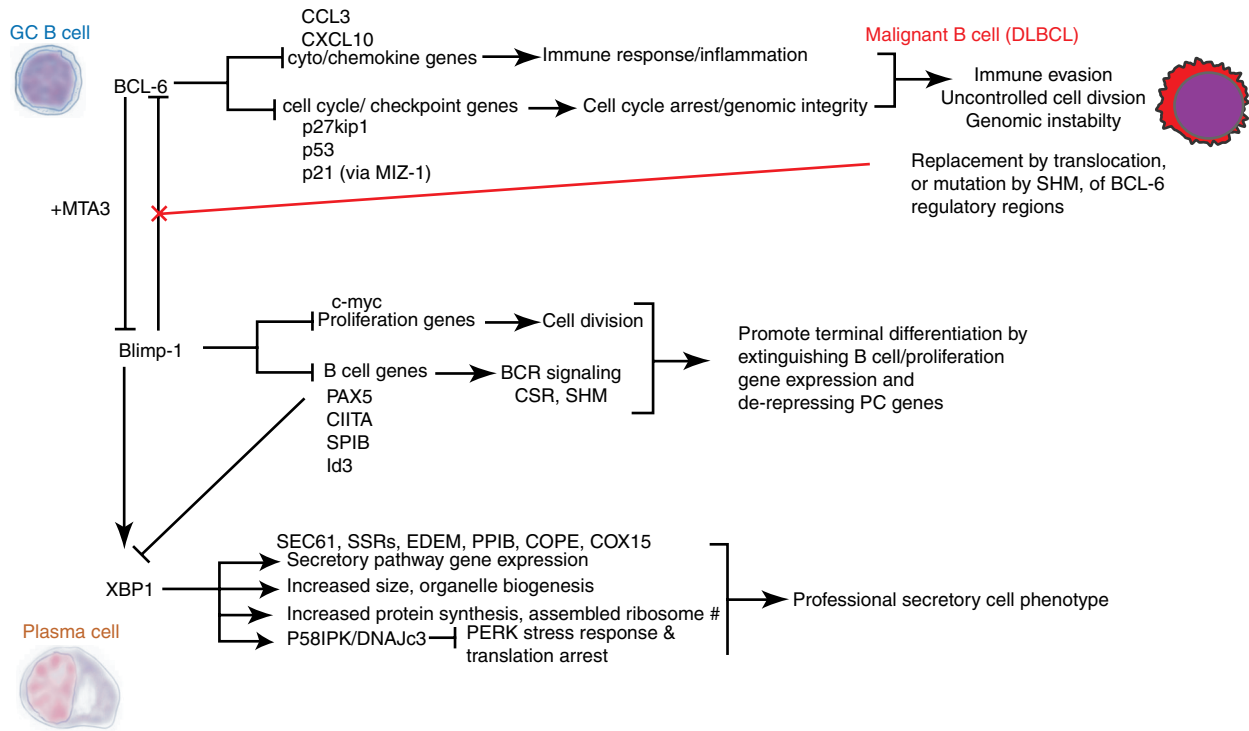


Fig. 2. Transcription factors in B-cell differentiation and disease. A model of the regulatory biology of terminal B-cell differentiation is shown. Recent work has demonstrated that a hierarchy of transcription factors controls the transition of a germinal center B cell (GCB) to a plasma cell. The repressor BCL-6 is highly expressed in GCB cells and suppresses immune response and cell cycle regulatory genes, allowing cells to participate in the germinal center reaction, divide rapidly, and avoid the inadvertent activation of a genomic damage response due to class switching or somatic hypermutation. BCL-6 expression, with its co-repressors MTA3 and SMRT/NcoR, also prevents the premature differentiation of B cells by repressing the expression of Blimp-1. Continued and inappropriate expression of BCL-6, due to promoter mutation or replacement by translocation, leads to sustained cell division and blocks terminal differentiation resulting in the development of diffuse large B-cell

lymphoma (DLBCL). Once BCL-6 expression diminishes in normal GCB cells, Blimp-1 may be expressed and feed back to further repress BCL-6 expression, locking cells into terminal differentiation. Blimp-1 acts as a master controller, extinguishing the gene expression programs of proliferation and mature B cells by repressing other transcription factors (e.g. c-myc, CIITA, SPIB, Id3, and PAX5). Blimp-1 also sets the stage for expression of XBP1 by repressing PAX5, which suppresses expression of the XBP1 gene. XBP1 acts to turn on secretory gene expression and to mediate the physiological changes (expansion of organelles) required for high-level immunoglobulin secretion. XBP1 induction of P58IPK/DNAJc3 also blocks the action of a parallel ER-stress sensing pathway mediated by PERK, which would normally reduce global translation and lessen the protein burden on the ER. Thus, XBP1 expression bestows a professional secretory cell phenotype on terminally differentiated plasma cells.

and immune responses, by interfering with cell cycle inhibitors and checkpoint enforcement, and by preventing terminal differentiation (15) (Fig. 2). BCL-6-knockout mice succumb to a T-helper 2 (Th2)-driven inflammatory disease, and cells derived from these mice overexpress many inflammatory mediators, including interleukin-4 (IL-4), IL-5, IL-10, and IL-13 in T cells (16, 17) and IL-6, IL-18, monocyte chemoattractant protein-1 (MCP-1), MCP-3, and macrophage inflammatory protein-related protein-1 (MRP-1) in macrophages (18–20). The study of BCL-6 targets in B cells showed that the expression of both macrophage inflammatory protein-1 α (MIP-1 α) (CCL3) and interferon-inducible protein-10 (IP-10) (CXCL10) is repressed by BCL-6 in B cells (15). The dysregulated expression of BCL-6 in tumors would thus aid these cells in evading an immune response by preventing the expression of chemokines that recruit macrophages, natural killer (NK) cells, and T cells. BCL-6 also represses the expression of CD80 in B cells, which may also protect tumors by minimizing their interaction with T cells (21) (Fig. 2).

A more direct role for BCL-6 expression in driving B-cell tumorigenesis may stem from its ability to repress cell cycle regulators. Gene expression studies first identified the G1 cell cycle regulator p27kip1 (22) as a target of BCL-6 (15). p27kip1 expression is absent in GCBs, which is consistent with their rapid transit through the cell cycle (23). It is possible that one role for BCL-6 in GCBs is to lower the level of p27kip1 to promote clonal expansion, and the continued expression of BCL-6 may therefore contribute to maintaining rapid cell division in malignant cells. Recent work also ties BCL-6 to the repression of another cdk inhibitor, p21 (24, 25). p21 is not a direct BCL-6 target in that the p21 promoter lacks a BCL-6 binding site. Rather, BCL-6 expression suppresses p21 expression via its association with MIZ1, a homeobox transcription factor (26, 27). MIZ1 binds the p21 promoter and recruits BCL-6, thereby repressing p21 expression. Further work has highlighted the effect of BCL-6 on cell cycle progression in B cells via its ability to directly bind to the p53 gene and repress its expression (28). BCL-6 can therefore suppress the DNA damage checkpoint that could be triggered by genomic alterations caused by somatic hypermutation and class switching in the GCBs (28). Interestingly, BCL-6 can repress replicative senescence in fibroblasts, suggesting that it makes cells unresponsive to the p19ARF/p53 pathway (29). Together, these data suggest that BCL-6 may affect several cell cycle regulatory pathways that allow B cells to escape normal proliferative restrictions (Fig. 2).

Tumor growth can also be promoted by arresting terminal differentiation, and one BCL-6 target gene, B lymphocyte-induced maturation protein-1 (Blimp-1), is particularly remarkable in this regard. Blimp-1 is a master regulator of PC differentiation (30, 31), was repressed in cell lines by BCL-6 expression, and, conversely, blocking BCL-6 expression in B cells led to an increase in Blimp-1 expression and the expression of genes and surface markers characteristic of PC differentiation (32, 33) (Fig. 2). As will be discussed in detail below, the expression of BCL-6 is likely to be a natural mechanism by which GCBs prevent premature terminal differentiation. It is also likely that translocations that deregulate BCL-6 expression prevent cells from moving on to the PC stage, thereby trapping the cells at the GCB cell stage in which rapid proliferation is combined with active mechanisms that remodel the genome (class switching and somatic hypermutation), a recipe for oncogenic disaster.

Although target gene studies have provided clues regarding the function of BCL-6 as an oncogene, recent studies have shown that BCL-6 misexpression in mice can promote lymphomagenesis (34). Mice that have BCL-6 inserted into the immunoglobulin heavy-chain (IgH) locus, mimicking a human BCL-6 translocation, develop a disease very similar to human DLBCL. This model system should be useful in assessing the role of secondary oncogenic events in DLBCL pathogenesis.

The disruption of BCL-6 action by pharmacological agents *in vitro* holds promise as a means to block its many oncogenic effects. The acetylation of BCL-6 has been shown to block its ability to bind co-repressors, thereby blocking its ability to repress its targets (35–37). However, drugs that interfere with the acetylation of BCL-6, such as histone deacetylase inhibitors, are likely to affect numerous cellular targets. A more specific approach to BCL-6 inhibition is the use of a cell-permeable peptide that disrupts the ability of BCL-6 to interact with a co-repressor complex (38). The treatment of cells with this peptide interferes with the ability of the BCL-6 POZ domain to recruit the co-repressors SMRT/NcoR and leads to cell death. This approach could have more specificity for BCL-6 and should be investigated further for its potential in the treatment of lymphomas.

Blimp-1

The gene encoding the transcriptional repressor PRDM1, more popularly known as Blimp-1, was identified as a gene induced during PC differentiation. When transfected into B-cell lines and primary B cells, Blimp-1 drives phenotypic and functional

changes associated with plasmacytic differentiation, including the acquisition of syndecan-1 surface expression and Ig secretion (39–41).

Given the role of Blimp-1 as a master regulator of plasmacytic differentiation, a concerted search was made for Blimp-1 target genes (Fig. 2). A primary aspect of terminal plasmacytic differentiation is exit from the cell cycle. The key growth-regulating transcription factor, *c-myc*, is directly repressed by Blimp-1, and the loss of *c-myc* is instrumental in the withdrawal of PCs from the cell cycle (42, 43). Another hallmark of PC differentiation is the loss of expression of many B cell-restricted genes. Major histocompatibility complex (MHC) class II is typically expressed on B cells and is critical for their cognate interaction with T cells. Blimp-1 directly represses class II transactivator (CIITA), a transcription factor essential for MHC class II expression (44). Blimp-1 also directly represses PAX5, a transcription factor that regulates many B cell-restricted genes and that is essential for commitment to the B-cell lineage (45). By repressing PAX5, Blimp-1 indirectly represses a host of B cell-restricted genes such as CD19 and B-cell linker protein (BLNK). Indeed, Blimp-1-mediated repression of PAX5 is critical for the *in vitro* differentiation of B cells to Ig-secreting cells.

Gene expression profiling was used to further explore the ability of Blimp-1 to drive PC differentiation (31). Unlike expression of BCL-6, which represses a modest number of genes, Blimp-1 expression extinguishes the entire program of mature B-cell gene expression, encompassing hundreds of genes. Most of these effects are indirect and are caused by the direct repression by Blimp-1 of key transcription factors. In addition to *c-myc*, CIITA, and PAX5, Blimp-1 also directly represses the genes encoding the transcription factors SpiB and Id3 (31) (Fig. 2). SpiB, an ets family transcription factor, is critical for signaling and survival in mature B cells (46–48), while Id3 represses E2A, which is a key transcription factor regulating B-cell development, survival, and Ig diversification (49–51). Finally, the role of Blimp-1 as the master regulator of PC differentiation was firmly established with the creation of a B cell-specific knockout of Blimp-1 in mice, which showed no PC development and no serum Ig production in the absence of Blimp-1 (30).

Blimp-1 has a further role in locking post-GCBs into their terminally differentiated fate, which involves its ability to repress, directly or indirectly, BCL-6 expression. The mutual repression of BCL-6 and Blimp-1 creates a double-negative-feedback loop that can drive cells toward the germinal center or the PC phenotype (Fig. 2). As Blimp-1 accumulates in differentiating cells, it represses the expression of its target,

BCL-6. Less BCL-6 allows for more Blimp-1 expression, creating an amplification loop that locks cells into their terminally differentiated state (52). Several independent observations support this reciprocal regulation of differentiation by these factors. First, BCL-6 and Blimp-1 are generally expressed in mutually exclusive B-cell subsets (53, 54). Second, B cells lacking BCL-6 exhibit elevated PC differentiation *in vitro* and *in vivo*, whereas the ectopic expression of BCL-6 blocks *in vitro* PC differentiation (15, 33, 55). Third, transgenic mice in which BCL-6 is dysregulated by insertion in the IgH locus show a partial defect in PC differentiation (34). Fourth, BCL-6 can directly repress Blimp-1 expression by binding transcriptional regulatory elements in the Blimp-1 gene (55). Finally, a remarkable study has shown that the enforced expression of BCL-6, along with a critical co-repressor MTA3, can repress Blimp-1 and cause a plasmacytic cell line to adopt a mature B-cell phenotype (37).

It is likely that this delicate balance between BCL-6 and Blimp-1 in late B-cell development has been exploited by human lymphoid malignancies. As mentioned above, the continued expression of BCL-6 by translocation or mutation is one means by which tumors may avoid terminal differentiation. In PC malignancies, the cell cycle-arresting effects of Blimp-1 may be overcome by multiple mechanisms. In mouse plasmacytomas, the translocation of *c-myc* to the IgH locus is likely to make the rearranged *c-myc* allele resistant to Blimp-1-mediated repression (56, 57). In human multiple myeloma, complex amplifications and rearrangements of the *c-myc* locus may also result in *c-myc* expression despite Blimp-1 expression (58). Thus, while Blimp-1 expression in myeloma may continue to repress other genes involved in B-cell function, such as PAX5 and SpiB, it no longer represses *myc*-driven proliferation and growth. Finally, an alternatively spliced variant of Blimp-1 has been identified that lacks the ability to repress target genes and could function in a dominant-negative fashion (59). Interestingly, a high expression of this splice variant of Blimp-1 has been observed in several PC tumor lines, suggesting that the alteration of Blimp-1 function might represent another way in which PCs can escape the limits of terminal differentiation.

XBPI

A second powerful transcription factor acting during plasmacytic differentiation is X-box binding protein-1 (XBPI), and gene expression profiling again proved useful in unraveling its mode of action (3). XBPI is the mammalian homolog of the

yeast protein Hac1p, which controls the transcriptional response to the accumulation of unfolded proteins in the endoplasmic reticulum (ER). In unstressed cells, a form of XBP1 mRNA is expressed that encodes an unstable transcription factor. When cells experience ER stress, the chaperone protein BiP is recruited to unfolded client proteins in the ER. As a result, the kinase/nuclease IRE1 is no longer bound by BiP and can multimerize and autophosphorylate. This results in the activation of IRE1's intracytoplasmic nuclease domain, which acts on the mRNA of XBP1 to excise a 26-nucleotide internal sequence. After religation, the spliced XBP1 mRNA encodes the stable active form of XBP1 that can activate downstream target genes (60–66).

In mammals, two other signaling pathways also mediate the ER-stress response. Another ER-transmembrane kinase, PERK, is also activated by ER stress and mediates the phosphorylation of eIF2 α , which results in the global repression of protein translation. Phosphorylated eIF2 α also mediates the use of an alternative translation initiation site in the mRNA encoding the transcription factor ATF4. ATF4 then acts to induce ER chaperone proteins and other ER-stress genes (67–69). A third signaling pathway in the ER-stress response involves ATF6, an ER membrane-spanning protein. On ER stress, the cytoplasmic portion of ATF6 is cleaved, and the liberated ATF6 acts as a transcriptional activator that may mediate the induction of some stress-response genes and PC genes (65, 70–74).

The role of XBP1 in B-cell differentiation has been clearly demonstrated through the generation and analysis of XBP1-knockout mice. Mice lacking XBP1 in the B-cell compartment have normal mature B-cell numbers but fail to generate Ig-secreting PCs (75). The introduction of an active form of XBP1 into XBP1-deficient B cells restored their ability to secrete Ig upon differentiation *in vitro* (76).

Gene expression profiling was used to identify the critical targets of XBP1 and Blimp-1 during PC differentiation of mouse B cells (3). B cells from wildtype, XBP1-knockout, and Blimp-1-knockout mice were differentiated using lipopolysaccharide. A discrete set of genes failed to be induced in the absence of Blimp-1 and XBP1. Those that were not induced in the absence of Blimp-1 included XBP1 itself as well as many XBP1 target genes, arguing that XBP1 is functionally downstream of Blimp-1 (3). This may be the result of Blimp-1 repression of PAX5, which itself represses XBP1. Additionally, other regulatory factors such as ATF6 (66) or signal transducer and activator of transcription (STAT) factors (76) may contribute to the transcriptional activation of XBP1.

The identification of XBP1 target genes has revealed the key role that this factor plays in promoting high-level Ig secretion in PCs (Fig. 2). B cells lacking XBP1 did not express known ER-stress-response genes. However, XBP1-deficient B cells were additionally defective in the expression of most genes involved in ER and Golgi function, as well as other secretion-related genes. These XBP1 target genes encode proteins required for protein translocation into the ER, protein folding and glycosylation in the ER, ER protein quality control (i.e. EDEM), and vesicular trafficking. Given the nature of XBP1 target genes, it was not surprising that the expression of XBP1 causes expansion of the ER and Golgi (3). Rather unexpectedly, XBP1 was also found to increase the abundance of other cellular organelles, including mitochondria and lysosomes, and increase cell size, total protein synthesis, and mitochondrial function (3).

These pleiotropic effects of XBP1 all serve to prepare PCs for the high rate of Ig synthesis that is their hallmark. XBP1 is also expressed highly in various exocrine organs such as the pancreas and salivary gland, where it is likely to play a similar role in coordinating protein secretion (77). XBP1 appears to have diverged functionally from its yeast homolog HAC1, which is devoted primarily to responding to ER stress. In this regard, a notable XBP1 target gene is p58IPK/DNAJc3, which encodes p58IPK, an inhibitor of PERK. XBP1 induction of DNAJc3 expression would antagonize the arrest of protein translation that is mediated by PERK, thereby supporting high-level Ig secretion in PCs (3, 70, 78, 79) (Fig. 2). With the evolution of other ER-stress pathways in mammalian cells (PERK/ATF4 and ATF6), it appears that XBP1 has been selected to function as a master regulator of professional secretory cell differentiation (Fig. 2).

As mentioned previously, XBP1 activation is controlled post-transcriptionally by the IRE1-mediated splicing of the XBP1 mRNA, and there remains some controversy as to how this process is initiated in PCs. One hypothesis posits that an initial burst of Ig expression in differentiating cells initiates an ER-stress response that activates IRE1 and leads to XBP1 mRNA processing. This was supported by the demonstration that deletion of the IgH in B cells undergoing plasmacytic differentiation reduces the accumulation of processed XBP1 protein (76). However, the kinetic studies of differentiating B cells suggest that XBP1 processing is initiated before the upregulation of Ig expression (80, 81), suggesting that the post-transcriptional processing of XBP1 is a programmed aspect of plasmacytic differentiation that is at least partially independent of the increase in Ig expression.

A library of gene expression signatures

We have previously discussed the utility of gene expression signatures in extracting biological insights from gene expression profiling data (1). Gene expression signatures are sets of coordinately expressed genes that reflect a particular aspect of cellular biology such as cell lineage, differentiation stage, or activation of a regulatory pathway. We have collated a large number of gene expression signatures that have been defined using a variety of molecular and statistical approaches (SignatureDB, available at: <http://lymphochip.nih.gov/signaturedb/>). Currently, SignatureDB contains 147 gene expression signatures, which are populated with a total of 6078 genes (Table 1). The majority of the curated signatures in SignatureDB are derived from hematopoietic cells, making this database particularly useful for the study of the normal and malignant immune processes.

The raw data for these signatures come from gene expression profiling experiments of two general types: (i) analysis of static gene expression states in sorted cell populations and (ii) analysis of dynamic changes in gene expression induced by various perturbations of cells. Several groups have profiled expression in isolated cell populations representing various lineages and differentiation states and have made these data sets publicly available (82, 83). We have reanalyzed such data by selecting genes that are associated with a particular cell type using standard statistical methods such as the t-test. Other signatures were harvested using the hierarchical clustering algorithm to organize genes according to their expression patterns across all of the samples (84). These signatures sometimes reflect lineage relationships but can also reflect variable biological attributes of cells that cross lineage boundaries. Examples of such signatures include the 'proliferation' and 'quiescence' signatures that reflect progression through the cell cycle and arrest in the G0 stage of the cell cycle, respectively. Other signatures were defined by manipulating lymphocytes *in vitro*, either by exposure to activating stimuli such as cytokines (85) or by treatment with small-molecule inhibitors of specific signaling pathways, such as cyclosporine (86) or NF- κ B inhibitors (87). As detailed earlier in this review, a variety of molecular approaches can be used to define signature genes whose expression levels are contingent on a particular transcription factor.

A seemingly trivial but important value of SignatureDB lies in its curation of gene lists from published articles. Often, a critical gene expression signature is represented in a figure, but the component genes are not provided electronically. As a result, many important results from gene expression profiling

papers languish in obscurity. Furthermore, the static representation of gene names in a figure does not allow more recent and informative gene annotations to be supplied. SignatureDB keeps all gene annotations in sync with the Gene (88) and Unigene (89) systems provided by the National Center for Biotechnology Information.

Although most of the original expression data sets represented in SignatureDB were derived from human cells, many were derived from mouse cells because of the ease with which the mouse genome can be manipulated. For example, the IRF3 target gene signature was generated by analyzing knockout mice for this transcription factor (3, 90). Mouse genetics was also used to develop a signature of regulatory T cells, by expressing green fluorescent protein under the influence of the FOXP3 promoter (91). In SignatureDB, the mouse genes in such signatures are related to their orthologous human genes.

It is critical to remember that signatures are defined in specific contexts and may only translate in part or not at all to gene expression data derived from disparate systems. One caveat *emport* is that orthologous genes need not have the same regulatory behavior in different species. Of course, the same caveat applies to the analysis of a signature outside of the cell types in which it was defined, because regulatory factors can alter gene expression in a highly cell type-specific fashion. For these reasons, failure of a signature to be differentially expressed between two biological specimens is not readily interpretable.

It is important to contrast the content of SignatureDB with other large efforts to annotate gene expression data functionally. The gene ontology (GO) curates published data concerning the function or subcellular localization of gene products (92). Likewise, many efforts have been mounted to collate all proteins reported as components of particular signaling or biochemical pathways. Many of these functionally annotated gene sets have recently been brought together in a publicly available database, MSigDB1.0 (93). In some sense, such gene sets should not be considered 'signatures', because their mRNAs may not be coordinately regulated. In many cases, proteins within a biochemical or signaling pathway may be primarily regulated by post-translational mechanisms, not by the relative abundance of their mRNAs. If only a few (or none) of the genes within a gene set are coordinately regulated as mRNAs, the gene set will have little ability to infer biological mechanisms from gene expression data sets. By contrast, an important feature of SignatureDB is that it focuses solely on gene expression signatures that are comprised of genes that have been shown to be coordinately

Table 1. Gene expression signatures in the SignatureDB

Signature type	Number of signatures	Number of genes	Reference
Transcription factor target			
XBPI targets	3	77	(3)
Blimp-1 targets	2	160	(31)
BCL-6 targets	1	19	(21, 25, 28, 32, 124)
NF-κB targets	5	73	(87)
c-myc targets	1	54	(95)
KLF2 targets	2	217	(125)
IRF3 targets	1	165	(90)
Signaling pathway			
T-cell cytokine signaling	9	170	(85)
T-cell calcium signaling	12	187	(86)
B-cell CD40 signaling	2	174	(109)
Dendritic cell TLR4/8 signaling	1	35	(126)
mTOR pathway	2	416	(127)
KRAS signaling	2	150	(128)
Cellular process			
Proliferation	11	1845	(1, 82, 129–131)
Quiescence	4	270	(82)
B-cell anergy	2	23	(132)
T-cell anergy	1	17	(133)
Immediate early	1	9	(1)
Response to metal ion	1	7	(82)
Fibroblast serum response	2	416	(134)
Amino acid starvation	8	845	(127)
Cellular differentiation			
B cell			
Blood pan-B cell	1	86	(82)
Germinal center B cell	1	388	(95)
Splenic marginal zone B cell	2	62	(104)
Bone marrow plasma cell	1	38	(102)
T/NK cell			
Blood pan-T cell	2	89	(1, 82)
CD4 T-cell differentiation	6	420	(83)
CD8 T-cell differentiation	5	112	(135)
Regulatory T cell	3	47	(91)
Germinal center T-helper cell	5	534	(136, 137)
Blood NK cell	1	69	(82)
Blood T/NK cell	1	22	(82)
Other hematopoietic/immune			
Blood monocyte/myeloid lineage	2	191	(82)
Plasmacytoid dendritic cell	1	92	(82)
Follicular dendritic cell	1	4	(138, 139)
Erythrocytic lineage	2	271	(82)
Bone marrow endothelial precursor	1	47	(82)
Pan-hematopoietic	2	103	(82)
Non-hematopoietic	14	727	(82)
Cancer differential			
Diffuse large B-cell lymphoma	14	972	(98, 102, 108)
Mantle cell lymphoma	2	49	(140)
Follicular lymphoma	2	66	(95)
Chronic lymphocytic leukemia	2	43	(141, 142)
Multiple myeloma	2	176	(143)
Total signatures	147		
Total unique genes		6078	

We have created an online resource, SignatureDB (<http://lymphochip.nih.gov/signaturedb>), by curating public and private gene expression profiles' data for gene expression signatures. A general description of the signature, along with the number of independently derived signatures describing the same phenotype, the number of genes that comprise each signature, and primary references for each signature are shown.

regulated in normal or experimentally manipulated cell types.

We imagine therefore that the signatures within SignatureDB will be more apt to identify informative biological differences within new gene expression data sets.

Defining and mining SignatureDB

Fig. 3 depicts four signatures that illustrate the biological insights that can be gleaned from the gene expression signatures that populate SignatureDB (<http://lymphochip.nih.gov/signaturedb/>).

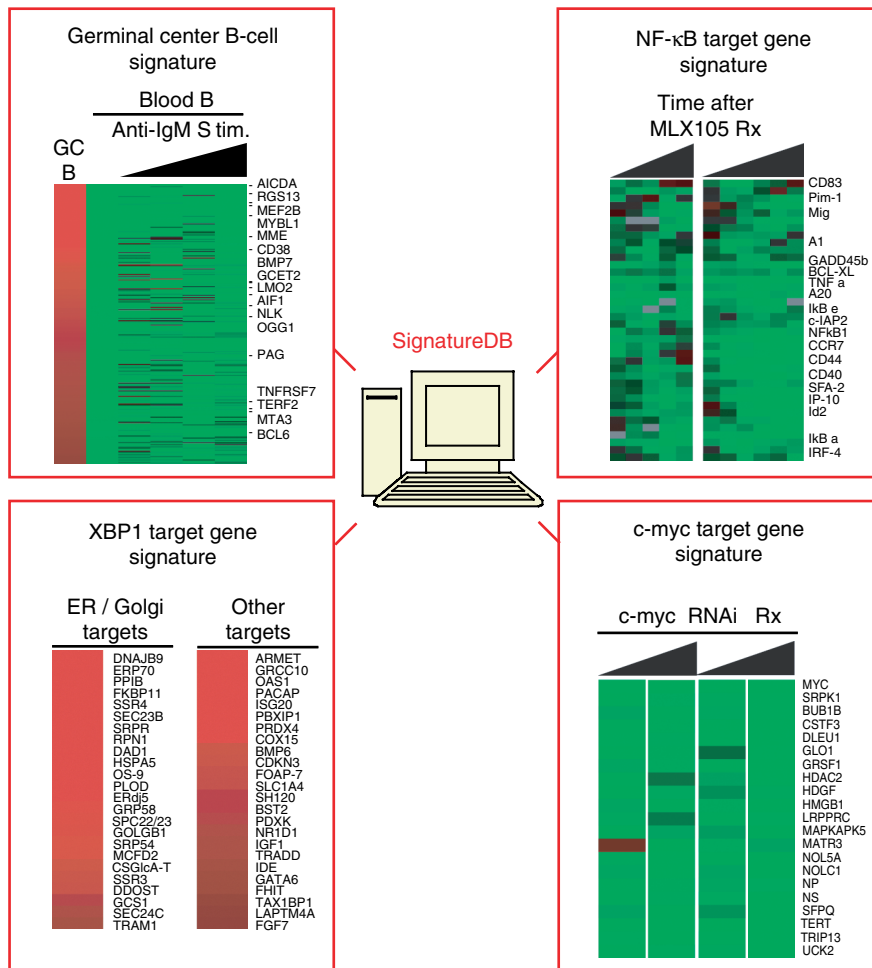


Fig. 3. Defining gene expression signatures: germinal center B cell (GCB), XBP1, nuclear factor-κB (NF-κB), and c-myc.

Gene expression signatures for developmental stage (GCB) and transcription factor-regulated genes (XBP1, NF-κB, and c-myc) were defined by the various approaches described in Fig. 1. The GCB cell signature was defined by comparison of primary B-cell populations and cell lines; the XBP1 signature by expression of the factor in non-expressing cells; the NF-κB signature by the use of a small-molecule inhibitor of IKK β , and the c-myc signature by RNA interference. See text and <http://lymphochip.nih.gov/signaturedb/> for details.

The GCB signature is an example of a signature that defines a particular state of cellular differentiation. The XBP1, NF-κB, and c-myc signatures are examples of signatures that comprise the regulatory targets of particular transcription factors. Each of these latter signatures was defined using a distinct experimental approach, including the overexpression of a transcription factor, pathway inhibition using a small molecule, and knockdown of transcription factor expression using RNAi. Such approaches reveal the molecular wiring of cells and enable the development of a systems biology view of gene expression (Fig. 1).

GCBs are a unique developmental subset that forms in secondary lymphoid organs on encounter with T cell-dependent antigens. Their many distinguishing characteristics include downregulation of surface Ig, proliferation at an exceedingly rapid rate, and initiation of somatic hypermutation and class-switch recombination, enabling affinity maturation of the B-cell receptor and acquisition of new antibody effector functions (94). It is not surprising then that GCBs express a large cohort of genes specific to this stage of differentiation. To define this signature, the gene expression of tonsillar GCBs

was compared with gene expression from resting and anti-IgM-activated blood B cells. GCBs are highly proliferative, but this characteristic is not specific to this cell type. Therefore, the GCB signature was further refined by eliminating proliferation-associated genes that were also induced by IgM cross-linking in primary blood B cells. The final GCB cell signature contains many genes that are known to be upregulated or specifically expressed in GCB cells (95) (Fig. 3), such as CD38, AID, and BCL-6, providing a molecular definition of this stage of B-cell differentiation.

As discussed above, XBP1 controls PC differentiation by regulating the acquisition of the professional secretory cell phenotype. By overexpressing the active form of XBP1 in a non-expressing B-cell line, it was possible to define a list of XBP1 signature genes (3). This list was further focused on genes relevant to plasmacytic differentiation by including only genes more highly expressed in primary human PCs than in primary B-cell populations (3) (Fig. 3). Approximately half of the XBP1 signature is comprised of genes known to be involved in secretory processes – entry into the ER, ER processing, vesicular transport – including SSR4, PPIB, DAD1,

SEC23B, and GOLGB1. The remaining genes may either be involved in secretion in some unknown fashion or be required for the other metabolic changes that accompany upregulation of the secretory pathway (e.g. increased protein synthesis and increased mitochondrial and lysosomal content and function). Although the XBP1 gene expression signature was initially defined in the context of PC differentiation, it also provides a basis for identifying other cells in which XBP1 may promote a secretory phenotype, such as pancreatic cells.

It has been observed that constitutive activation of the NF- κ B pathway is required for survival of the activated B cell-like (ABC) subgroup of DLBCL (96). We have shown that the administration of a small-molecule inhibitor of IKK kinase (IKK), MLX105, is toxic for ABC DLBCL cell lines but not for cell lines derived from the GCB subgroup of DLBCL (87). MLX105 is a member of the β -carboline class of IKK inhibitors (97). MLX105 specifically inhibits the IKK β subunit with an IC₅₀ of 27 nM and showed negligible cross-inhibition of 19 other kinases (87). Moreover, the toxicity of MLX105 for ABC DLBCL cells could be rescued by an inducible form of the p65 subunit of NF- κ B, arguing that this inhibitor has negligible off-target toxicity (87). The treatment of ABC DLBCL cell lines with this inhibitor resulted in a series of rapid gene expression changes that were attributable to cessation of constitutive IKK activity, similar to the changes observed following the acute expression of genetic inhibitors of NF- κ B (87). Several well-known NF- κ B target genes populate this signature, including IKB α , A20, NFKB2, A1, CD44, I-FLICE, IRF4, and tumor necrosis factor α (TNF α), but other genes in this list have not been previously described as NF- κ B targets.

In DLBCL, high expression of the proliferation signature in a tumor biopsy was found to predict unfavorable response to chemotherapy (98). This signature included c-myc, a transcription factor that promotes cell growth and proliferation and whose expression levels are increased in response to mitogenic stimuli. To define a c-myc target gene signature, RNAi-mediated knockdown of c-myc expression was performed by transfecting a pool of four double-stranded RNAi oligonucleotides into an ABC DLBCL cell line, followed by gene expression profiling (95) (Fig. 3). Using this method, we defined a set of c-myc signature genes that were repressed by c-myc RNAi and were correlated with c-myc expression in DLBCL tumor samples. As expected, the c-myc signature includes genes involved in proliferation (e.g. CDK4), metabolism (e.g. enolase and guanine monophosphate synthetase), protein translation (e.g. EIF4G1 and EIF3S9), and nucleolar function (e.g. GNL3, NOLC1, and NOL5A), many of which

have been identified as c-myc targets in other studies (2, 99). This example demonstrates the ability of RNAi directed at a transcription factor to define its downstream targets.

We have begun to use the SignatureDB to mine existing gene expression data sets for novel insights into the biology of B-cell malignancies. The initial question we have asked is whether two cancer types differ in the expression of a signature, and the gene expression data set we have used is derived from a large number of biopsy samples representing each of the two cancer types. Two statistical approaches have proven useful in this effort: signature average analysis and gene set enrichment analysis (GSEA). In signature average analysis, the average expression of the genes in a signature is calculated for each sample, and a t-test is used to identify whether the signature averages differ between two groups of samples. This approach was used to demonstrate that certain gene expression signatures were differentially expressed between subgroups of DLBCL (98).

In contrast, GSEA measures whether the individual genes in a signature are differentially expressed in a consistent fashion between two groups of samples (100). GSEA begins with a ranking of all of the genes in a microarray data set based on the difference in their expression levels between the two sample sets. Typically, the T-statistic from the t-test is used as the ranking parameter. Next, the position of each signature gene within this ranked list is determined, and the degree to which the signature genes are bunched near the top (or bottom) of the ranked list is measured using a Kolmogorov-Smirnoff (KS) statistic. In the null hypothesis, the signature genes will be evenly distributed in the ranked list. The probability that a result is significant in GSEA is estimated by repeating the entire procedure many times using permuted data sets in which each sample is randomly assigned to one of the two cancer types being investigated. A nominal P-value is assigned based on the frequency with which permuted data give a KS statistic that is larger than that obtained with the original unpermuted data. This method was used to demonstrate that genes involved in oxidative phosphorylation are decreased in expression in diabetic muscle (100). A recent enhancement of the GSEA method places added emphasis on those genes in a signature that are the most differentially expressed between two groups (93).

Signature average analysis and GSEA test subtly different hypotheses regarding the expression of a signature in two groups. In signature average analysis, it is not necessary that all the genes in the signature be among the most differentially expressed, just that the signature average is different between

the two groups. In contrast, GSEA assesses the extent to which the genes in the signature appear consistently among those genes that are most differentially expressed. If only a subset of the signature genes is differentially expressed between two groups, GSEA is less likely to produce a significant result. Even if all of the signature genes are differentially expressed between the groups, GSEA will not yield a significant result if there is a large set of non-signature genes that are ranked higher in the list of differentially expressed genes.

Fig. 4 shows how gene expression signature analysis can be used to tease out biological differences between the ABC and GCB subgroups of DLBCL. Four signatures from SignatureDB that showed significant differences between the DLBCL subgroups by both signature average analysis ($P < 0.0001$) and GSEA ($P < 0.05$) are shown. For the signature average analysis, the signature average is displayed for each tumor sample among a collection of ABC and GCB DLBCLs, with red indicating relatively high expression and green relatively low expression. A bar graph is also provided that displays the mean of the signature averages within each DLBCL subgroup along with the P -value for the significance of the distinction. For the GSEA, the T -statistic that measures differential expression between the DLBCL subgroups is plotted for each gene on the microarray in rank order, generating the sigmoidal curve shown (Fig. 4). The position of each signature gene within

this ranked list of all genes is indicated along with a nominal P -value for the significance to which the signature genes are bunched near the top or bottom of the ranked list. Both the signature average analysis and GSEA methods revealed that these four signatures are differentially expressed between ABC DLBCL and GCB DLBCL.

At first glance, the differential expression of the GCB signature between ABC DLBCL and GCB DLBCL may appear circular, because this distinction was evident when these subgroups were originally defined (98, 101). However, the Lymphochip cDNA microarray that was used in the original studies measures the expression of roughly 5000 genes, whereas the Affymetrix U133+ microarray used in the current analysis measures the expression of over 25 000 genes. Consequently, the GCB cell signature has grown from 64 genes to 352 genes. The current GCB cell signature was derived by comparing the gene expression profiles of tonsillar GCBs with those of resting blood B cells and anti-IgM-stimulated blood B cells (95). Well-known genes that affect germinal center function, such as BCL-6 (8, 9) and MTA3 (37), are found in this signature, but the impressive breadth of this signature illustrates how little is known of the biology of GCBs.

While these data demonstrate that GCB DLBCL arises from normal GCBs, the cell of origin of ABC DLBCL remains a

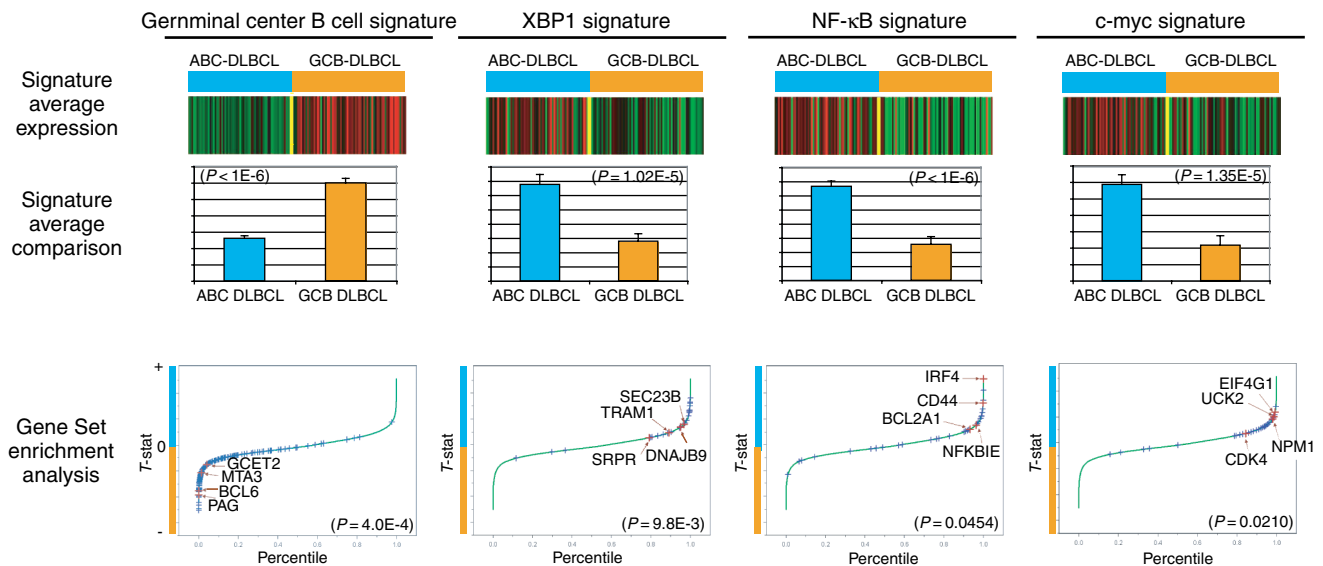


Fig. 4. Signature analysis of activated B cell-like (ABC) versus germinal center B-cell (GCB) diffuse large B-cell lymphoma (DLBCL). Gene expression signatures defined by independent experiments (Fig. 3 and <http://lymphochip.nih.gov/signaturedb>) were used to compare tumors representing two subtypes of DLBCL (ABC versus GCB, see text). Both signature average comparison ($P < 0.0001$) and gene set enrichment analysis (GSEA, $P < 0.05$) identified signatures that distinguished ABC DLBCL from GCB DLBCL. Shown for each signature are the

signature average expression, the mean signature expression in each tumor type, the P -value for the difference in the average expression of the signature, the GSEA Kolmogorov–Smirnov P -value, and a plot of the T -statistic for each gene versus the proportion of genes with a higher T -statistic (percentile). On the T -statistic plot, positive values identify genes expressed higher in ABC DLBCL, and negative values indicate genes higher in GCB DLBCL. Blue crosses show genes from a given signature, and red crosses identify exemplar genes from the signature.

mystery. ABC DLBCLs have mutated Ig genes but do not have ongoing somatic hypermutation, as is the case with GCB DLBCLs (101). Therefore, one possibility is that ABC DLBCL arises from a post-germinal center cell whose differentiation toward a PC has been interrupted by the mechanisms of malignant transformation. In line with this possibility, ABC DLBCLs express some genes typical of PCs at higher levels than GCB DLBCLs, particularly those that are targets of XBP1 (102). By both signature average analysis and GSEA, it is clear that ABC DLBCLs express XBP1 target genes at higher levels than GCB DLBCLs (Fig. 4). As detailed above, this signature includes many genes that encode regulators of the ER-stress response and secretion, such as DNAJB9, SEC23B, TRAM1, and SRPR (3). Thus, gene expression signature analysis revealed a previously unknown aspect of the regulatory biology of ABC DLBCL, namely activation of the ‘physiological’ unfolded protein response that is coordinated by XBP1 (3).

These data support the hypothesis that ABC DLBCL arises from a B cell that has differentiated partially toward a PC. However, many other hallmarks of PC differentiation are missing from ABC DLBCLs, such as the high expression of syndecan-1 and downmodulation of mature B-cell genes. This later observation may suggest that the action of Blimp-1 is suppressed or altered in ABC DLBCLs. Because ABC DLBCLs have mutated Ig genes, the AID-dependent somatic hypermutational mechanism was activated sometime during their natural history. Although this may imply that ABC DLBCLs arise from B cells that have transited the germinal center, some somatic hypermutation can occur in the absence of CD40/CD40 ligand interactions and outside of the germinal center (103–106). Further work is needed therefore to illuminate the origin of ABC DLBCL more precisely.

ABC DLBCLs have constitutive activation of the nuclear factor- κ B (NF- κ B) pathway due to constitutive IKK activation, and this pathway is required for their survival (87, 96). This conclusion was initially based on the observation that a small set of published NF- κ B target genes is more highly expressed in ABC DLBCL than in GCB DLBCL (96). As detailed above, we have used a small-molecule inhibitor of IKK to define a more comprehensive list of NF- κ B target genes (87), and this signature is differentially expressed between ABC DLBCL and GCB DLBCL, according to both signature average analysis and GSEA (Fig. 4). This list of NF- κ B targets can be used to demonstrate that other B-cell malignancies have constitutive activity of the NF- κ B pathway, such as primary mediastinal B-cell lymphoma (87, 107, 108). Conversely, we have used this signature to demonstrate that Burkitt lymphoma is particularly deficient in the expression of NF- κ B target genes (95).

Burkitt lymphomas arise from GCBs, which have exceedingly low expression of NF- κ B target genes compared B with cells at other stages of differentiation (1, 95, 109). Burkitt lymphoma is characterized by a higher rate of spontaneous apoptosis and is highly curable using intensified chemotherapy regimens. Given the potent ability of NF- κ B to suppress apoptosis, these two phenotypes might reflect the low NF- κ B activity in Burkitt lymphomas (95).

GCBs proliferate at an astonishing rate *in vivo*, yet they express very low levels of c-myc and its target genes (1, 110). This is a surprising observation, because proliferation is associated with elevated c-myc expression in most cell types. Although c-myc is likely to play a role in cell cycle progression, much of its regulatory biology is geared to enhancing cell growth and metabolism (111). Therefore, low c-myc levels in normal GCBs may be a mechanism to devote the majority of cellular energy to cell cycle progression instead of cell growth. As presented above, we have defined a signature of c-myc target genes using RNAi (95) and find that ABC DLBCL has a higher expression of this signature than GCB DLBCL, using both signature average analysis and GSEA (Fig. 4). The reasons for the higher activity of c-myc in ABC DLBCL are unclear, but it is well known that a variety of signaling pathways can induce c-myc, including NF- κ B. The expression of c-myc and some of its target genes (e.g. GNL3 and NPM3) in DLBCL tumors is associated with poor outcome following chemotherapy (98). This survival association is due, in part, to the comparatively high expression of this signature in the tumors of patients with ABC DLBCL, who have a worse overall survival than patients with GCB DLBCL (98, 101, 102).

These examples indicate how SignatureDB could prove useful in the understanding of disease pathogenesis. Biopsies or peripheral blood samples contain a variety of cell types, and this heterogeneity can be measured using the cell type-specific gene expression signatures in SignatureDB. Furthermore, the pathological activity of signaling pathways in a disease process can be inferred by the expression of pathway-specific gene expression signatures. There are, of course, limitations to this methodology. For example, some gene expression signatures may be expressed in an overlapping fashion in multiple cell types within a clinical sample, confounding attempts to clearly resolve the role of the signature in the disease process. This is a knotty problem that may sometimes require the use of cell separation technologies to help deconvolute the gene expression profiles of clinical samples. A further complexity is provided by the cross-regulation of many genes by different transcription factors. Despite these caveats, it is likely that the

power of gene expression signature analysis to illuminate disease biology will grow as additional curated gene expression signatures are added to SignatureDB.

Achilles' heel RNAi screens

RNAi has emerged as one of the most powerful genetic tools for the dissection of gene-regulatory networks. The implementation of RNAi technology in mammalian cells by either short interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs) has made the selective knockdown of gene expression possible (112–116). The transfection of double-stranded (siRNA) oligonucleotides leads to a transient knockdown of gene expression, making it difficult to observe phenotypic changes for longer than 2–3 days in most instances. By contrast, shRNAs can be stably expressed in cells, leading to an extended knockdown of gene expression and the possibility to observe phenotypic changes over several weeks. As described below, we have used shRNA technology to perform genetic screens for genes required for the proliferation and survival of human lymphoma cells.

The design of RNAi experiments must take into account several potential limitations of this technology (116). The first hurdle in these experiments is the identification of si/shRNA sequences that effectively reduce target gene expression. Although advances have been made in defining the rules for the design of RNAi sequences, many iterations are often needed to obtain an acceptable degree of gene knockdown. In our experience with shRNAs, it is often possible to identify an shRNA that knocks down gene expression by more than 50% among three candidate sequences. However, for some genes, 10–20 shRNAs need to be screened to find one that is effective. Second, it is important to verify that the phenotype elicited by an si/shRNA is due to the knockdown of the targeted gene and not to off-target effects. It is therefore important to observe the same phenotypic changes using a second si/shRNA targeting the same mRNA. Perhaps the best control is to target the 3' or 5' UTR of an mRNA with an si/shRNA and then demonstrate that the phenotypic effects of the si/shRNA can be reversed by co-expression of a cDNA containing the coding region but lacking the UTR of the targeted gene. Third, the persistence of the knockdown of gene expression is also critical in the design and interpretation of RNAi experiments. As mentioned previously, siRNAs typically knockdown genes for shorter periods of time than shRNAs. However, we have also observed a waning in the effect of shRNAs over several weeks in some instances, particularly when the shRNA confers a toxic phenotype. Finally,

controversy exists as to whether the introduction of si/shRNAs into cells *in vitro* or *in vivo* evokes an unwanted, Toll-like receptor-based innate immune response (reviewed in 117). Such off-target effects appear to be cell type specific; we have introduced shRNAs into B-cell lines and profiled the attendant gene expression changes using microarrays and thus far have not observed induction of interferon response genes, which can be induced by Toll-like receptor signaling in response to double-stranded RNAs.

One of the most exciting prospects of RNAi technology is the possibility of mammalian somatic cell genetics. In this approach, libraries are created to express si/shRNAs targeting thousands of genes, and screens are conducted to identify which si/shRNAs target genes that control a specific cellular phenotype. Two retroviral shRNA expression libraries have been successfully developed by the Bernards and Hannon groups (118, 119). To monitor the abundance of each shRNA within the population of retrovirally transduced cells, each group developed library-specific DNA microarrays. For the Bernards library, the microarray consists of the shRNA sequences themselves. The Hannon group included a random 60-bp sequence in each shRNA vector and generated a microarray of these so-called molecular 'bar code' sequences. Both shRNA libraries have been used successfully to screen for phenotypes in cancer cells, leading to the discovery of five targets of p53 that regulate p53-dependent cell cycle arrest (118) and a new tumor suppressor gene for the phosphatidylinositol-3 kinase (PI3K) pathway (120). Both of these screens employed a 'positive' selection experimental scheme: a selective pressure was exerted that prevented cells from growing and/or surviving, and shRNAs that allowed cells to circumvent these inhibitory constraints were positively selected.

We have developed an shRNA library to identify genes that are required for abnormal proliferation and survival of cancer cells, two of the hallmarks of the malignant phenotype (121). Because such genes might encode potential anti-cancer drug targets, we refer to this approach as an Achilles' heel RNAi screen (Fig. 5). To identify shRNAs that knockdown the expression of proliferation or survival genes, we needed to develop an inducible system so that the expression of such toxic shRNAs could be switched on at will, which we have accomplished using tetracycline (TET) repressor technology (122). Our library vector is based on pRetro-Super, which expresses an shRNA using the histone H1 polIII promoter (113, 123). We modified this vector by cloning two binding sites for the bacterial TET repressor into the H1 promoter, as previously described (122). This TET-inducible,

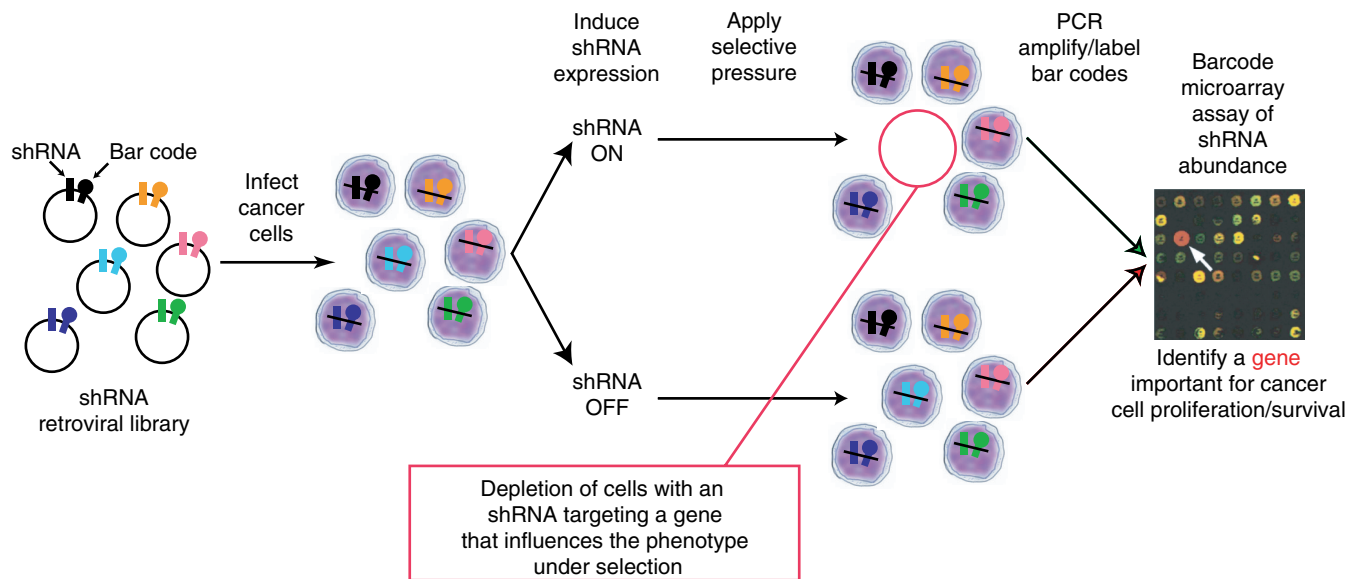


Fig. 5. An Achilles' heel-inducible short hairpin RNA (shRNA) library screen for genes controlling cancer phenotypes. Infection of an shRNA retroviral library into a cancer cell line produces a 'library' with each cell carrying one or more shRNAs. Each shRNA is tagged by a known, unique bar code. Infected cells are divided into two subpopulations: one induced for expression of shRNAs and the other serving as the control. The inducibility of the shRNA library is important to prevent the loss of shRNA

species that are acutely deleterious to infected cells. Abundance in the library of an shRNA targeting a gene that influences the phenotype under selection can be quantified by comparing its relative abundance in the two subpopulations. Genomic DNA fragments carrying the bar code representing each shRNA are amplified by PCR from each subpopulation, labeled with different fluorescent dyes, and co-hybridized to a microarray containing complementary bar code oligonucleotides.

shRNA-expressing retrovirus is used to infect target cells that are engineered to constitutively express the TET repressor, which suppresses the expression of the shRNA until TET or doxycycline is added. Using this vector, we have created a library of shRNAs that target over 2500 human genes with three to six different shRNAs for each gene. The library targets genes encoding all protein kinases, PI3Ks, and deubiquitinating enzymes along with most known regulators of the NF- κ B pathway. In addition, the library targets hundreds of genes which are differentially expressed between lymphoma types but whose function is not clear. The vectors in the library have all been sequence verified, and each includes a different 60-bp molecular bar code sequence.

The experimental outline for an Achilles' heel RNAi screen is shown in Fig. 5. The shRNA library plasmids are turned into a retroviral stock that is used to infect a cancer cell line. On average, each cell within the population is transduced with a single retroviral vector, and therefore, on induction of shRNA expression, a single gene will be knocked down in expression in each cell. After infected cells are selected for the presence of the shRNA library, the cells are divided into two subpopulations, one that is induced for shRNA expression by doxycycline addition and the other serving as the uninduced negative control. After doxycycline addition, cells that harbor an shRNA that knocks down a gene required for proliferation

or survival will disappear from the infected population over the course of days to weeks. Genomic DNA is harvested from the induced and uninduced cells, and the shRNA and bar code sequences are retrieved by polymerase chain reaction (PCR) amplification. The PCR products from uninduced and induced cultures are labeled with separate fluorochromes and co-hybridized to a microarray containing the 60-mer bar code sequences. Scanning the microarray yields the relative fluorescent signal intensities, which reflect the relative abundance of each shRNA in the uninduced and induced cultures (Fig. 5).

Using this shRNA library, we have performed Achilles' heel screens using cell lines that are models for ABC DLBCL and GCB DLBCL (144). As expected, a number of shRNAs were identified that affected the growth and survival of both ABC DLBCL and GCB DLBCL cell lines. These shRNAs targeted genes that encode various cell cycle regulators and general transcription and splicing factors. More importantly, we have identified several shRNAs that are selectively toxic for ABC DLBCL cell lines or for GCB DLBCL cell lines. For instance, in screens of ABC DLBCLs, we have isolated two independent shRNAs that target the gene encoding the IKK β subunit, but these shRNAs were not identified in screens of GCB DLBCL cell lines. This result is consistent with our previous finding that IKK activity is required for the survival of ABC DLBCLs

but not GCB DLBCLs (96). These proof-of-principle data demonstrate the ability of this approach to identify regulatory pathways that control the proliferation and survival of specific cancer cell types. Such pathways could serve as targets for the development of drugs that are tailored for particular cancer types. We envision a comprehensive effort to perform Achilles' heel screens in a large panel of cancer cell lines. This effort could yield a functional classification of cancer that will ultimately be used to choose the appropriate targeted therapy for each patient's cancer.

Concluding remarks

As illustrated in this review, each genomics technology provides a different vantage point from which to view the function of normal and malignant lymphocytes. In the future, the fusion of these technologies will provide the deepest insights.

For example, an Achilles' heel RNAi screen in a lymphoma cell line might yield a potent shRNA specific to a particular regulatory factor. A gene expression signature of this regulatory factor could then be generated by inducing the expression of this shRNA and monitoring changes in the gene expression using DNA microarrays. In parallel, ChIP-on-chip experiments could be used to determine which genes in this signature are bound by the regulatory factor *in vivo* and are thus direct target genes of the factor. This refined gene expression signature could be used to identify which lymphomas have the activity of pathway and thus which patients might benefit from therapies targeting the pathway. Although this specific example illustrates the potential utility of this 'parallel processing' approach in cancer, these methods could also be used to deconvolute the complexities of autoimmune, inflammatory, and infectious disease processes.

References

- Shaffer AL, et al. Signatures of the immune response. *Immunity* 2001;**15**:375–385.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005;**37**:382–390.
- Shaffer AL, et al. XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* 2004;**21**:81–93.
- Ren B, et al. Genome-wide location and function of DNA binding proteins. *Science* 2000;**290**:2306–2309.
- Shaffer AL, Rosenwald A, Staudt LM. Lymphoid malignancies: the dark side of B-cell differentiation. *Nat Rev Immunol* 2002;**2**:920–932.
- Staudt LM, Dent AL, Shaffer AL, Yu X. Regulation of lymphocyte cell fate decisions and lymphomagenesis by BCL-6. *Int Rev Immunol* 1999;**18**:381–403.
- Dalla-Favera R, et al. Molecular pathogenesis of B cell malignancy: the role of BCL-6. *Curr Top Microbiol Immunol* 1999;**246**: 257–263.
- Ye BH, et al. The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. *Nat Genet* 1997;**16**:161–170.
- 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.
- Fukuda T, et al. Disruption of the Bcl6 gene results in an impaired germinal center formation. *J Exp Med* 1997;**186**:439–448.
- Baron BW, Nucifora G, McCabe N, Espinosa R, Le Beau MM, McKeithan TW. Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas. *Proc Natl Acad Sci USA* 1993;**90**:5262–5266.
- Ye BH, et al. Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. *EMBO J* 1995;**14**:6209–6217.
- Gaidano G, et al. Aberrant somatic hypermutation in multiple subtypes of AIDS-associated non-Hodgkin lymphoma. *Blood* 2003;**102**:1833–1841.
- Pasqualucci L, Migliazza A, Basso K, Houldsworth J, Chaganti RS, Dalla-Favera R. Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* 2003;**101**:2914–2923.
- Shaffer AL, Yu X, He Y, Boldrick J, Chan E, Staudt L. BCL-6 represses genes that function in lymphocyte differentiation, inflammation and cell cycle control. *Immunity* 2000;**13**:199–212.
- Arima M, et al. A putative silencer element in the IL-5 gene recognized by Bcl6. *J Immunol* 2002;**169**:829–836.
- Kusam S, Toney LM, Sato H, Dent AL. Inhibition of Th2 differentiation and GATA-3 expression by BCL-6. *J Immunol* 2003;**170**:2435–2441.
- Takeda N, et al. Bcl6 is a transcriptional repressor for the IL-18 gene. *J Immunol* 2003;**171**:426–431.
- Yu RY, et al. BCL-6 negatively regulates macrophage proliferation by suppressing autocrine IL-6 production. *Blood* 2005;**105**:1777–1784.
- Toney LM, et al. BCL-6 regulates chemokine gene transcription in macrophages. *Nat Immunol* 2000;**1**:214–220.
- Niu H, Cattoretti G, Dalla-Favera R. BCL6 controls the expression of the B7-1/CD80 costimulatory receptor in germinal center B cells. *J Exp Med* 2003;**198**:211–221.
- Sherr CJ, Roberts JM. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* 1999;**13**:1501–1512.
- Sanchez-Beato M, et al. Cyclin-dependent kinase inhibitor p27KIP1 in lymphoid tissue: p27KIP1 expression is inversely proportional to the proliferative index. *Am J Pathol* 1997;**151**:151–160.
- Dalla-Favera R. FASEB Summer Research Conference, Lymphocytes & the Immune System: Molecular, Cellular & Integrative Mechanisms, Tucson, AZ, July 9–14, 2005.
- Phan RT, Saito M, Basso K, Niu H, Dalla-Favera R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nat Immunol* 2005;**6**:1054–1060.

26. Wu S, et al. Myc represses differentiation-induced p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter. *Oncogene* 2003;**22**:351–360.
27. Wu L, et al. Miz1, a novel zinc finger transcription factor that interacts with Msx2 and enhances its affinity for DNA. *Mech Dev* 1997;**65**:3–17.
28. Phan RT, Dalla-Favera R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* 2004;**432**:635–639.
29. Shvarts A, et al. A senescence rescue screen identifies BCL6 as an inhibitor of anti-proliferative p19(ARF)-p53 signaling. *Genes Dev* 2002;**16**:681–686.
30. Shapiro-Shelef M, Lin KL, McHeyzer-Williams LJ, Liao J, McHeyzer-Williams MG, Calame K. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* 2003;**19**:607–620.
31. Shaffer AL, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* 2002;**17**:51–62.
32. Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity* 2000;**13**:199–212.
33. Reljic R, Wagner SD, Peakman LJ, Fearon DT. Suppression of signal transducer and activator of transcription 3-dependent B lymphocyte terminal differentiation by BCL-6. *J Exp Med* 2000;**192**:1841–1848.
34. Cattoretti G, et al. Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer Cell* 2005;**7**:445–455.
35. Bereshchenko OR, Gu W, Dalla-Favera R. Acetylation inactivates the transcriptional repressor BCL6. *Nat Genet* 2002;**32**:606–613.
36. Dhordain P, et al. The LAZ3(BCL-6) oncoprotein recruits a SMRT/mSIN3A/histone deacetylase containing complex to mediate transcriptional repression. *Nucleic Acids Res* 1998;**26**:4645–4651.
37. Fujita N, et al. MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* 2004;**119**:75–86.
38. Polo JM, et al. Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat Med* 2004;**10**:1329–1335.
39. Schliephake DE, Schimpl A. Blimp-1 overcomes the block in IgM secretion in lipopolysaccharide/anti-mu F(ab')₂-co-stimulated B lymphocytes. *Eur J Immunol* 1996;**26**:268–271.
40. Turner CA Jr, Mack DH, Davis MM. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* 1994;**77**:297–306.
41. Messika EJ, et al. Differential effect of B lymphocyte-induced maturation protein (Blimp-1) expression on cell fate during B cell development. *J Exp Med* 1998;**188**:515–525.
42. Lin Y, Wong K, Calame K. Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science* 1997;**276**:596–599.
43. Lin KI, Lin Y, Calame K. Repression of c-myc is necessary but not sufficient for terminal differentiation of B lymphocytes in vitro. *Mol Cell Biol* 2000;**20**:8684–8695.
44. Lin K-I, Angelin-Duclos C, Kuo TC, Calame K. Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to IgM secreting plasma cells. *Mol Cell Biol* 2002;**22**:4771–4780.
45. Piskurich JF, Lin KI, Lin Y, Wang Y, Ting JP, Calame K. BLIMP-1 mediates extinction of major histocompatibility class II transactivator expression in plasma cells. *Nat Immunol* 2000;**1**:526–532.
46. Brass AL, Kehrl E, Eisenbeis CF, Storb U, Singh H. Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1. *Genes Dev* 1996;**10**:2335–2347.
47. Garrett-Sinha LA, et al. PU.1 and Spi-B are required for normal B cell receptor-mediated signal transduction. *Immunity* 1999;**10**:399–408.
48. Hu CJ, et al. PU.1/Spi-B regulation of c-rel is essential for mature B cell survival. *Immunity* 2001;**15**:545–555.
49. Kee BL, Rivera RR, Murre C. Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF-beta. *Nat Immunol* 2001;**2**:242–247.
50. Quong MW, Harris DP, Swain SL, Murre C. E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination. *EMBO J* 1999;**18**:6307–6318.
51. Sayegh CE, Quong MW, Agata Y, Murre C. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol* 2003;**4**:586–593.
52. Staudt LM. Cancer: negative feedback for B cells. *Nature* 2004;**431**:919–920.
53. Angelin-Duclos C, Cattoretti G, Lin KI, Calame K. Commitment of B lymphocytes to a plasma cell fate is associated with Blimp-1 expression in vivo. *J Immunol* 2000;**165**:5462–5471.
54. Cattoretti G, Angelin-Duclos C, Shaknovich R, Zhou H, Wang D, Alobeid B. PRDM1/Blimp-1 is expressed in human B-lymphocytes committed to the plasma cell lineage. *J Pathol* 2005;**206**:76–86.
55. Tunyaplin C, Shaffer AL, Angelin-Duclos CD, Yu X, Staudt LM, Calame KL. Direct repression of prdm1 by Bcl-6 inhibits plasmacytic differentiation. *J Immunol* 2004;**173**:1158–1165.
56. Park SS, et al. Insertion of c-Myc into IgH induces B-cell and plasma-cell neoplasms in mice. *Cancer Res* 2005;**65**:1306–1315.
57. Potter M. Neoplastic development in plasma cells. *Immunol Rev* 2003;**194**:177–195.
58. Kuehl WM, Bergsagel PL. Multiple myeloma: evolving genetic events and host interactions. *Nat Rev Cancer* 2002;**2**:175–187.
59. Gyory I, Fejer G, Ghosh N, Seto E, Wright KL. Identification of a functionally impaired positive regulatory domain I binding factor 1 transcription repressor in myeloma cell lines. *J Immunol* 2003;**170**:3125–3133.
60. Reimold AM, et al. Transcription factor B cell lineage-specific activator protein regulates the gene for human X-box binding protein 1. *J Exp Med* 1996;**183**:393–401.
61. Calfon M, et al. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* 2002;**415**:92–96.
62. Fewell SW, Travers KJ, Weissman JS, Brodsky JL. The action of molecular chaperones in the early secretory pathway. *Annu Rev Genet* 2001;**35**:149–191.
63. Harding HP, et al. Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell* 2000;**6**:1099–1108.
64. Mori K, Ogawa N, Kawahara T, Yanagi H, Yura T. mRNA splicing-mediated C-terminal replacement of transcription factor Hac1p is required for efficient activation of the unfolded protein response. *Proc Natl Acad Sci USA* 2000;**97**:4660–4665.
65. Shen X, et al. Complementary signaling pathways regulate the unfolded protein response and are required for *C. elegans* development. *Cell* 2001;**107**:893–903.
66. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 2001;**107**:881–891.
67. Harding HP, Zhang Y, Ron D. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* 1999;**397**:271–274.
68. Harding HP, et al. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* 2003;**11**:619–633.

69. Harding HP, et al. Translational regulation in the cellular response to biosynthetic load on the endoplasmic reticulum. *Cold Spring Harb Symp Quant Biol* 2001;**66**:499–508.
70. Lee AH, Iwakoshi NN, Glimcher LH. XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol Cell Biol* 2003;**23**:7448–7459.
71. Okada T, Yoshida H, Akazawa R, Negishi M, Mori K. Distinct roles of activating transcription factor 6 (ATF6) and double-stranded RNA-activated protein kinase-like endoplasmic reticulum kinase (PERK) in transcription during the mammalian unfolded protein response. *Biochem J* 2002;**366**:585–594.
72. Yoshida H, et al. ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response. *Mol Cell Biol* 2000;**20**:6755–6767.
73. Yoshida H, Matsui T, Hosokawa N, Kaufman RJ, Nagata K, Mori K. A time-dependent phase shift in the mammalian unfolded protein response. *Dev Cell* 2004;**4**:265–271.
74. Gunn KE, Gifford NM, Mori K, Brewer JW. A role for the unfolded protein response in optimizing antibody secretion. *Mol Immunol* 2004;**41**:919–927.
75. Reimold AM, et al. Plasma cell differentiation requires the transcription factor XBP-1. *Nature* 2001;**412**:300–307.
76. Iwakoshi NN, Lee AH, Vallabhajosyula P, Otipoby KL, Rajewsky K, Glimcher LH. Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. *Nat Immunol* 2003;**4**:321–329.
77. Clauss IM, Gravalles EM, Darling JM, Shapiro F, Glimcher MJ, Glimcher LH. In situ hybridization studies suggest a role for the basic region-leucine zipper protein hXBP-1 in exocrine gland and skeletal development during mouse embryogenesis. *Dev Dyn* 1993;**197**:146–156.
78. Yan W, et al. Control of PERK eIF2alpha kinase activity by the endoplasmic reticulum stress-induced molecular chaperone P58IPK. *Proc Natl Acad Sci USA* 2002;**99**:15920–15925.
79. van Huizen R, Martindale JL, Gorospe M, Holbrook NJ. P58IPK, a novel endoplasmic reticulum stress-inducible protein and potential negative regulator of eIF2alpha signaling. *J Biol Chem* 2003;**278**:15558–15564.
80. van Anken E, et al. Sequential waves of functionally related proteins are expressed when B cells prepare for antibody secretion. *Immunity* 2003;**18**:243–253.
81. Gass JN, Gifford NM, Brewer JW. Activation of an unfolded protein response during differentiation of antibody-secreting B cells. *J Biol Chem* 2002;**277**:49047–49054.
82. Su AI, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004;**101**:6062–6067.
83. Lee MS, Hanspers K, Barker CS, Korn AP, McCune JM. Gene expression profiles during human CD4⁺ T cell differentiation. *Int Immunol* 2004;**16**:1109–1124.
84. 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.
85. Kovanen PE, et al. Analysis of gamma c-family cytokine target genes. Identification of dual-specificity phosphatase 5 (DUSP5) as a regulator of mitogen-activated protein kinase activity in interleukin-2 signaling. *J Biol Chem* 2003;**278**:5205–5213.
86. Feske S, Giltman J, Dolmetsch R, Staudt LM, Rao A. Gene regulation mediated by calcium signals in T lymphocytes. *Nat Immunol* 2001;**2**:316–324.
87. Lam LT, et al. Small molecule inhibitors of IkappaB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin Cancer Res* 2005;**11**:28–40.
88. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2005;**33**:D54–D58.
89. Schuler GD. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med* 1997;**75**:694–698.
90. Ogawa S, et al. Molecular determinants of crosstalk between nuclear receptors and Toll-like receptors. *Cell* 2005;**122**:707–721.
91. Fontenot JD, Rasmussen JP, Williams LM, Dooley JL, Farr AG, Rudensky AY. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* 2005;**22**:329–341.
92. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–29.
93. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;**102**:15545–15550.
94. McHeyzer-Williams LJ, Driver DJ, McHeyzer-Williams MG. Germinal center reaction. *Curr Opin Hematol* 2001;**8**:52–59.
95. Dave S, et al. Molecular diagnosis of Burkitt lymphoma. *N Engl J Med* 2006 (in press).
96. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001;**194**:1861–1874.
97. Castro AC, et al. Novel IKK inhibitors: beta-carbolines. *Bioorg Med Chem Lett* 2003;**13**:2419–2422.
98. Rosenwald A, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;**346**:1937–1947.
99. Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol* 2003;**4**:R69.
100. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;**34**:267–273.
101. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**:503–511.
102. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci USA* 2003;**100**:9991–9996.
103. Weller S, et al. CD40-CD40L independent Ig gene hypermutation suggests a second B cell diversification pathway in humans. *Proc Natl Acad Sci USA* 2001;**98**:1166–1170.
104. Weller S, et al. Human blood IgM 'memory' B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004;**104**:3647–3654.
105. William J, Euler C, Christensen S, Shlomchik MJ. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. *Science* 2002;**297**:2066–2070.
106. Matsumoto M, et al. Affinity maturation without germinal centres in lymphotoxin-alpha-deficient mice. *Nature* 1996;**382**:462–466.
107. Savage KJ, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood* 2003;**102**:3871–3879.
108. Rosenwald A, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med* 2003;**198**:851–862.

109. Basso K, et al. Tracking CD40 signaling during germinal center development. *Blood* 2004;**104**:4088–4096.
110. Klein U, et al. Transcriptional analysis of the B cell germinal center reaction. *Proc Natl Acad Sci USA* 2003;**100**:2639–2644.
111. Levens D. Disentangling the MYC web. *Proc Natl Acad Sci USA* 2002;**99**:5757–5759.
112. Brummelkamp TR, Bernards R. New tools for functional mammalian cancer genetics. *Nat Rev Cancer* 2003;**3**:781–789.
113. Brummelkamp TR, Bernards R, Agami R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* 2002;**2**:243–247.
114. Paddison PJ, Caudy AA, Sachidanandam R, Hannon GJ. Short hairpin activated gene silencing in mammalian cells. *Methods Mol Biol* 2004;**265**:85–100.
115. Silva J, Chang K, Hannon GJ, Rivas FV. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* 2004;**23**:8401–8409.
116. Sandy P, Ventura A, Jacks T. Mammalian RNAi: a practical guide. *Biotechniques* 2005;**39**:215–224.
117. Robbins MA, Rossi JJ. Sensing the danger in RNA. *Nat Med* 2005;**11**:250–251.
118. Berns K, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004;**428**: 431–437.
119. Paddison PJ, et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* 2004;**428**:427–431.
120. Westbrook TF, et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell* 2005;**121**:837–848.
121. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;**100**:57–70.
122. van de Wetering M, et al. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. *EMBO Rep* 2003;**4**:609–615.
123. Brummelkamp TR, Bernards R, Agami R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* 2002;**296**:550–553.
124. Li Z, et al. BCL-6 negatively regulates expression of the NF-kappaB1 p105/p50 subunit. *J Immunol* 2005;**174**:205–214.
125. Haaland RE, Yu W, Rice AP. Identification of LKLF-regulated genes in quiescent CD4⁺ T lymphocytes. *Mol Immunol* 2005;**42**: 627–641.
126. Napolitani G, Rinaldi A, Bertoni F, Sallusto F, Lanzavecchia A. Selected Toll-like receptor agonist combinations synergistically trigger a T helper type 1-polarizing program in dendritic cells. *Nat Immunol* 2005;**6**:769–776.
127. Peng T, Golub TR, Sabatini DM. The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation. *Mol Cell Biol* 2002;**22**:5575–5584.
128. Sweet-Cordero A, et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 2005;**37**:48–55.
129. Whitfield ML, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002;**13**:1977–2000.
130. Liu D, Umbach DM, Peddada SD, Li L, Crockett PW, Weinberg CR. A random-periods model for expression of cell-cycle genes. *Proc Natl Acad Sci USA* 2004;**101**:7240–7245.
131. Cho RJ, et al. Transcriptional regulation and function during the human cell cycle. *Nat Genet* 2001;**27**:48–54.
132. Glynne R, Ghandour G, Rayner J, Mack DH, Goodnow CC. B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays. *Immunol Rev* 2000;**176**:216–246.
133. Macian F, Garcia-Cozar F, Im SH, Horton HF, Byrne MC, Rao A. Transcriptional mechanisms underlying lymphocyte tolerance. *Cell* 2002;**109**:719–731.
134. Chang HY, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *Plos Biol* 2004;**2**:E7.
135. 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.
136. Chtanova T, et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *J Immunol* 2004;**173**:68–78.
137. Kim CH, Lim HW, Kim JR, Rott L, Hillsamer P, Butcher EC. Unique gene expression program of human germinal center T helper cells. *Blood* 2004;**104**:1952–1960.
138. Marshall AJQ, Draves KE, Shikishima Y, HayGlass KT, Clark EA. FDC-SP, a novel secreted protein expressed by follicular dendritic cells. *J Immunol* 2002;**169**: 2381–2389.
139. Huber C, et al. Lymphotoxin-beta receptor-dependent genes in lymph node and follicular dendritic cell transcriptomes. *J Immunol* 2005;**174**:5526–5536.
140. Rosenwald A, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 2003;**3**:185–197.
141. Rosenwald A, et al. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 2001;**194**:1639–1647.
142. Wiestner A, et al. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. *Blood* 2003; **101**:4944–4951.
143. Moreaux J, et al. The level of TACI gene expression in myeloma cells is associated with a signature of microenvironment dependence versus a plasmablastic signature. *Blood* 2005;**106**:1021–1030.
144. Ngo V, et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* 2006;(in press)