Leukemia (2008), 1–9

www.nature.com/leu

ORIGINAL ARTICLE

Pathway activation patterns in diffuse large B-cell lymphomas

S Bentink[1], S Wessendorf[2], C Schwaenen[2], M Rosolowski[3], W Klapper[4], A Rosenwald[5], G Ott[5,13], AH Banham[6], H Berger[3], AC Feller[7], M-L Hansmann[8], D Hasenclever[3], M Hummel[9], D Lenze[9], P Möller[10], B Stuerzenhofecker[11], M Loeffler[3], L Truemper[11], H Stein[9], R Siebert[12] and R Spang[1] for the Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe

[1]*Institute of Functional Genomics, University of Regensburg, Regensburg, Germany;* [2]*Cytogenetic and Molecular Diagnostics, Internal Medicine III, University Hospital Ulm, Ulm, Germany;* [3]*Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany;* [4]*Institute of Hematopathology, University Hospital Schleswig-Holstein, Christian-Albrechts University Kiel, Kiel, Germany;* [5]*Institute of Pathology, University of Würzburg, Würzburg, Germany;* [6]*Nuffield Department of Clinical Laboratory Sciences, University of Oxford, Oxford, UK;* [7]*Institute of Pathology, University Hospital Schleswig-Holstein, Lübeck, Germany;* [8]*Institute of Pathology, University Hospital Frankfurt, Frankfurt, Germany;* [9]*Institute of Pathology, Campus Benjamin Franklin, Charité–Universitätsmedizin Berlin, Berlin, Germany;* [10]*Institute of Pathology, University Hospital Ulm, Ulm, Germany;* [11]*Department of Hematology and Oncology, Georg-August University Göttingen, Göttingen, Germany;* [12]*Institute of Human Genetics, University Hospital Schleswig-Holstein, Christian-Albrechts University Kiel, Kiel, Germany and* [13]*Department of Clinical Pathology, Robert-Bosch-Krankenhaus, Stuttgart, Germany*

Deregulation of cell signaling pathways controlling cell growth and cell survival is a common feature of all cancers. Although a core repertoire of oncogenic mechanisms is widely conserved between various malignancies, the constellation of pathway activities can vary even in patients with the same malignant disease. Modern molecularly targeted cancer drugs intervene in cell signaling compensating for pathway deregulation. Hence characterizing tumors with respect to pathway activation will become crucial for treatment decisions. Here we have used semi-supervised machine learning methodology to generate signatures of eight oncogene-inducible pathways, which are conserved across epithelial and lymphoid tissues. We combined them to patterns of pathway activity called PAPs for pathway activation patterns and searched for them in 220 morphologically, immunohistochemically and genetically well-characterized mature aggressive B-cell lymphomas including 134 cases with clinical data available. Besides Burkitt lymphoma, which was characterized by a unique pattern, the PAPs identified four distinct groups of mature aggressive B-cell lymphomas across independent gene expression studies with distinct biological characteristics, genetic aberrations and prognosis. We confirmed our findings through cross-platform analysis in an independent data set of 303 mature aggressive B-cell lymphomas.
Leukemia advance online publication, 26 June 2008;
doi:10.1038/leu.2008.166
Keywords: lymphoma; oncogenic pathways; transcriptional modules; cancer

Introduction

Bild *et al.*[1] recently showed for carcinomas that oncogenic pathway activation patterns (PAPs) artificially induced in nonmalignant breast epithelial cells can predict outcome and treatment efficiency. They transfected quiescent primary human mammary epithelial cells (HMECs) singly with each of the five human oncogenes: *MYC*, activated *RAS*, *SRC*, *E2F3* and activated β-*catenin*. They trained five discriminatory classifiers of oncogenic pathway activation on expression profiles from transfected HMECs and controls using a supervised Bayesian classification model.[2] The classifiers were successfully applied to expression profiles of epithelial neoplasms including ovarian, breast and lung cancer to predict the activity of each of the pathways in the tumors. So far this approach has not been followed for transcriptional profiles of lymphoid malignancies like diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL).

Molecular BL is a homogenous disease across patients with respect to large-scale gene expression profiling whereas DLBCL is shown to be heterogeneous.[3–8] A prominent molecular distinction is made between the germinal center- and the activated B-cell-like gene expression phenotype of DLBCL (GCB/ABC).[3,7] Moreover, primary mediastinal B-cell lymphomas[8,9] as well as a subset of BLs with DLBCL morphology can be identified using gene expression profiling.[4,5] DLBCL can also be grouped into three distinct 'consensus clusters' named after the most prominent genes in the signatures ('oxidative phosphorylation': OxPhos, 'B-cell receptor/proliferation': BCR and 'host response': HR).[6]

We have recently described a gene expression signature providing an index of 'Burkitt-likeness' (mBL-index) for mature aggressive B-cell lymphomas allowing their classification into molecular Burkitt lymphomas (mBL, mBL-index >0.95) and non-mBL (mBL-index <0.05). In-between cases were assigned an 'intermediate' status ($0.05 \leqslant$ mBL-index $\leqslant 0.95$). The majority of intermediate lymphomas and non-mBL displayed DLBCL morphology whereas mBL consisted of cases with and without classical BL morphology. Most mBLs generally exhibited a fusion of the *MYC* oncogene to an immunoglobulin (*IG*) locus in the absence of translocations involving *BCL2* or *BCL6*. In contrast, with only few exceptions, non-mBLs were *MYC* breakpoint negative. Interestingly, the intermediate cases were enriched for lymphomas with *MYC* fused to a non-*IG* locus or carrying an *IG-MYC* fusion in the presence of many imbalances or *BCL2* or *BCL6* translocations.[5]

Correspondence: Professor R Spang, Computational Diagnostics Group, Institute of Functional Genomics, University of Regensburg, Josef Engertstr. 9, 93053 Regensburg, Germany.
E-mail: rainer.spang@klinik.uni-regensburg.de or Professor R Siebert, Institute of Human Genetics, University Hospital Schleswig-Holstein, Campus Kiel, Schwanenweg 24, 24105 Kiel, Germany.
E-mail: rsiebert@medgen.uni-kiel.de
Received 18 December 2007; revised 7 May 2008; accepted 20 May 2008

Here we complement the molecular characterization of mature aggressive B-cell lymphomas by grouping 220 morphologically, immunohistochemically and genetically well-characterized mature aggressive B-cell lymphomas including 134 cases with clinical data available[5] according to patterns of pathway activity called PAPs. We identified besides BL four distinct groups of mature aggressive B-cell lymphomas across independent gene expression studies with distinct biological characteristics, genetic aberrations and prognosis, which are defined by different constellations of pathway activity.

Materials and methods

Data

Affymetrix (Affymetrix, Santa Clara, CA, USA) raw intensity files (CEL-files) were obtained at Gene Expression Omnibus[10] GEO-accession GSE4475 (DLBCL and BL,[5] Affymetrix HGU133A), http://data.cgt.duke.edu/oncogene.php (breast epithelial cell line data,[1] Affymetrix HGU133 plus 2.0), http://llmpp.nih.gov/BL (DLBCL and BL,[4] Affymetrix HGU133 plus 2.0 and custom Affymetrix LymphDx 2.7k). For each data set varying sample covariates were available from the respective accompanying websites. This included labels from gene expression signatures for BL,[4,5] GCB/ABC[4,5] and primary mediastinal B-cell lymphoma (PMBL).[4]

Data preprocessing

Affymetrix HGU133 raw data sets were merged based on common probe sequences between HGU133A and HGU133 plus 2.0 gene chips and were preprocessed together. Probe intensities were normalized using a variance stabilization method (vsn).[11] Gene expression levels were estimated by fitting an additive model employing a median polish routine.[12] HGU133A parameters for vsn (trim selection, sample scale and offset) and the medianpolish (probe effects) were estimated on the combined data set from Hummel *et al*.[5] and Bild *et al*.[1] HGU133 plus 2 samples from Dave *et al*.[4] were added to this core data set without changing the normalization of the original.[13] To adjust for study-specific effects (scanner generation, calibration, platform and so on), we followed the strategy described in Wright *et al*.[14] and additionally scaled and shifted each gene in the data set of Dave *et al*.[4] to have the same mean and variance across patients as in the respective reference samples from Hummel *et al*.[5] Affymetrix LymphDx data were preprocessed separately using vsn[11] and the medianpolish.[12]

Semi-supervised class detection and PAPs

Transcriptional modules conserved between mammary epithelial cell lines and lymphomas were identified as described in the Supplementary protocol S1. We used a semi-supervised extension of the established class finding algorithm ISIS.[15] By definition, we knew that for a transfected cell sample, all transcriptional modules associated with the transfected oncogene were activated, and for a control sample, we knew that they were not (labeled samples). For the lymphoma data, we did not know the activation status of the modules *a priori* (unlabeled samples). ISIS searches for bipartitions of the unlabeled samples into two groups, which can be clearly separated by subsets of coherently expressed genes (transcriptional modules). We applied ISIS to the combined data set of HMECs and lymphomas and restricted the search space to bipartitions of unlabeled lymphoma samples that were consistent with the labeled data. All control epithelial samples needed to fall in one group of lymphomas and all transfected epithelial samples in the other. We could then classify each lymphoma according to activity (similar to oncogene-transfected HMECs) or nonactivity (similar to control-transfected HMECs) of each oncogenic transcriptional module. We refer to the combination of the activity status of multiple pathway modules in individual lymphomas as PAP for pathway activation patterns.

Cross-platform PAP signatures

Oncogenic pathway modules were derived from data generated with Affymetrix HGU133 gene chips. In contrast, the series of Dave *et al*.[4] contains 303 transcriptional profiles generated with a custom Affymetrix oligonucleotide microarray (LymphDx 2.7k) with 2524 unique genes that are expressed differentially among the various forms of non-Hodgkin's lymphoma. This array holds only a fraction of the probe-sets of the HGU133-arrays. In addition, Dave *et al*.[4] hybridized 99 cases to both arrays. We used these cases to transfer our signatures from the HGU133 to the custom LymphDx chips. We applied our initial pathway signatures to add pathway activation labels to the 99 Affymetrix HGU133 plus 2 profiles and then trained new LymphDx-chip-based classifiers using these labels. Each signature includes 25 probe-sets, which are present on both arrays. Finally, we applied these signatures to all 303 LymphDx-chip profiles of Dave *et al*.[4]

Survival analysis

Clinical data including information on therapy and the two parameters age and Ann Arbor stage both used in the international prognostic index (IPI)[16] was available for 134 cases of our own study and for 220 cases of Dave *et al*.[4] We analyzed the association of survival with eight oncogenic modules and their combination to four recurrent PAPs (PAPs 1–4) by fitting multivariate Cox proportional hazard models. The analysis was carried out for each study separately and for the pooled data taking both studies together. For the pooled data we used stratified Cox models assuming separate baseline hazard functions for both studies. The ABC signature is an established prognostic indicator and age and Ann Arbor stage are part of the IPI. In our analysis we have included the presence of an ABC signature, age (age $>$ 59 years) and stage (stage = III or IV) as covariates in the multivariate models, such that the estimated hazards are independent of them. Furthermore, the analysis was restricted to patients having received a treatment with a combination of chemotherapy based on cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) or similar and not belonging to a further PAP, which is strongly associated with mBL (BL-PAP) as defined in this manuscript. Altogether, the survival analysis includes 81 of our own cases and 186 cases from Dave *et al*.[4]

Detection of chromosomal imbalances

Data on genomic imbalances from 183 cases in the GSE4475 data set determined by array-based comparative genomic hybridization (CGH) have been described previously.[5] Every clone on an array-CGH was either classified as showing genomic gain, loss, normal copy number, or it was called missing if it could not be classified.[17] A χ^2-test statistic was computed for each clone in order to test for overrepresentation of gains and losses in PAP groups. Missing values were removed before the computation of each statistic. Multiple testing was performed with the step-down minP method implemented in the R package multtest.[18]

Immunohistochemistry

Immunohistochemistry for forkhead box P1 (FOXP1) protein expression (antibody JC12) was performed on paraffin embedded tumor tissue from 27 samples of the PAP-1 and 27 samples of the PAP-2 groups according to standard protocols. Evaluation was performed twice, first by one and then by two different experienced hematopathologists (WK, evaluation 1; AR and GO, evaluation 2). The FOXP1 protein expression score was calculated by multiplying the score for the percentage of tumor cells showing FOXP1 protein expression (0%, 0; 1–25%, 1; 26–50%, 2; 51–75%, 3; 76–100%, 4) by the score describing the intensity of the staining (weak, 1; moderate, 2; strong, 3). Intensity scores were computed separately for each evaluation. Intensities in PAP-1 and PAP-2 were compared using a one-sided Wilcoxon's test for a shift of PAP-2 intensities toward higher expression. We performed a separate test for each expert evaluation and conservatively reported the larger of the *P*-values.

Results

Detection of conserved transcriptional modules

We analyzed the original oncogenic pathway classifiers from Bild *et al*.[1] derived from primary HMEC cultures in a microarray data set of 220 mature aggressive B-cell lymphomas. In contrast to the situation in epithelial neoplasms, the purely supervised application of the HMEC signatures to human tumor samples[1] proved insufficient to identify signatures of pathway activation in our lymphoma data set. Most likely, this failure was due to global expression differences between epithelial and lymphoid tissues (Figure 1a). In order to bridge the gap, we extended the computational framework from Bild *et al*.[1] by searching for oncogene-inducible modules—sets of genes—which were differentially expressed between oncogene-activated and control HMECs, and subdivided our own lymphoma cases into two groups: one with expression levels characteristic for the nonactivated state of the pathway and one with expression levels typical for the activated state (Figure 1a). For several oncogenes, we found multiple conserved oncogene-inducible modules (Figure 1b).

For the *MYC*, *RAS*, *SRC* and *E2F3* oncogenes, but not for β-*catenin*, we were able to identify eight nonredundant modules in a training set of $n=100$ mature aggressive B-cell lymphomas (MYC.1, E2F3.1, E2F3.2, SRC.2, SRC.10, RAS.1, RAS.4 and RAS.6; Supplementary Tables S1–S8). Each of these oncogene-inducible modules was either active or nonactive in a lymphoma. The modules characterize lymphomas but they do not group them, as they overlap. For this reason, we combined the eight activation states to a binary pattern, which we called a *PAP* (Supplementary Table S9). PAPs define nonoverlapping groups of lymphomas from the perspective of pathway activation. Figure 2 shows the patterns together with the underlying expression data. Expression characteristics from the training set reemerged well in the test samples ($n=120$), such that virtually no signs of overfitting were evident. Importantly, 158 of the 220 (72%) mature aggressive B-cell lymphomas showed only five of the possible $2^8=256$ PAPs. The remaining 62 samples displayed distinct, rarely or nonrecurrent patterns (recurrence in $<5\%$ of the lymphomas). They were subsumed in a heterogeneous pool called mind-L for molecularly individual lymphomas.

Burkitt lymphoma is characterized by a distinct pathway activity pattern

Recently, we and others have described mBL as a homogeneous lymphoma entity with respect to molecular, genetic and clinical features.[4,5] Also here, we found only one recurrent PAP (BL-PAP) among the mBL cases. This pattern was expressed in 39 of the 44 (89%) mBL cases. *Vice versa*, 39 of the 41 (95%) lymphomas displaying the pattern showed the mBL signature (Table 1). Consistent with activity of the MYC.1 module in the BL-PAP, in 38 of 40 of these lymphomas, where data from fluorescence *in situ* hybridization were available, a *MYC* breakpoint was detectable (Table 1).

Pathway activation patterns and lymphoma stratification

In addition to the BL-PAP, we identified four recurrent patterns in DLBCL, which we termed PAP-1 to PAP-4 (Table 1; Figure 2). A total of 42 DLBCLs showed the most frequent pattern PAP-1, which was exactly the inverse of the BL-PAP. The second pattern, PAP-2, was present in 32 DLBCLs. PAP-1 and -2 lymphomas frequently expressed the BCL6 protein (28/37, or 76% and 25/31, or 81%, respectively; Table 1) but rarely CD10 (7/33, or 18% and 4/27, or 13%, respectively; Table 1). PAP-3 was the only pattern, except for BL-PAP, that displayed activation of the MYC.1 module, although this did not commonly arise through *MYC* translocation, as a break was detected in only 3 of 27 cases suggesting alternative means of pathway activation. PAP-4, which was displayed by 16 lymphomas, was the only activation pattern more prevalent in females (11/16). Unlike BL, neither activated B-cell-like DLBCLs (ABC-DLBCL) nor germinal center-like DLBCLs (GCB-DLBCL)[7] displayed a unique PAP, nor did the consensus clusters defined by Monti *et al*.[6] (Table 1; Supplementary Protocol S1). Finally, mind-L included 69% (33/48) of intermediate cases[5] with an mBL-index between mBL and non-mBL. In line with this, a high number of lymphomas with *MYC* breakpoints were included in this pool (28/61, or 46% of mind-L cases), which turned out to be particularly enriched for lymphomas with non-*IG-MYC* fusion and *MYC* complex status (11/15, or 73% and 22/33, or 66% of all non-*IG-MYC* and *MYC* complex cases, respectively).

The PAP-groups show distinct patterns of chromosomal changes and FOXP1 expression

Despite PAP-2 differing transcriptionally from PAP-1 only with respect to the first E2F3 inducible module, E2F3.1, it showed a profile of chromosomal changes markedly distinct, not only from PAP-1, but also from PAP-3 and -4 (Figure 3). The most frequent chromosomal changes in PAP-2 detected by array-CGH were gains at 18q21 and 3q27 ($n=14/29$ PAP-2 cases, or 48%). Gains at 18q21 containing the *BCL2* and *MALT1* oncogenes and 3q27 containing the *BCL6* oncogene have been shown to be associated with ABC-DLBCL.[19] Notably, in PAP-2 also GCB-DLBCLs showed *MALT1*/18q21 gains (2/5) and lacked *IGH-BCL2* fusions ($n=5$) like ABC-DLBCLs, indicating genetic similarity of GCB- and ABC-like lymphomas in PAP-2. The most specific chromosomal change in PAP-2 was a gain at 3p13. Indeed, of nine lymphomas displaying a 3p13 gain, eight belonged to PAP-2 and only one to PAP-3. The specifically gained region in PAP-2 at 3p13 contains *FOXP1*, which has been proposed as an oncogene involved in lymphomagenesis as well as a prognostic marker in diffuse large B-cell lymphoma.[20–24] Immunohistochemical staining of the FOXP1 protein in 54 cases (PAP-1, $n_1=27$; PAP-2, $n_2=27$) revealed a significantly increased expression of the transcription factor in PAP-2 as compared to PAP-1 ($P=0.006$). Furthermore, the 3p13-gain-positive cases express the FOXP1 protein at high levels indicating an association between FOXP1 expression, 3p13 gains and the PAP stratification.

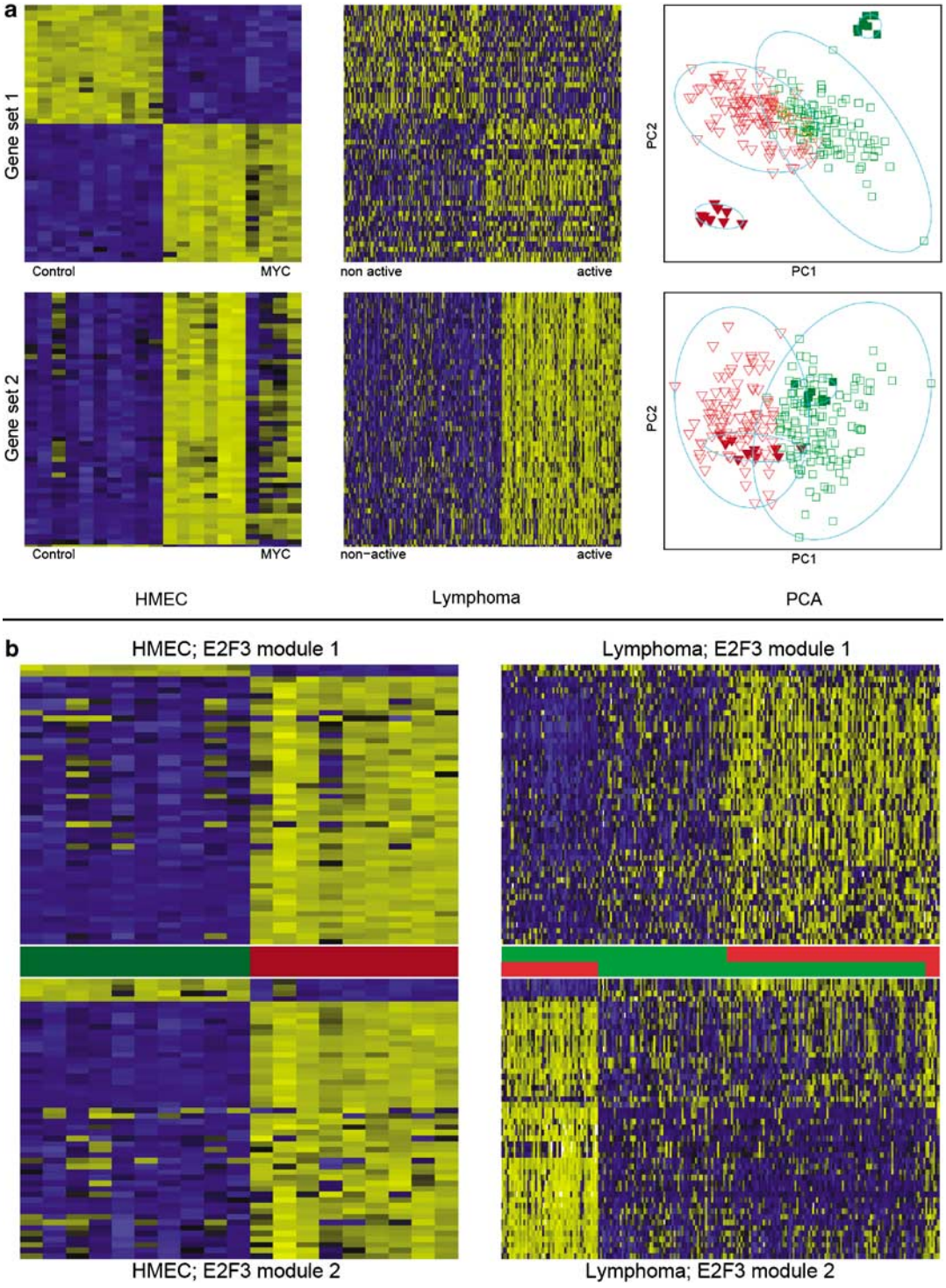

Figure 1 Conserved expression modules across epithelial and lymphoid tissues. The heat maps display expression levels of different gene sets (rows) in human primary mammary epithelial cells (HMECs) and lymphoma samples (columns). Yellow encodes expression above the overall mean of a gene, blue encodes expression below that mean and black encodes average expression levels. (**a**) The top heat map on the left shows the 50 most differentially expressed genes between control and *MYC*-transfected HMECs. The corresponding heat map in the middle column shows the expression of the same genes in lymphomas, where only little structure becomes apparent. The plot on the top right displays all samples in the space spanned by the first two principal components generated by this gene set. Dominant is the difference between HMECs (filled symbols) and lymphomas (empty symbols). Green squares indicate *MYC*-transfected HMECs as well as corresponding lymphomas with an activated *MYC* expression module, red triangles indicate the control-transfected status, respectively. Below we plotted the same for the gene set identified by our semi-supervised selection procedure. The induced structure in the lymphoma samples is stronger and module activity becomes the dominant feature in the principal component analysis, showing that for this gene set, hardly any systematic expression difference between HMECs and lymphomas exists. (**b**) The heat maps represent two different conserved modules, which are jointly activated in HMECs in response to *E2F3* overexpression but can be activated independently from each other in lymphomas. The E2F3.1 module is active in patients grouped to the right of the plots indicated by the top red bar between the plots, an E2F3.2 is activated in lymphomas grouped to the left and indicated by the bottom red bar.

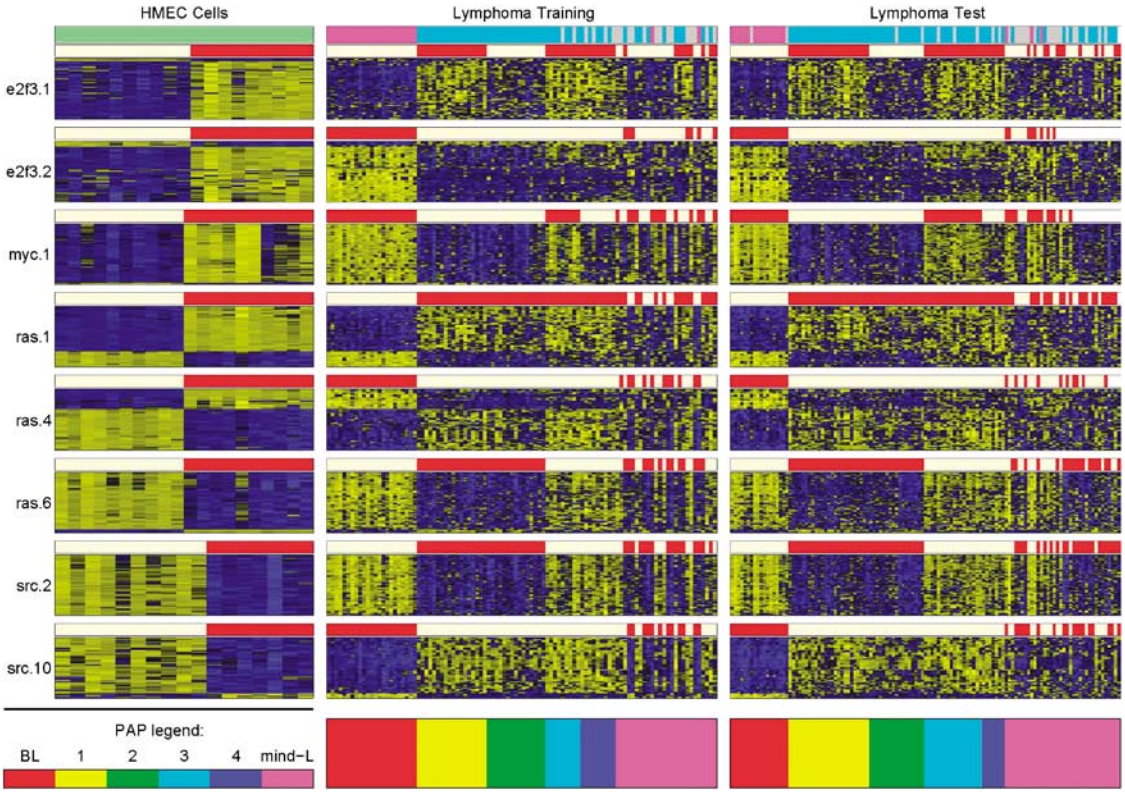

Figure 2 Stratification of lymphomas based on pathway activation patterns (PAPs): the eight rows correspond to the conserved oncogenic expression modules. In each row the left heat map shows expression of the module genes in human primary mammary epithelial cells (HMECs), the middle heat map shows the expression of the same genes in the training samples of the lymphoma data, and the right one refers to the test samples of the lymphoma data. The samples are sorted by module activation patterns starting with Burkitt lymphoma PAP (BL-PAP) on the left and ending with mind-L on the right (see the color coding in the bar below the heat maps). Above each row of heat maps is a bar indicating module activation in red. The pattern of module activation is constant in each of the groups BL-PAP, PAP-1, PAP-2, PAP-3 and PAP-4 but heterogeneous in the pool mind-L. The horizontal bar on top of all plots encodes the type of samples (lightgreen, HMECs; cyan, non-mBL; magenta, mBL; gray, intermediate).

The PAP groups are also present in an independent data set

Most notably, the four recurring DLBCL-PAPs and the BL-PAP found in our data set were also the most recurrent PAPs in the independent study on 303 mature aggressive B-cell lymphomas of Dave *et al*.[4] (Supplementary Table S10). This indicates that the five PAPs widely cover the spectrum of mature aggressive B-cell lymphomas from the perspective of pathway activation constellations. Furthermore, 49 of the 55 lymphomas exhibiting the BL-PAP in the data set from Dave *et al*.[4] also showed the BL-signature defined in this paper confirming that the BL-PAP pattern is characteristic for BL.

Survival analysis

Figure 4 summarizes the results of survival analysis of BL-PAP-negative cases from both data sets. Hazards estimated from multivariate analysis adjusting for ABC status, age, and Ann Arbor stage together with their 95% confidence intervals (CIs) are shown. In our own collection of lymphomas the PAP-1 group has a significantly better prognosis then non-PAP-1 DLBCLs (hazard ratio for death, 0.25; 95% CI: 0.1–0.65; $P=0.004$), whereas PAP-2 is a group with a significantly worse prognosis (hazard ratio for death, 2.45; 95% CI: 1.16–5.17; $P=0.019$). The same trends are observed in the study of Dave *et al.*,[4] although statistical significance is not reached. Addressing the question, as to which molecular features are responsible for the prognostic differences, we have extended the survival analysis to the level of individual oncogene-inducible modules. In the pooled analysis of both data sets, patients with lymphomas showing activity of the first E2F3 inducible module, E2F3.1, had a significantly better prognosis (hazard ratio for death, 0.47; 95% CI: 0.33–0.67; $P=0.00003$). This prognostic effect, which notably is independent of the cell of origin signature and the well-established clinical risk factors age and Ann Arbor stage, is also visible in both individual studies. In contrast, patients with RAS.4 active lymphomas, which are restricted to and comprise almost one-third of the mind-L lymphomas, display consistently worse outcome across both studies.

Discussion

By introducing conserved oncogenic transcriptional modules to the molecular pathology of lymphomas we have structured this disease from the perspective of oncogenic pathway activation. The PAPs identified four novel biologically homogenous subgroups among the DLBCLs. In contrast, our approach describes mBL as a lymphoma with a uniform oncogenic pathway pattern, in line with our recent molecular definition of this lymphoma type. Thus, mBL truly represents a single lymphoma entity.

Table 1 Characterization of patients in the PAP groups

	All	*BL-PAP*	*PAP-1*	*PAP-2*	*PAP-3*	*PAP-4*	*Mind-L*	P-*value*
Total	220 (100%)	41 (19%)	42 (19%)	32 (15%)	27 (12%)	16 (7%)	62 (28%)	
Modules		E2F3.1	**E2F3.1**	E2F3.1	**E2F3.1**	**E2F3.1**	21 (34%)	
		E2F3.2	E2F3.2	E2F3.2	E2F3.2	E2F3.2	15 (24%)	
		MYC.1	MYC.1	MYC.1	**MYC.1**	MYC.1	27 (44%)	
		RAS.1	**RAS.1**	**RAS.1**	**RAS.1**	**RAS.1**	38 (61%)	
		RAS.4	RAS.4	RAS.4	RAS.4	RAS.4	21 (34%)	
		RAS.6	**RAS.6**	**RAS.6**	RAS.6	RAS.6	33 (53%)	
		SRC.2	**SRC.2**	**SRC.2**	SRC.2	SRC.2	40 (65%)	
		SRC.10	SRC.10	SRC.10	SRC.10	SRC.10	31 (50%)	
Morphologic diagnosis								0.94393
BL	8 (4%)	8 (20%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Atypical BL	28 (13%)	22 (54%)	1 (2%)	1 (3%)	1 (4%)	1 (6%)	2 (3%)	
DLBCL	165 (75%)	7 (17%)	36 (86%)	30 (94%)	25 (93%)	14 (88%)	53 (85%)	
Aggressive B-nHL unclassifiable	18 (8%)	4 (10%)	5 (12%)	1 (3%)	1 (4%)	1 (6%)	6 (10%)	
BL-leukemia	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	
GCB or ABC signature								0.00002
ABC	58 (33%)	0 (0%)	8 (19%)	20 (63%)	14 (52%)	5 (31%)	11 (19%)	
GCB	79 (45%)	2 (100%)	19 (45%)	5 (16%)	7 (26%)	9 (56%)	37 (65%)	
Unclassified	39 (22%)	0 (0%)	15 (36%)	7 (22%)	6 (22%)	2 (13%)	9 (16%)	
Molecular Burkitt lymphoma								0.00001
mBL	44 (20%)	39 (95%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (8%)	
Non-mBL	128 (58%)	0 (0%)	42 (100%)	30 (94%)	22 (81%)	10 (63%)	24 (39%)	
Intermediate	48 (22%)	2 (5%)	0 (0%)	2 (6%)	5 (19%)	6 (38%)	33 (53%)	
MYC partner								0.00001
IG-MYC	59 (28%)	37 (93%)	1 (3%)	2 (7%)	2 (7%)	0 (0%)	17 (28%)	
Non-IG-MYC	15 (7%)	1 (3%)	0 (0%)	2 (7%)	1 (4%)	0 (0%)	11 (18%)	
MYC-negative	140 (65%)	2 (5%)	39 (98%)	26 (87%)	24 (89%)	16 (100%)	33 (54%)	
Genetic group								0.00001
MYC simple	35 (17%)	28 (80%)	0 (0%)	1 (3%)	1 (4%)	0 (0%)	5 (8%)	
MYC complex	33 (16%)	5 (14%)	1 (3%)	3 (10%)	2 (7%)	0 (0%)	22 (37%)	
MYC-negative	140 (67%)	2 (6%)	39 (98%)	26 (87%)	24 (89%)	16 (100%)	33 (55%)	
Consensus cluster								0.00001
BCR	123 (56%)	33 (80%)	11 (26%)	21 (66%)	11 (41%)	9 (56%)	38 (61%)	
HR	63 (29%)	0 (0%)	30 (71%)	11 (34%)	4 (15%)	5 (31%)	13 (21%)	
OxPhos	34 (15%)	8 (20%)	1 (2%)	0 (0%)	12 (44%)	2 (13%)	11 (18%)	

Abbreviations: ABC, activated B-cell-like; BCR, B-cell receptor; BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma; GCB, germinal center B-cell-like; HR, host response; IG, immunoglobulin; mBL, molecular Burkitt lymphoma; Mind-L, molecularly individual lymphoma; OxPhos, oxidative phosphorylation; PAP, pathway activation pattern.
The first row gives the total number of patients within each PAP group. The second row indicates in bold which modules are active in the respective PAP. The following rows contain morphological and molecular characteristics of patients in the PAP groups.[5] *P*-values reflect overrepresentation of a characteristic in a DLBCL-PAP group and were calculated using to a two-sided χ^2-test including PAP-1 to 4 and mind-L, but excluding BL-PAP cases. A full version can be found as supplementary material (Supplementary Table S13).

More importantly, the PAPs identified four novel biological subgroups among the DLBCLs with homogenous pathway activation constellations. Most remarkable are the differences between the two largest groups PAP-1 and PAP-2. First, we observed strong prognostic effects. Second, PAP-2 is characterized by accumulated genetic aberrations on several chromosomes, which are found only on baseline frequencies in PAP-1 lymphomas. Moreover, protein expression of FOXP1 is significantly higher in PAP-2, in line with frequently observed gains on chromosome 3p13 around the locus of this gene. Notably, FOXP1 constitutes a target for *IGH*-translocations in DLBCL and MALT-type lymphomas[20–22] and expression of the FOXP1 protein, a member of the forkhead box (FOX) transcription factor family, has also been reported to be associated with poor prognosis in DLBCL.[23,24] These differences are even more striking given that PAP-1 and PAP-2 only differ with respect to the activity of the E2F3.1 module.

We have introduced two concepts: oncogene-inducible modules and PAPs. Both have merits of their own. The first E2F3-inducible module, E2F3.1, and the RAS-inducible module, RAS.4, appear to be the strongest prognostic markers. However, modules do not group patients, as they overlap. A patient is not either E2F3.1-positive or RAS.4-positive, but can also have both features or none at all. Moreover, no single module on its own characterizes BLs. In contrast, PAPs define nonoverlapping lymphoma groups; a feature that is important in view of treatment decisions or molecularly stratified clinical studies.

All oncogene-inducible pathways are active in some DLBCL and inactive in others, supporting the hypothesis that DLBCL as a whole is a biologically heterogeneous lymphoma entity. However, the accumulation of 72% of BL and DLBCL cases in only 5 of the 256 possible PAPs shows that the biological processes underlying the conserved modules are not independently regulated in lymphomas. In contrast, their regulatory

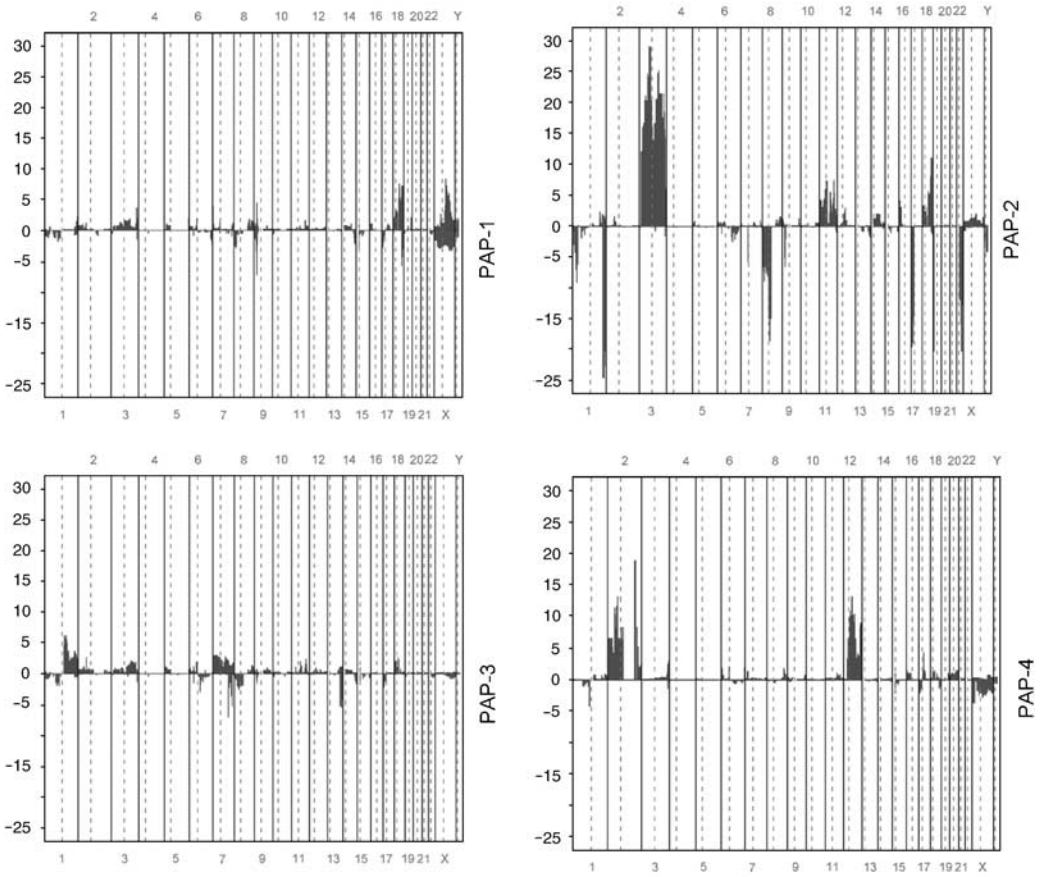

Figure 3 Specificity of chromosomal imbalances detected by array-CGH (comparative genomic hybridization) for pathway activation pattern (PAP)-1 to 4. The *x* axis shows the genomic position of a clone. Chromosomes are separated by solid and chromosome arms by dashed vertical lines. The *y* axis shows the χ^2 score measuring overrepresentation of gains and losses in the respective PAP groups. Scores corresponding to losses are shown with a negative sign for clarity. Overrepresentation tests were computed for the contrasts PAP-1 (top left), PAP-2 (top right), PAP-3 (bottom left) and PAP-4 (bottom right) vs the respective rest of the cases without the Burkitt lymphoma PAP (BL-PAP) group.

interaction characterizes mBL and the four biologically homogenous groups of DLBCL.

Among DLBCLs (PAPs 1–4), our survival analysis suggests that oncogene module activation patterns have clinical significance and are associated with overall survival. Similar effects were observed in an independent data set. Although statistical significance and reproducibility in a second data set were achieved, it is important to note that sample sizes are small, studies are retrospective and patients were not treated with today's state of the art treatment combining rituximab with CHOP (R-CHOP). Thus definitive conclusions concerning the prognostic value of both modules and PAPs require further studies, although certain trends are visible.

The oncogenes analyzed here have been chosen in the original study by Bild *et al*.[1] due to their prominent role in breast, ovarian and lung cancer. Although they do not give a complete picture of oncogenic pathway activity in lymphomas (for example BCL6, BCL2, MUM1 and BLIMP1, which have not been analyzed), the strong conservation of downstream transcriptional modules is remarkable and underlines the general importance of these pathways in tumor genesis.

The PAPs identified four novel biologically homogenous subgroups among the DLBCLs, which could guide the design of stratified prospective randomized studies on the efficiency of treatment modalities. In the future, PAPs could direct the development of inhibitors specific for oncogene-driven pathways characteristically activated in our pathway-defined lymphoma subgroups. Based on the conservation of the oncogenic modules across various solid and hematological cancers, targeted molecular-based therapies might well be effective in different kinds of tumors irrespective of localization or tissue derivation. Finally, conservation of oncogenic modules across cancers may also help to explain why some widely used anticancer drugs are potent in different cancers.

Acknowledgements

This work was carried out within the framework of the research network 'Molecular Mechanisms in Malignant Lymphoma' (MMML), supported by the Deutsche Krebshilfe (70-3173-Tr3). In addition R Spang announces support by the Bavarian Genome Network *BayGene* and AHB is supported by the Leukaemia Research Fund.

Members of MMML project are: *Pathology group*: Wolfram Klapper, Monika Szcepanowski, Thomas Barth, Wolfram Bernd, Alfred Feller, Martin-Leo Hansmann, Peter Möller, German Ott, Hans-Konrad Müller-Hermelink, Andreas Rosenwald, Hans-Heinrich Wacker, Sergio Cogliatti, Michael Hummel, Harald Stein. *Genetics group*: Carsten Schwaenen, Swen Wessendorf, Heiko Trautmann, Jose-Ignacio Martin-Subero, Eugenia Haralambieva, Judith Dierlamm, German Ott, Andreas Rosenwald, Thomas

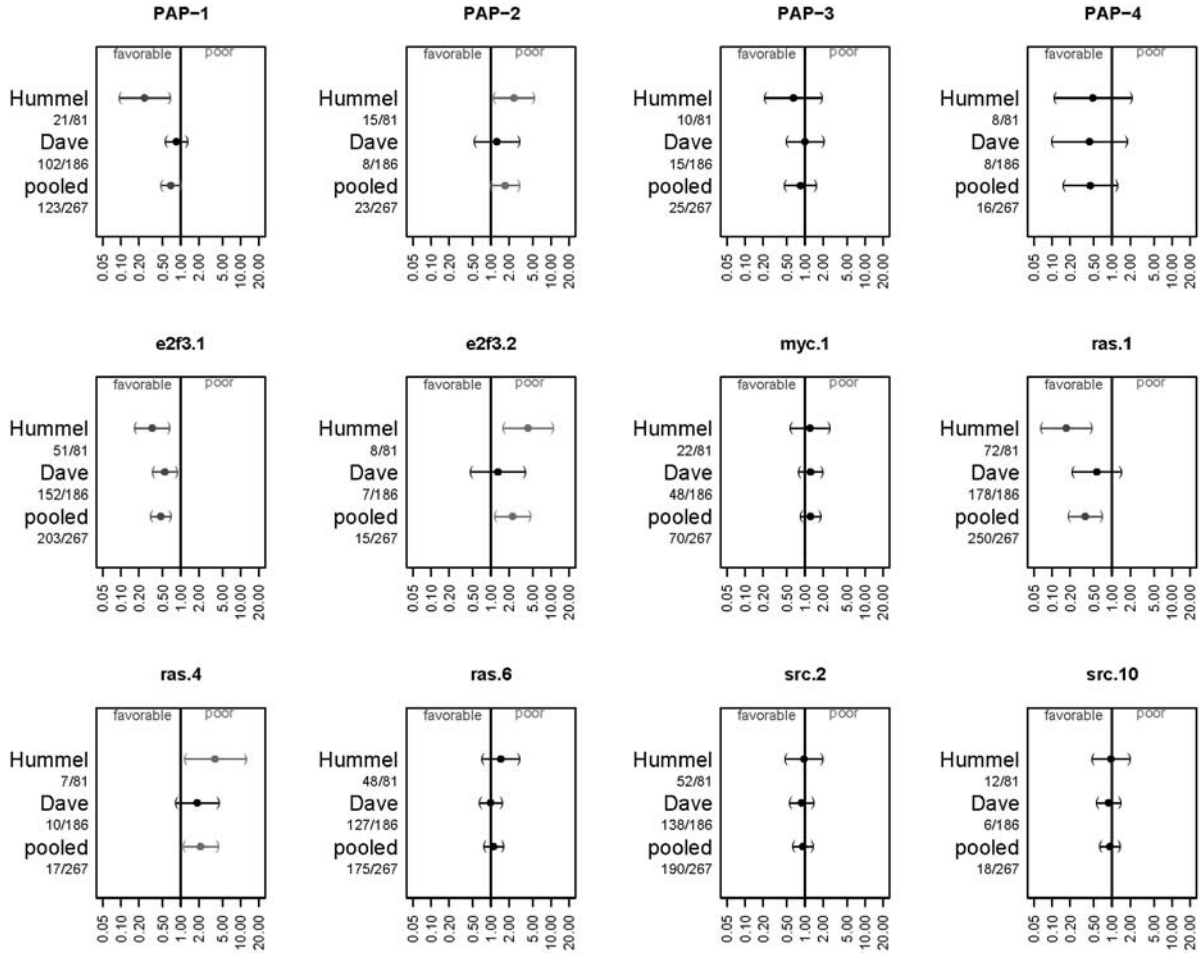

Figure 4 Survival analysis of Burkitt lymphoma pathway activation pattern (BL-PAP)-negative lymphomas adjusted for activated B-cell (ABC) status, age and Ann Arbor stage. Shown are estimated hazard ratios (*x* axes) and their 95% confidence intervals associated with PAPs 1–4 (top row) and the underlying eight conserved oncogenic modules (bottom rows). Within each plot rows refer to the two individual studies and a pooled analysis, respectively. Hazard ratios result from multivariate Cox models including ABC status, age and Ann Arbor stage and, thus, give estimations of risks independent of these known risk factors. For the pooled data we used stratified Cox models assuming separate baseline hazard functions for each study. The group sizes as well as the total number of samples are given on the left of each plot (*n* (group)/*n* (total)).

Barth, Christiane Pott, Ralf Küppers, Reiner Siebert. *Bioinformatics group*: Maciej Rosolowski, Rainer Spang, Hilmar Berger, Stefan Bentink, Dirk Hasenclever, Markus Loeffler. *Project coordination*: Benjamin Stürzenhofecker, Hilmar Berger, Michael Hummel, Lorenz Trümper. *Steering Committee*: Reiner Siebert, Harald Stein, Markus Loeffler, Lorenz Trümper.

References

1 Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006; **439**: 353–357.
2 West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R *et al.* Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA* 2001; **98**: 11462–11467.
3 Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403**: 503–511.
4 Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma EJ *et al.* Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 2006; **354**: 2431–2442.
5 Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF *et al.* A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med* 2006; **354**: 2419–2430.
6 Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 2005; **105**: 1851–1861.
7 Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI *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.
8 Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD *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.
9 Savage KJ, Monti S, Kutok JL, Cattoretti G, Neuberg D, De Leval L *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.
10 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; **30**: 207–210.

11 Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002; **18** (Suppl 1): 96–104.
12 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al*. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**: 249–264.
13 Kostka D, Spang R. Microarray based diagnosis profits from better documentation of gene expression signatures. *PLoS Comput Biol* 2008; **4**: e22.
14 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.
15 von Heydebreck A, Huber W, Poustka A, Vingron M. Identifying splits with clear separation: a new class discovery method for gene expression data. *Bioinformatics* 2001; **17** (Suppl 1): S107–S114.
16 Shipp MA. Prognostic factors in aggressive non-Hodgkin's lymphoma: who has 'high-risk' disease? *Blood* 1994; **83**: 1165–1173.
17 Kreuz M, Rosolowski M, Berger H, Schwaenen C, Wessendorf S, Loeffler M *et al*. Development and Implementation of an analysis tool for array-based comparative genomic hybridization. *Methods Inf Med* 2007; **46**: 608–613.
18 Pollard KS, Dudoit S, van der Laan MJ. Multiple testing procedures: the multtest package and applications to genomics. In: Gentleman R, Carey VJ, Huber W, Irizarry R, Dudoit S (eds). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer: New York, 2005. pp 249–271.
19 Bea S, Zettl A, Wright G, Salaverria I, Jehn P, Moreno V *et al*. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* 2005; **106**: 3183–3190.
20 Fenton JA, Schuuring E, Barrans SL, Banham AH, Rollinson SJ, Morgan GJ *et al*. t(3;14)(p14;q32) results in aberrant expression of FOXP1 in a case of diffuse large B-cell lymphoma. *Genes Chromosomes Cancer* 2006; **45**: 164–168.
21 Streubel B, Vinatzer U, Lamprecht A, Raderer M, Chott A. T(3;14)(p14.1;q32) involving IGH and FOXP1 is a novel recurrent chromosomal aberration in MALT lymphoma. *Leukemia* 2005; **19**: 652–658.
22 Wlodarska I, Veyt E, De Paepe P, Vandenberghe P, Nooijen P, Theate I *et al*. FOXP1, a gene highly expressed in a subset of diffuse large B-cell lymphoma, is recurrently targeted by genomic aberrations. *Leukemia* 2005; **19**: 1299–1305.
23 Banham AH, Connors JM, Brown PJ, Cordell JL, Ott G, Sreenivasan G *et al*. Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma. *Clin Cancer Res* 2005; **11**: 1065–1072.
24 Barrans SL, Fenton JA, Banham A, Owen RG, Jack AS. Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood* 2004; **104**: 2933–2935.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)