Gene expression profiling and statistical analysis

To identify NF-kB target genes using the IκBsr, mRNA from uninduced cells was labeled with the Cy3 dye and mRNA from doxycyline-treated, IκBsr-expressing cells was labeled with the Cy5 dye. Data was selected such that the Cy5 signal intensity was >200 relative fluorescent units (RFU). To identify target genes of cytokines, mRNA from untreated cells was labeled with the Cy3 dye and mRNA from IL-6-treated (50ng/ml) cells or IL-10-treated (50ng/ml) cells was labeled with the Cy5 dye. Data was selected such that the Cy5 signal intensity was >200 relative fluorescent units (RFU). To identify STAT3 target genes by RNA interference, mRNA from control cells was labeled with the Cy3 dye and mRNA from siRNA-transfected cells was labeled with the Cy5 dye. Data was selected such that the Cy3 signal intensity was >200 RFU. To identify genes regulated by JAK signaling, mRNA from untreated cells was labeled with the Cy3 dye and mRNA from JAK inhibitor-treated cells was labeled with the Cy5 dye. Data was selected such that the Cy3 signal intensity was >150 RFU.

In the IκBsr time course studies, a gene was selected as an NF-kB target gene in OCI-Ly3 or OCI-Ly10 cells if IκBsr decreased expression of the gene by >1.4-fold in at least two time points in the time course. A gene was selected as an NF-kB target gene in both OCI-Ly3 and OCI-Ly10 cells if IκBsr decreased expression of the gene by >1.4-fold in at least two time points in time courses from both cell lines. In the IL-6 time course studies in OCI-Ly10 cells, a gene was selected as an IL-6 target gene if IL-6 increased expression of the gene by >1.4-fold in at least two time points in the time course. In the IL-10 time course studies in OCI-Ly3 cells, a gene was selected as an IL-10 target gene if IL-10 increased expression of the gene by >3-fold in at least four time points in the time course. In the STAT3 siRNA time course studies in OCI-Ly10 cells, a gene was selected as an STAT3 target gene if STAT3 siRNA decreased expression of the gene by >1.4-fold in at least two time points in the time course and was an IL-10 target in OCI-Ly3 cells. In the JAK inhibitor time course studies in OCI-Ly10 cells, a gene was selected as an JAK inhibitor target gene if the JAK inhibitor decreased expression of the gene by >1.4-fold in at least two time points in the time course and was an IL-10 target in OCI-Ly3 cells.

The identification of ABC DLBCL signature genes that were not related to the STAT3 subset distinction was performed as follows. Genes that were more highly expressed in ABC DLBCL than GCB DLBCL were selected (>1.4-fold, $p<0.005$). From these, we removed all genes that were differentially expressed between the STAT3-high vs. STAT3-low subsets at $p<0.05$.

STAT3 predictor and gene set enrichment analysis

Hierarchical clustering was used to group the ABC DLBCL cases based on their expression of the STAT3 target genes (Figure 3E), leading to an initial definition of “STAT3-high” and “STAT3-low” subsets with relatively high or low expression of the STAT3 target genes, respectively (data not shown). Next, for each STAT3 target gene, a T-statistic was calculated as a measure of the difference in its expression between these two subsets. For each ABC DLBCL case, a “linear predictor score” was calculated as a weighted sum of the gene expression measurements for each STAT3 target gene, in which the weighting coefficient was the corresponding T-statistic for that gene. Based on the distribution of the linear predictor scores among the STAT3-high and STAT3-low subsets, Bayes’ rule was used to assign a probability

that a given case belonged to one or the other subset. If a case had more than a 90% probability of belonging to a subset, it was classified as such; all other cases were declared unclassified. Gene set enrichment analysis (GSEA) [1,2] was performed using a curated set of gene expression signatures (available at http://lymphochip.nih.gov/signaturedb/) [3], with the following modifications. For each signature in the database, the number of the signature genes that were correlated in expression ($r > 0.4$) across the entire set of ABC DLBCL tumors was determined. At least 50% of the genes in a signature were required to be correlated in expression for inclusion of the signature in the analysis.

The procedure for determining a nominal p-value for enrichment of a signature was modified as follows. For all the genes within each signature, we computed the t-statistics between the STAT3-high and STAT3-low subsets and averaged them to form a "true test statistic" for that signature. For the permutation test, we wished to create a similar statistic for data with permuted class labels. However, we realized that since there were many differentially expressed genes, the t-statistics on the permuted data would have a much different scale, which was addressed as follows. t-statistics between the STAT3-high and STAT3-low cases were calculated for all genes and ordered from least to greatest. We then randomly permuted the class labels (i.e. STAT3-high and STAT3-low), recalculated the t-tests for each gene, and used these t-statistics to rank the genes from least to greatest. Based on the rank of each gene in the permuted data, we associated it with a t-statistic of the same rank from the unpermuted data. These values were than averaged for all genes in a given signature to arrive at a permuted test statistic for that signature. This was repeated 1000 times. Finally the true test statistics and the permuted test statistics for each signature were normalized to match those of other signatures by subtracting their mean and dividing by their standard deviation. Based on these normalized statistics, calculation of a nominal permutation p-value and a false discovery rate (FDR) proceeded as described [1,2]. Only signatures with a FDR of <0.1 were reported.

References

1. Subramanian A, Tamayo P, Mootha VK, et al. From the Cover: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-15550.
2. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34:267-273.
3. Shaffer AL, Wright G, Yang L, et al. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. Immunol Rev. 2006;210:67-85.