Supplemental Information

Large scale microarray profiling reveals four stages of immune escape in Non-Hodgkin Lymphomas

Authors:

Marie Tosolini, Christelle Algans, Frédéric Pont, Bernard Ycart, & Jean-Jacques Fournié

Supplemental Methods

Datasets

We screened the Gene Expression Omnibus (GEO)[1] repository for gene expression profiles of B-cell Non-Hodgkin lymphomas obtained with the Affymetrix HGU133 plus 2.0 microarray and downloaded 47 series of RMA-normalized datasets (GSE) as txt files, together with gene sets from Gene Ontology,[2] KEGG (`http://www.genome.jp/kegg/pathway.html`) from the Molecular Signatures Database (MSigDB 3.0)[3] as well as from TM22[4,5] and gene sets defined in text. Each series of transcriptomes were collapsed to common protein-encoding genes (HUGO nomenclature) and the compatibility for merging of the datasets was assessed. The gene expression data of each GSM were transformed into van der Waerdens scores,[6,7] each dataset was reduced to a per gene median of samples and the pairwise correlation matrix of these 47 datasets was computed (Supplemental Figure 1A).[8] The 33 datasets with correlations >0.8 were deemed compatible for merging while the less homogeneous others were discarded (Supplemental Table 1). Since these 33 datasets comprised malignant samples as well as various normal cell controls, the compatibility screening was repeated at the sample level for the group of GSM from all malignant samples on the one hand, and for the group of all non-cancer cell controls on the other hand (Supplemental Figure 1B). The samples deemed irrelevant for this meta-analysis (e.g. d microarray-based study of invitro drug activity on cell lines) were rejected. This yielded a dataset composed of 1446 samples from 7 distinct malignant histologies. These included 489 diffuse large B cell lymphoma (DLBCL), 149 follicular lymphoma (FL), 125 mantle cell lymphoma (MCL), 12 small marginal zone lymphoma (MZL), 4 Burkitts lymphomas (BL), 630 chronic lymphocytic leukemia (CLL), 5 hairy cell leukemia (HCL), and control cells which encompassed 32 normal B cell samples including CD20 B cell purified from peripheral blood, nave B cells, centrocytes and centroblasts purified from tonsils of non-cancer donors.

Sample enrichment scores (SES)

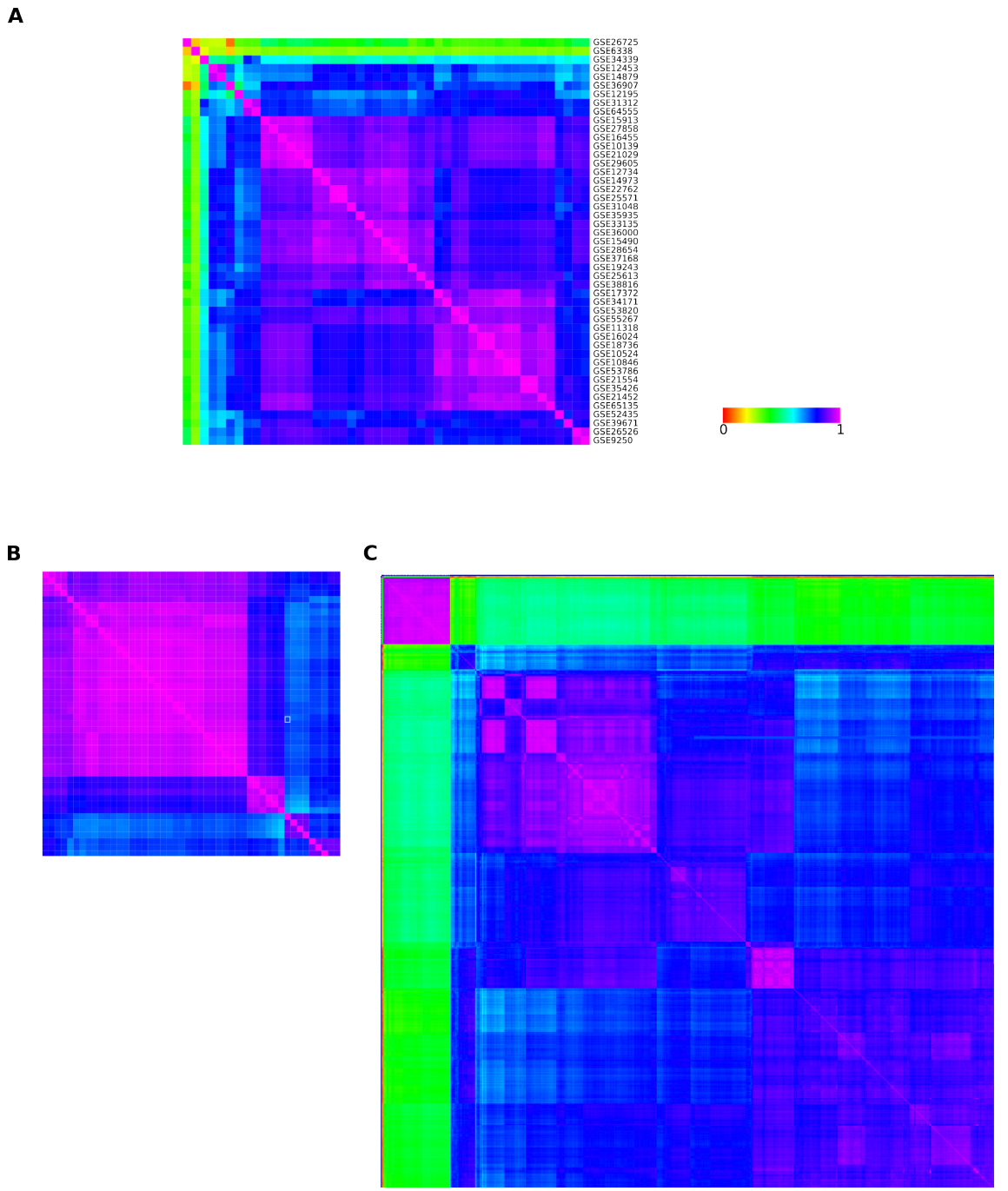
The RMA-normalized expression data from each collapsed sample (GSM) were rank-ordered by decreasing expression level along the [0-1] segment (highest expr. level=0, lowest expr. level=1) yielding v, a vector of ranks. Since each series from the NHL data set had been pre-selected for high pairwise correlation (see above), their rank-ordered gene distributions were highly correlated as well, despite distinct batch RMA normalizations (Supplemental Figure 3). Let gene set g be a subset of v, then assessing enrichment of g means testing g *versus* v. However the human genome comprises many pleiotropic genes involved in several different pathways, functions, and cell compartment more frequently up-regulated in any sample, and over-represented in gene sets and databases.[9] Since genes are far from equiprobable in gene sets, the adequate null hypothesis for g versus v gives genes with probabilities proportional to their frequency in the gene set database[10] and requires Kolmogorov-Smirnov (KS), T-test (T) and Wilcoxon (W) tests with frequency-corrected null hypothesis (see[11–13] for review). In addition scoring enrichment of the $\sim$1500 KEGG and GO gene sets in $>$1000 samples means computing $\sim$1.5 million tests with high accuracy for the most significant results. So, algoritmic optimization was necessary i) to compute scores in a reasonable time, ii) to perform a precise computing of the most significant (smallest) P values, iii) to allow the gene frequency correction, and iv) to allow testing of the 3 alternatives "greater" (enriched geneset),"less" (down-regulated geneset) and "two.sided" (the geneset is enriched or down-regulated). We developped an R script which current version computes enrichment of $\sim$5000 gene sets for a single sample within $\sim$5 seconds (Statistical tests have been implemented as R functions `http://www.R-project.org/`. Source codes are available here: `https://sites.google.com/site/fredsoftwares/products/autocompare_ses`).

Validation of SES by distribution analysis

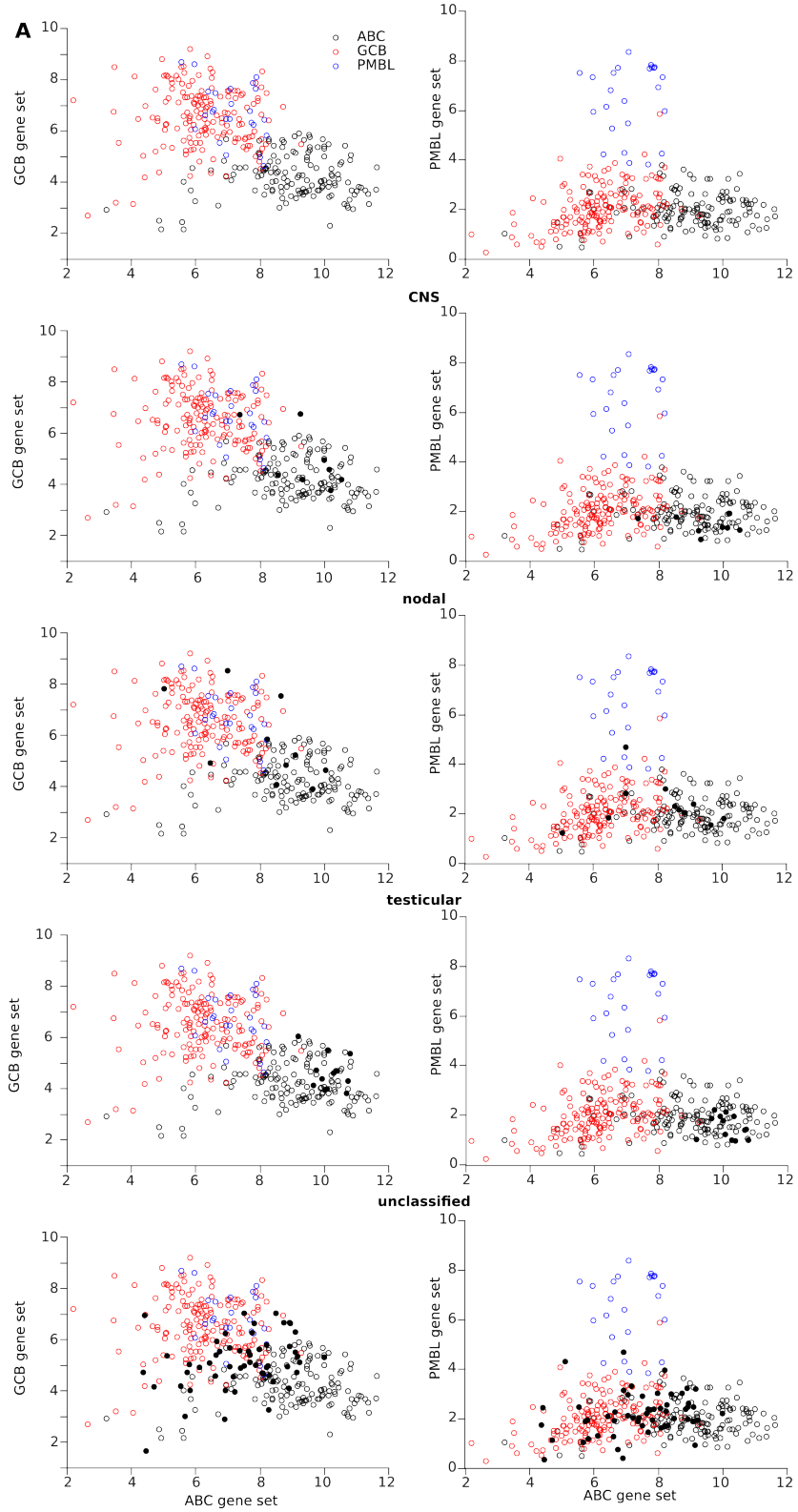
We considered the issue of false discovery rates and checked whether the test discriminates biologically relevant and irrelevant sets of data. To address this, we compared the scores of actual NHL samples for all gene sets defined in GO and KEGG to irrelevant scores for enrichment of (n=1000, size 10-1000) random gene sets in (n=500) simulated samples. All the corresponding SES were computed and the respective distributions were compared. The right skewed distribution of scores from empirical data compared to all others (e.g. 29% of SES$>$1.3 for KEGG and GO gene sets in NHL samples vs$<$5% of SES in all other settings) demonstrated this test discriminates relevant and irrelevant sets of data (Supplemental Figure 2A). Despite the expected effect of the data conversion to ranks, we considered the confounding influence of batch effects for assigning consistent SES though distinct datasets. To determine whether the independent normalization from different studies introduced significant differences of gene expression rank in the same type of samples, we computed the correlation of all gene ranks though three independent studies of CLL samples. The high Pearson correlation (0.895 to 0.966, Supplemental Figure 2B) of these data ruled out an elusive confounding influence of independent normalization of the datasets. Furthermore, both consistency and homogeneity of scoring for gender (Figure 1), proliferation,

(Figure 1, and not shown) or metabolic gene sets (Figure 2) in the 1446 samples from the 33 different GSE datasets demonstrated that the SES are minimally affected by the independent pre-processing (data normalization and gene collapsing) of each series composing of the NHL dataset.

Supplemental Figures



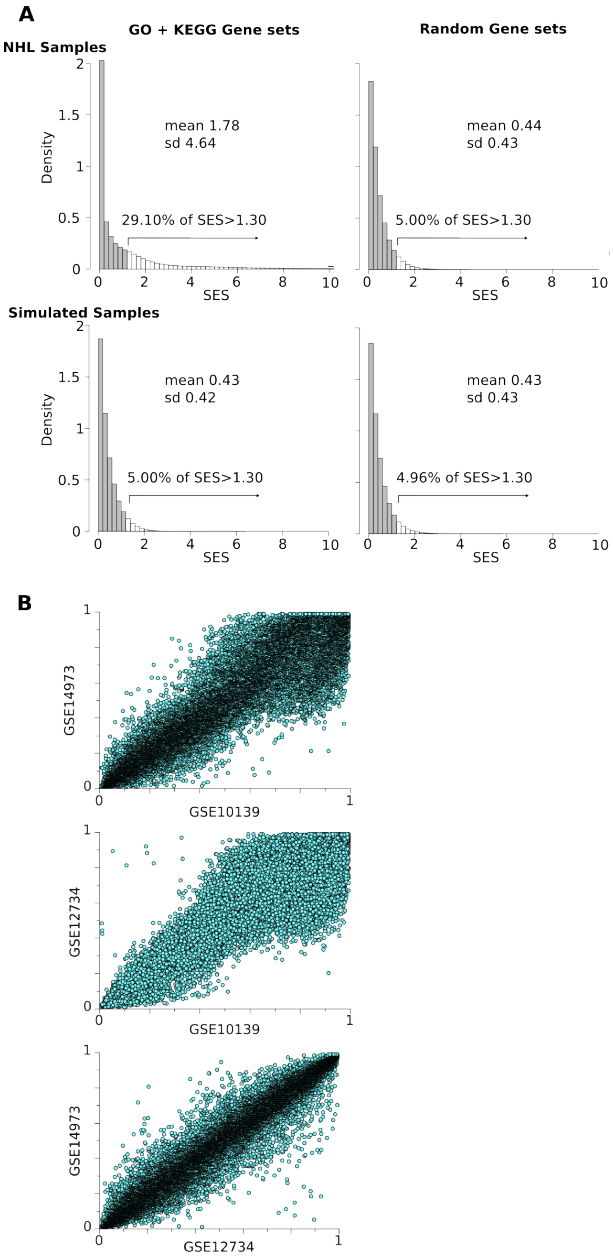
Supplemental Figure 1: Screening datasets and sample sets compatibility for merging
A Pairwise correlation matrix plots for visualization homogeneity of the NHL datasets. For each of the 47 GSE data matrices of Supplemental Table 1, the median column value of each of the 21228 gene symbol was computed. This gave 47 dataset median columns which pairwise correlations are shown as colour plots. Only 33 datasets with $r^2 > 0.8$ (group from GSE12195 to GSE65135) were considered homogeneous and kept for merging. **B,C** Pairwise correlation matrix at the sample level for respectively, normal B cells (**B**) and NHL samples (**C**) from the above 33 datasets. The samples with correlation above 0.8 were kept from each group, and pooled together in the integrated NHL dataset.



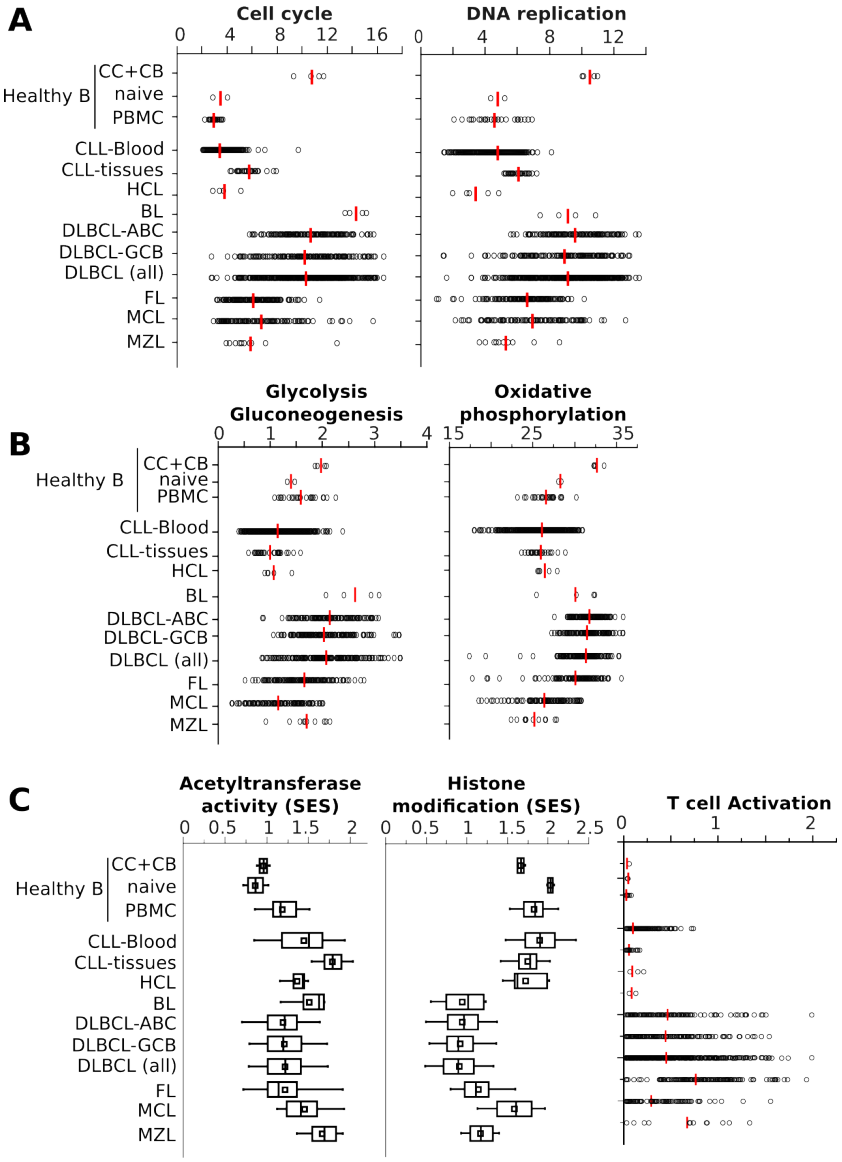
B

ABC	GCB	PMBL
BATF	ACP1	BST2
BCL2	CSTB	CCL17
CCND2	DCK	CDKN1A
CLINT1	FAM3C	E2F2
CSNK1E	GPR18	FAS
CXCR4	HLA-DMB	FCER2
FUT8	HLA-DQ1A	FGFR1
GOT2	HMG14	FLJ10420
IGHM	ITPKB	FNBP1
MARCKS	KRT8	IL13RA1
P2RX5	LMO2	IL23A
PIM1	LRMP	IL4I1
PIM2	MYBL1	ITGAM
PRKCB1	PLEK	JAK2
PTPN1	PSIP2	LY75
SLA	REL	MAL
SP140	RFTN1	MEF2A
SPIB	RPL21	MST1R
TCF4	RTVP1	NFKB2
TPM2	SCA1	PDL2
XBP1	SLAM	RGS9
	SPINK2	SAMSN1
	VCL	SETD1A
		SLAM
		SNFT
		SPINT2
		TLR7
		TNFRSF17
		TNFRSF6
		TNFRSF8
		TNFSF4
		TRAF1

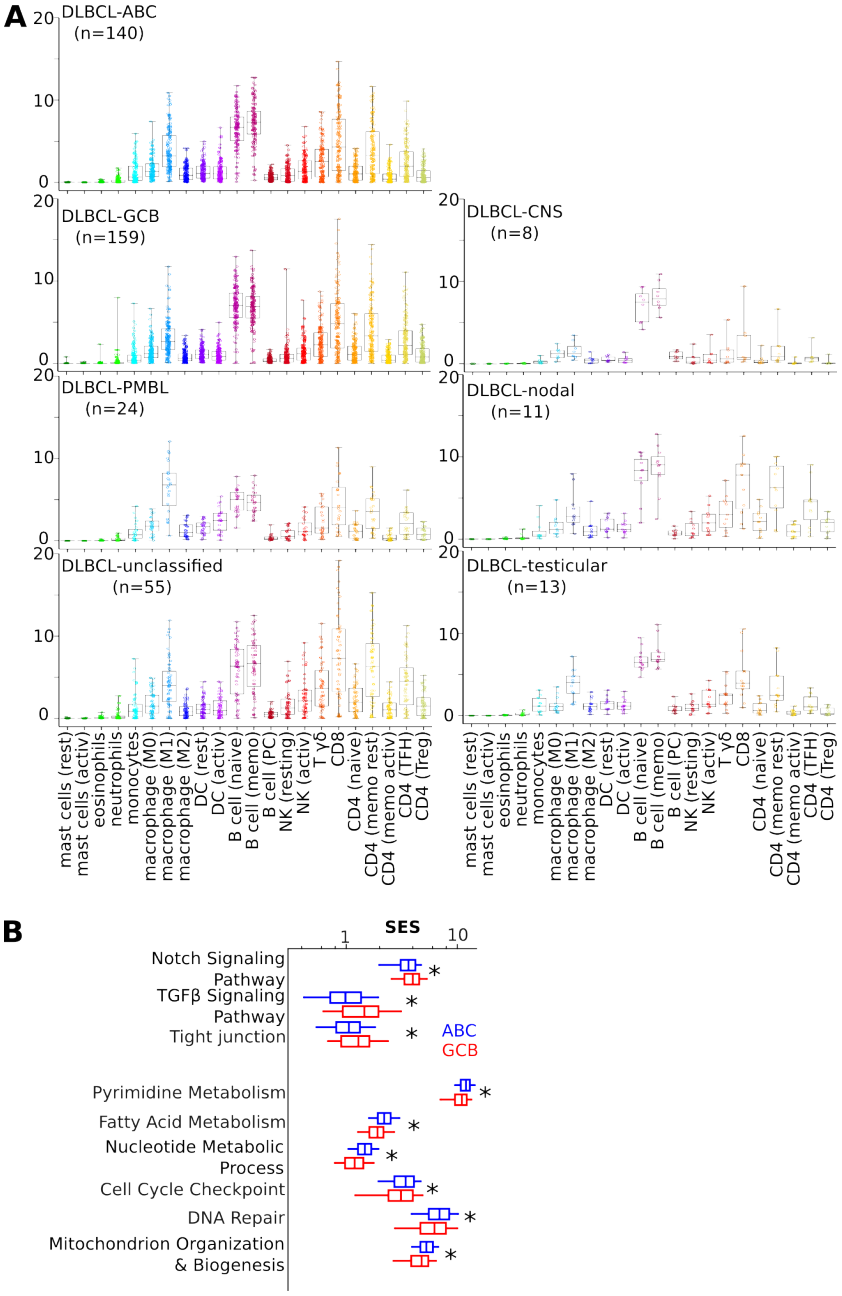
Supplemental Figure 2: SES scores of DLBCL subtype gene sets
A Plots of SES for the ABC gene set vs GCB gene set (left) and ABC gene set vs PMBL gene set (right) in annotated subtypes of DLBCL samples: ABC (empty black dots), GCB (empty red dots) and PMBL (empty blue dots). Each line of panel specifies the clinical annotation of samples shown in pannels as full black dots. This indicates that all but one CNS DLBCL and all testicular DLBCL are of ABC type, and both nodal DLBCL and unclassified DLBCL comprise samples from both ABC and GCB subtypes. **B** The ABC, GCB and PMBL gene sets as defined in.[14]



Supplemental Figure 3: Scoring gene set enrichment in samples
A Distribution of Wilcoxon-based SES of the NHL and simulated samples for the actual (GO and KEGG) and random gene sets. Shown are the % of scores above 1.3 (corresponding to p ¡ 0.05). **B** Comparison of whole transcriptome gene expression ranks in three CLL datasets RMA-normalized independently of each other. Shown are *per* gene median of ranks from all samples of the specified series (GSE), genes are decreaseing-ordered along the [0-1] segment: highest expression level=0, lowest expression level=1. The Pearson correlation of ranks are GSE10139 *vs* GSE14973: R=0.895, GSE10139 *vs* GSE: R=0.904 and GSE14973 *vs* GSE12734: R=0.966.



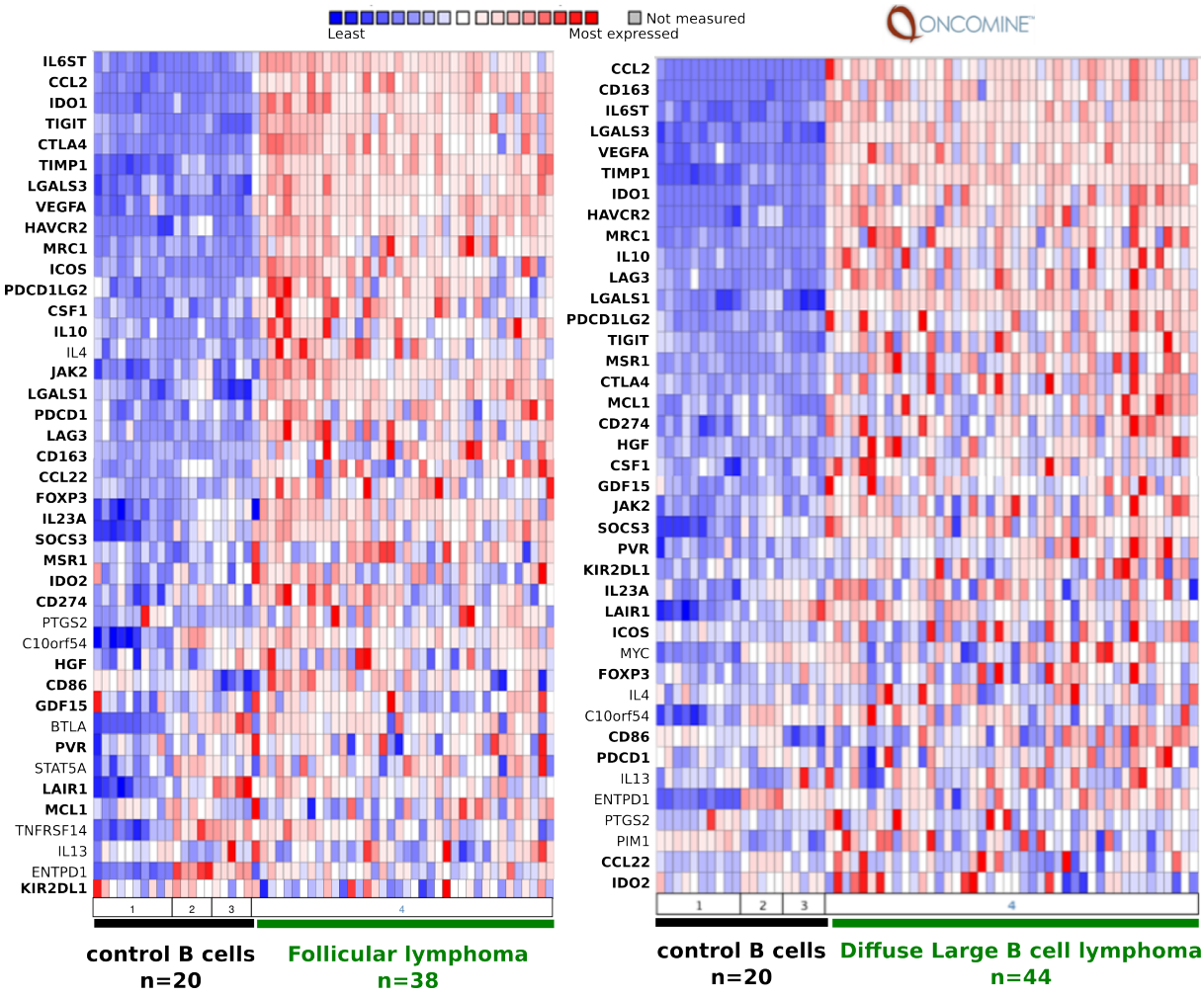
Supplemental Figure 4: SES of the NHL dataset match to the phenotype of these samples
Sample enrichment scores (SES) were computed for each of the 1446 samples of the NHL data set and for the gene sets indicated on the Y axis. Each sample is shown by a dot, red bars are means of the specified group. CC: centroblasts, CB: centrocytes. SES for gene sets related to proliferation (**A**), metabolic (**B**) phenotype. **C**: SES for 'acetyltransferase activity' and 'histone modification' and 'T cell activation' pathways. Nearly all SES for T cell activation were low in the dataset as expected for samples essentially composed of B cells, nevertheless, the scores of FL (group mean of SES= 0.8) and DLBCL (group mean of SES= 0.5) suggested the presence of some activated T cells in these samples. On the other hand, the SES for histone modifications and acetyl transferase activities are higher in CLL than in DLBCL and FL (not shown), possibly reflecting the frequent genomic deletions or inactivating mutations of histone acetyl transferases and histone methyl transferases in these lymphomas.[15,16]



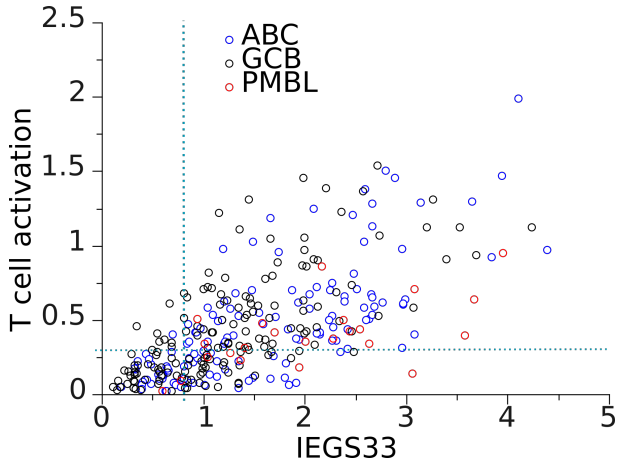
Supplemental Figure 5:Comparison of DLBCL subtypes

The SES of the various DLBCL subtypes for the 22 leucocyte subsets evidenced the B cell origin of these lymphomas, while plots of scores for B cell-nave versus B cell-memory illustrated the distinct signature of each subtype except for ABC and testicular DLBCL (**A**). Plots for myeloid vs (non-B) lymphoid subsets confirmed the relatedness of ABC and testicular DLBCL, and refined the analysis of the subtypes. Summarizing, ABC, testicular, and GCB have very similar lymphoid and myeloid compositions (e.g. mean of SES for CD8 T cells = 4.8 in ABC versus 4.9 in GCB, Wilcoxon P = 0.9). PMBL have a similar lymphoid level too, but a myeloid enrichment of M1-type macrophages and activated DC (unpaired Student P= 0.0033). Both nodal and unclassified DLBCL have similar myeloid profiles as ABC, GCB and testicular DLBCL, but twice higher scores for all lymphoid subsets (both NK, CD4, CD8 and gd) (Wilcoxon P=0.003 to 0,0001 according to subset), suggesting they have the richest TILs compartment. By contrast, the CNS subtype of DLBCL has the same B cell signature as the other DLBCL but the SES of all lymphoid subsets are strikingly lower, suggesting they have little immune cell infiltrate in line with brains unique lymphoid composition.[17,18]

We refined the analysis of GCB and ABC DLBCL by comparing their SES for various metabolic and signaling gene sets (**B**). Many KEGG and GO gene sets scored similarly in both subtypes however, including B cell receptor signaling pathway (P= 0.36), oxidative phosphorylation (P= 0.11) or glycolysis and gluconeogenesis (P= 0.06). Nevertheless, the DLBCL ABC scored significantly higher than DLBCL GCB for gene sets reflecting activated state[19,20] such as: pyrimidin metabolism (mean of SES = 12 versus 10.8, respectively, unpaired Student P= 10^{-9}), fatty acid metabolism (mean of SES = 2.3 versus 1.9, respectively, unpaired Student P= $3x10^{-9}$),nucleotide metabolic process (mean of SES = 1.5 versus 1.2, respectively, unpaired Student P=$7x10^{-13}$),cell cycle checkpoint (mean of SES = 3.4 versus 3.1, respectively, unpaired Student p= 0.02), DNA repair (mean of SES = 7.0 versus 6.3, respectively, unpaired Student p= 0.005) and mitochondrion organization and biogenesis (mean of SES = 5.4 versus 4.7, respectively, unpaired Student p= 10^{-8}), among others. On the other hand, the DLBCL GCB scored significantly higher than DLBCL ABC for Notch signaling (mean of SES = 4.0 versus 3.6, respectively, P= 0.0001) TGFb signaling (mean of SES = 1.6 versus 1.1, respectively, P= 10^{-9}), two pathways recently found activated in DLBCL[21–23] and Tight junction (mean of SES = 1.4 versus 1.1, respectively, P= $8x10^{-5}$) in line with its germinal center cell of origin.[19]



Supplemental Figure 6: Per-gene analysis of immune escape gene up-regulation in FL and DLBCL samples (n=82).
Oncomine-based comparison (`https://www.oncomine.org/resource/login.html`) of log2 median-centered intensities of the immune escape genes (defined in[24]) most differentially expressed between control B cells, FL and DLBCL samples from GEO dataset GSE12195.[25] These immune escape genes are shown above by decreasing order of significance when comparing samples groups, TIMP1 is up-regulated by all the FL and DLBCL samples relative to normal control B cells, LGALS3 by 98%, IDO1 by 90%, HAVCR2 by 91%, TIGIT by 83%, CTLA4 by 78%, PDCD1 by 67%, PDCD1LG1 by 70% and CD274 (PDCD1LG2) by 80% of FL and DLBCL samples. Bold text indicate genes selected for Immune Escape Gene Set 33 (IEGS33).



Supplemental Figure 7: Stages of immune escape in the DLBCL samples (n=385).
SES dot plots for IEGS33 vs T cell activation through the DLBCL samples. Groups of phenotypes were arbitrarily defined according to the dotted lines: group I (IEGS33- T cell activation-), group II (IEGS33- T cell activation+), group III (IEGS33+ T cell activation+), group IV (IEGS33+ T cell activation-).

Supplemental Tables

GSE number	Compatible	Cancer type	First author	reference
GSE10139	Yes	CLL	Friedman DR	[26]
GSE10524	Yes	DLBCL	Booman M	[27]
GSE10846	Yes	DLBCL	Lenz G	[28]
GSE11318	Yes	DLBCL	Lenz G	[29]
GSE12195	Yes	DLBCL	Compagno M	[25]
GSE12453	No	FL, BL, DLBCL	Brune V	[30]
GSE12734	Yes	DLBCL	Stamatopoulos B	[31]
GSE14879	No	ALCL	Eckerle S	[32]
GSE14973	Yes	CLL	Stamatopoulos B	[33]
GSE15490	Yes	CLL	Shehata M	[34]
GSE15913	Yes	CLL	Giannopoulos K	[35]
GSE16024	Yes	FL, MCL	Hamoudi R	
GSE16455	Yes	MCL	Fernandez V	[36]
GSE17372	Yes	BL, DLBCL	Deffenbacher KE	[37]
GSE18736	Yes	FL, MCL	Hamoudi RA	[38]
GSE19243	No	MCL	Leshchenko VV	[39]
GSE21029	Yes	CLL	Herishanu Y	[40]
GSE21452	Yes	MCL	Hartmann EM	[41]
GSE21554	Yes	MZL	Watkins AJ	[42]
GSE22762	Yes	CLL	Herold T	[43]
GSE25571	Yes	CLL	Herold T	[44]
GSE25613	No	MCL	Beltran E	[45]
GSE26526	No	CLL	Saiya-Cork K	[46]
GSE26725	No	CLL	Vargova K	[47]
GSE27858	Yes	CLL	Durig J	[48]
GSE28654	Yes	CLL	Trojani A	[49]
GSE29605	Yes	CLL	Davidson-Mocada JK	
GSE31048	Yes	CLL	Wang L	[50]
GSE31312	No	DLBCL	Frei E	[51]
GSE33135	Yes	CLL	Baptista MJ	[52]
GSE34171	Yes	DLBCL	Monti S	[53]
GSE34339	No	DLBCL	Hogfeldt T	
GSE35426	Yes	MZL	Watkins AJ	[42]
GSE35935	Yes	CLL	Chiaretti S	
GSE36000	Yes	MCL	Navarro A	[54]
GSE36907	No	CLL	Seifert M	[55]
GSE37168	Yes	CLL	Landau DA	[56]
GSE38816	Yes	FL	Green MR	[57]
GSE39671	No	CLL	Chuang HY	[58]
GSE52435	No	MCL	Rahal R	[59]
GSE53786	Yes	DLBCL	Scott DW	[60]
GSE53820	Yes	FL	Brodtkorb M	[61]
GSE55267	Yes	FL	Guo S	[62]
GSE6338	No	PTCL	Piccaluga PP	[63]
GSE64555	No	DLBCL	Linton K	
GSE65135	Yes	FL, DLBCL, MZL	Newman AM	[5]
GSE9250	No	CLL	Ouillette P	[64]

Supplemental table 1: List of studies tested for their compatibility

Gene	Protein	Chr.
AZI2	5-Azacytidine-Induced Gene 2	3
EREG	Epiregulin	4
GPM6B	Neuronal Membrane Glycoprotein M6b	X
XIST	X Inactivation-Specific Transcript	X

Supplemental table 2A: female specific gene set for showing highly sex-biased expression patterns in Affymetrix HG U133 plus microarrays from human tissues[65]

Gene	Protein	Chr.
CD24L4	CD24 molecule like 4	Y
CYORF14	Chromosome Y open reading frame 14	Y
CYORF15A	Chromosome Y open reading frame 15A	Y
CYORF15B	Chromosome Y open reading frame 15B	Y
DDX3Y	Dead/H Box 3, Y-Linked	Y
EIF1AY	Eukaryotic Translation Initiation Factor 1a, Y-Linked	Y
JARID1D	Jumonji, AT rich interactive domain Ad	Y
PRKY	Protein Kinase, Y-Linked	Y
RPS4Y1	Ribosomal Protein S4, Y-Linked	Y
USP9Y	Ubiquitin-Specific Protease 9	Y
UTY	Ubiquitously Transcribed Tetratricopeptide Repeat Gene On Y Chromosome	Y
ZFY	Zinc Finger Protein, Y-Linked	Y

Supplemental table 2B: male specific gene set for showing highly sex-biased expression patterns in Affymetrix HG U133 plus microarrays from human tissues[65]

GENE ID	protein, alias names
CCL2	Chemokine (CC motif) ligand 2, MCP1
CCL22	Chemokine (CC motif) ligand 22
CD163	CD163
CD274	Programmed cell death 1 ligand 1
CD86	CTLA4 ligand, B7.2
CSF1	”Colony stimulating factor 1 (macrophage), MCSF1”
CTLA4	Cytotoxic T-lymphocyte-associated protein 4, CD152
FOXP3	Forkhead box P3
GDF15	Growth differentiation factor 15
HAVCR2	TIM3
HGF	Hepatocyte growth factor
ICOS	CD278, Inducible T-cell co-stimulator
IDO1	Indoleamine 2,3 dioxygenase 1
IDO2	Indoleamine 2,3 dioxygenase 2
IL10	Interleukin 10
IL23A	Interleukin 23
IL6ST	Interleukin 6 signal transducer, CD130
JAK2	Janus kinase 2
KIR2DL1	CD158A, NK cell inhibitory receptor p58
LAG3	Lymphocyte-activation gene 3, CD223
LAIR1	Leukocyte-associated Ig-like receptor 1
LGALS1	Galectin 1
LGALS3	Galectin 3
MCL1	Myeloid cell leukemia 1, BCL2-related
MRC1	Mannose receptor C type 1, CD206
MSR1	Macrophage scavenger receptor 1, CD204
PDCD1	PD-1, programmed cell death 1, CD279
PDCD1LG2	CD273,PD-L2, PD-1L2, butyrophilin B7DC,
PVR	Polyovirus receptor, CD155
SOCS3	Puppressor of cytokine signaling 3
TIGIT	T cell immunoreceptor with Ig and ITIM domains
TIMP1	Tissue inhibitor of metalloproteinase 1
VEGFA	Vascular endothelial growth factor A

Supplemental table 3: The IEGS33 immune escape gene set

Group / DLBCL	I	II	III	IV
ABC	34	1	99	33
GCB	46	11	103	35
PMBL	2	0	16	5
Unclassified	6	2	42	13
NA	10	0	28	20

Fisher's Exact Test p-value=0.042

Group / IPI	I	II	III	IV
low	37	9	124	39
Low-intermediate	26	5	64	23
High-intermediate	17	0	31	11
High	2	0	7	1
NA	16	0	62	32

Fisher's Exact Test p-value=0.6222

Supplemental Table 4: Distribution of patients according to groups defined by IEGS33 and T-cell activation gene set scores. Tables show results according to different clinical parameters: type of DLBCL (top) and International Prognostic Index (IPI, bottom).

Supplemental References

References

[1] R. Edgar, M. Domrachev, and A. E. Lash, Gene expression omnibus: Ncbi gene expression and hybridization array data repository, Nucleic acids research 2002; 30:207–210.

[2] G. O. Consortium et al., The gene ontology (go) database and informatics resource, Nucleic acids research 2004; 32:D258–D261.

[3] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J. P. Mesirov, Molecular signatures database (msigdb) 3.0, Bioinformatics 2011; 27:1739–1740.

[4] A. J. Gentles, A. M. Newman, C. L. Liu, S. V. Bratman, W. Feng, D. Kim, V. S. Nair, Y. Xu, A. Khuong, C. D. Hoang *et al.*, The prognostic landscape of genes and infiltrating immune cells across human cancers, Nature medicine 2015; 21:938–945.

[5] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, Nature methods 2015; 12:453–457.

[6] A. Tsodikov, A. Szabo, and D. Jones, Adjustments and measures of differential expression for microarray data, Bioinformatics 2002; 18:251–260.

[7] P. Warnat, R. Eils, and B. Brors, Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes, BMC bioinformatics 2005; 6:265.

[8] B. Ycart, K. Charmpi, S. Rousseaux, and J.-J. Fournié, Large scale statistical analysis of geo datasets, arXiv preprint arXiv:1410.2585 2014; .

[9] B. Ycart, F. Pont, and J.-J. Fournié, Curbing false discovery rates in interpretation of genome-wide expression profiles, Journal of biomedical informatics 2014; 47:58–61.

[10] K. Charmpi and B. Ycart, Weighted kolmogorov smirnov testing: an alternative for gene set enrichment analysis, arXiv preprint arXiv:1410.1620 2014; .

[11] X. Cui, G. A. Churchill, et al., Statistical tests for differential expression in cdna microarray experiments, Genome Biol 2003; 4:210.

[12] R. A. Irizarry, C. Wang, Y. Zhou, and T. P. Speed, Gene set enrichment analysis made simple, Statistical methods in medical research 2009; 18:565–575.

[13] V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila *et al.*, Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes, Nature genetics 2003; 34:267–273.

[14] G. Wright, B. Tan, A. Rosenwald, E. H. Hurt, A. Wiestner, and L. M. Staudt, A gene expression-based method to diagnose clinically distinct subgroups of diffuse large b cell lymphoma, PNAS 2003; 100:9991–9996.

[15] R. D. Morin, M. Mendez-Lago, A. J. Mungall, R. Goya, K. L. Mungall, R. D. Corbett, N. A. Johnson, T. M. Severson, R. Chiu, M. Field *et al.*, Frequent mutation of histone-modifying genes in non-hodgkin lymphoma, Nature 2011; 476:298–303.

[16] L. Pasqualucci, V. Trifonov, G. Fabbri, J. Ma, D. Rossi, A. Chiarenza, V. A. Wells, A. Grunn, M. Messina, O. Elliot *et al.*, Analysis of the coding genome of diffuse large b-cell lymphoma, Nature genetics 2011; 43:830–837.

[17] A. Louveau, T. H. Harris, and J. Kipnis, Revisiting the mechanisms of cns immune privilege, Trends in immunology 2015; 36:569–577.

[18] A. Louveau, I. Smirnov, T. J. Keyes, J. D. Eccles, S. J. Rouhani, J. D. Peske, N. C. Derecki, D. Castle, J. W. Mandell, K. S. Lee *et al.*, Structural and functional features of central nervous system lymphatic vessels, Nature 2015; .

[19] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu *et al.*, Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling, Nature 2000; 403:503–511.

[20] A. Rosenwald, G. Wright, K. Leroy, X. Yu, P. Gaulard, R. D. Gascoyne, W. C. Chan, T. Zhao, C. Haioun, T. C. Greiner *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, The Journal of experimental medicine 2003; 198:851–862.

[21] L. Arcaini, D. Rossi, M. Lucioni, M. Nicola, A. Bruscaggin, V. Fiaccadori, R. Riboni, A. Ramponi, V. V. Ferretti, S. Cresta *et al.*, The notch pathway is recurrently mutated in diffuse large b-cell lymphoma associated with hepatitis c virus infection, Haematologica 2015; 100:246–252.

[22] S.-y. Lee, K. Kumano, K. Nakazaki, M. Sanada, A. Matsumoto, G. Yamamoto, Y. Nannya, R. Suzuki, S. Ota, Y. Ota *et al.*, Gain-of-function mutations and copy number increases of notch2 in diffuse large b-cell lymphoma, Cancer science 2009; 100:920–926.

[23] D. Rai, S.-W. Kim, M. R. McKeller, P. L. Dahia, and R. C. Aguiar, Targeting of smad5 links microrna-155 to the tgf-β pathway and lymphomagenesis, Proceedings of the National Academy of Sciences 2010; 107:3111–3116.

[24] C. Laurent, K. Charmpi, P. Gravelle, M. Tosolini, C. Franchet, L. Ysebaert, P. Brousset, A. Bidaut, B. Ycart, and J.-J. Fourni, Several immune escape patterns in non-hodgkin's lymphomas, OncoImmunology 2015; 4:e1026530.

[25] M. Compagno, W. K. Lim, A. Grunn, S. V. Nandula, M. Brahmachary, Q. Shen, F. Bertoni, M. Ponzoni, M. Scandurra, A. Califano *et al.*, Mutations of multiple genes cause deregulation of nfκb in diffuse large b-cell lymphoma, Nature 2009; 459:717–721.

[26] D. R. Friedman, J. B. Weinberg, W. T. Barry, B. K. Goodman, A. D. Volkheimer, K. M. Bond, Y. Chen, N. Jiang, J. O. Moore, J. P. Gockerman *et al.*, A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia, Clinical Cancer Research 2009; 15:6947–6955.

[27] M. Booman, K. Szuhai, A. Rosenwald, E. Hartmann, H. C. Kluin-Nelemans, D. de Jong, E. Schuuring, and P. Kluin, Genomic alterations and gene expression in primary diffuse large b-cell lymphomas of immune-privileged sites: the importance of apoptosis and immunomodulatory pathways, The Journal of pathology 2008; 216:209–217.

[28] G. Lenz, G. Wright, S. Dave, W. Xiao, J. Powell, H. Zhao, W. Xu, B. Tan, N. Goldschmidt, J. Iqbal *et al.*, Stromal gene signatures in large-b-cell lymphomas, New England Journal of Medicine 2008; 359:2313–2323.

[29] G. Lenz, G. W. Wright, N. T. Emre, H. Kohlhammer, S. S. Dave, R. E. Davis, S. Carty, L. T. Lam, A. Shaffer, W. Xiao *et al.*, Molecular subtypes of diffuse large b-cell lymphoma arise by distinct genetic pathways, Proceedings of the National Academy of Sciences 2008; 105:13520–13525.

[30] V. Brune, E. Tiacci, I. Pfeil, C. Döring, S. Eckerle, C. J. van Noesel, W. Klapper, B. Falini, A. von Heydebreck, D. Metzler *et al.*, Origin and pathogenesis of nodular lymphocyte–predominant hodgkin lymphoma as revealed by global gene expression analysis, The Journal of experimental medicine 2008; 205:2251–2268.

[31] B. Stamatopoulos, B. Haibe-Kains, C. Equeter, N. Meuleman, A. Sorée, C. De Bruyn, D. Hanosset, D. Bron, P. Martiat, and L. Lagneaux, Gene expression profiling reveals differences in microenvironment interaction between patients with chronic lymphocytic leukemia expressing high versus low zap70 mrna, haematologica 2009; 94:790–799.

[32] S. Eckerle, V. Brune, C. Döring, E. Tiacci, V. Bohle, C. Sundström, R. Kodet, M. Paulli, B. Falini, W. Klapper *et al.*, Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to hodgkin lymphoma, Leukemia 2009; 23:2129–2138.

[33] B. Stamatopoulos, N. Meuleman, C. De Bruyn, P. Mineur, P. Martiat, D. Bron, and L. Lagneaux, Antileukemic activity of valproic acid in chronic lymphocytic leukemia b cells defined by microarray analysis, Leukemia 2009; 23:2281–2289.

[34] M. Shehata, D. Demirtas, S. Schnabl, M. Hilgarth, R. Hubmann, C. Fonatsch, I. Schwarzinger, G. Hopfinger, K. Eigenberger, D. Heintel *et al.*, Sequential gene expression profiling during treatment for identification of predictive markers and novel therapeutic targets in chronic lymphocytic leukemia, Leukemia 2010; 24:2122–2127.

[35] K. Giannopoulos, A. Dmoszynska, M. Kowal, E. Wkasik-Szczepanek, A. Bojarska-Junak, J. Rolinski, H. Döhner, S. Stilgenbauer, and L. Bullinger, Thalidomide exerts distinct molecular antileukemic effects and combined thalidomide/fludarabine therapy is clinically effective in high-risk chronic lymphocytic leukemia, Leukemia 2009; 23:1771–1778.

[36] V. Fernàndez, O. Salamero, B. Espinet, F. Solé, C. Royo, A. Navarro, F. Camacho, S. Beà, E. Hartmann, V. Amador *et al.*, Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma, Cancer research 2010; 70:1408–1418.

[37] K. E. Deffenbacher, J. Iqbal, Z. Liu, K. Fu, and W. C. Chan, Recurrent chromosomal alterations in molecularly classified aids-related lymphomas: an integrated analysis of dna copy number and gene expression, JAIDS Journal of Acquired Immune Deficiency Syndromes 2010; 54:18–26.

[38] R. Hamoudi, A. Appert, H. Ye, A. Ruskone-Fourmestraux, B. Streubel, A. Chott, M. Raderer, L. Gong, I. Wlodarska, C. De Wolf-Peeters *et al.*, Differential expression of nf-κb target genes in malt lymphoma with and without chromosome translocation: insights into molecular mechanism, Leukemia 2010; 24:1487–1497.

[39] V. V. Leshchenko, P.-Y. Kuo, R. Shaknovich, D. T. Yang, T. Gellen, A. Petrich, Y. Yu, Y. Remache, M. A. Weniger, S. Rafiq *et al.*, Genomewide dna methylation analysis reveals novel targets for drug development in mantle cell lymphoma, Blood 2010; 116:1025–1034.

[40] Y. Herishanu, P. Pérez-Galán, D. Liu, A. Biancotto, S. Pittaluga, B. Vire, F. Gibellini, N. Njuguna, E. Lee, L. Stennett *et al.*, The lymph node microenvironment promotes b-cell receptor signaling, nf-κb activation, and tumor proliferation in chronic lymphocytic leukemia, Blood 2011; 117:563–574.

[41] E. M. Hartmann, E. Campo, G. Wright, G. Lenz, I. Salaverria, P. Jares, W. Xiao, R. M. Braziel, L. M. Rimsza, W.-C. Chan *et al.*, Pathway discovery in mantle cell lymphoma by integrated analysis of high-resolution gene expression and copy number profiling, Blood 2010; 116:953–961.

[42] A. J. Watkins, R. A. Hamoudi, N. Zeng, Q. Yan, Y. Huang, H. Liu, J. Zhang, E. Braggio, R. Fonseca, L. de Leval *et al.*, An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma, PLoS ONE 2012; 7.

[43] T. Herold, V. Jurinovic, K. Metzeler, A.-L. Boulesteix, M. Bergmann, T. Seiler, M. Mulaw, S. Thoene, A. Dufour, Z. Pasalic *et al.*, An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia, Leukemia 2011; 25:1639–1645.

[44] T. Herold, V. Jurinovic, M. Mulaw, T. Seiler, A. Dufour, S. Schneider, P. M. Kakadia, M. Feuring-Buske, J. Braess, K. Spiekermann *et al.*, Expression analysis of genes located in the minimally deleted regions of 13q14 and 11q22-23 in chronic lymphocytic leukemiaun-expected expression pattern of the rho gtpase activator arhgap20, Genes, Chromosomes and Cancer 2011; 50:546–558.

[45] E. Beltran, V. Fresquet, J. Martinez-Useros, J. A. Richter-Larrea, A. Sagardoy, I. Sesma, L. L. Almada, S. Montes-Moreno, R. Siebert, S. Gesk *et al.*, A cyclin-d1 interaction with bax underlies its oncogenic role and potential as a therapeutic target in mantle cell lymphoma, Proceedings of the National Academy of Sciences 2011; 108:12461–12466.

[46] K. Saiya-Cork, R. Collins, B. Parkin, P. Ouillette, E. Kuizon, L. Kujawski, H. Erba, E. Campagnaro, K. Shedden, M. Kaminski *et al.*, A pathobiological role of the insulin receptor in chronic lymphocytic leukemia, Clinical Cancer Research 2011; 17:2679–2692.

[47] K. Vargova, N. Curik, P. Burda, P. Basova, V. Kulvait, V. Pospisil, F. Savvulidi, J. Kokavec, E. Necas, A. Berkova *et al.*, Myb transcriptionally regulates the mir-155 host gene in chronic lymphocytic leukemia, Blood 2011; 117:3816–3825.

[48] J. Dürig, U. Dührsen, L. Klein-Hitpass, J. Worm, J. R. Hansen, H. Ørum, and M. Wissenbach, The novel antisense bcl-2 inhibitor spc2996 causes rapid leukemic cell clearance and immune activation in chronic lymphocytic leukemia, Leukemia 2011; 25:638–647.

[49] A. Trojani, B. Di Camillo, A. Tedeschi, M. Lodola, S. Montesano, F. Ricci, E. Vismara, A. Greco, S. Veronese, A. Orlacchio *et al.*, Gene expression profiling identifies arsd as a new marker of disease progression and the sphingolipid metabolism as a potential novel metabolism in chronic lymphocytic leukemia., Cancer biomarkers: section A of Disease markers 2010; 11:15–28.

[50] L. Wang, A. K. Shalek, M. Lawrence, R. Ding, J. T. Gaublomme, N. Pochet, P. Stojanov, C. Sougnez, S. A. Shukla, K. E. Stevenson *et al.*, Somatic mutation as a mechanism of wnt/β-catenin pathway activation in cll, Blood 2014; blood–2014.

[51] E. Frei, C. Visco, Z. Xu-Monette, S. Dirnhofer, K. Dybkær, A. Orazi, G. Bhagat, E. Hsi, J. van Krieken, M. Ponzoni *et al.*, Addition of rituximab to chemotherapy overcomes the negative prognostic impact of cyclin e expression in diffuse large b-cell lymphoma, Journal of clinical pathology 2013; 66:956–961.

[52] M. J. Baptista, A. Muntañola, E. Calpe, P. Abrisqueta, O. Salamero, E. Fernández, C. Codony, E. Giné, S. G. Kalko, M. Crespo *et al.*, Differential gene expression profile associated to apoptosis induced by dexamethasone in cll cells according to ighv/zap-70 status, Clinical Cancer Research 2012; 18:5924–5933.

[53] S. Monti, B. Chapuy, K. Takeyama, S. J. Rodig, Y. Hao, K. T. Yeda, H. Inguilizian, C. Mermel, T. Currie, A. Dogan *et al.*, Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large b cell lymphoma, Cancer cell 2012; 22:359–372.

[54] A. Navarro, G. Clot, C. Royo, P. Jares, A. Hadzidimitriou, A. Agathangelidis, V. Bikos, N. Darzentas, T. Papadaki, I. Salaverria *et al.*, Molecular subsets of mantle cell lymphoma defined by the ighv mutational status and sox11 expression have distinct biologic and clinical features, Cancer research 2012; 72:5307–5316.

[55] M. Seifert, L. Sellmann, J. Bloehdorn, F. Wein, S. Stilgenbauer, J. Dürig, and R. Küppers, Cellular origin and pathophysiology of chronic lymphocytic leukemia, The Journal of experimental medicine 2012; 209:2183–2198.

[56] D. A. Landau, S. L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M. S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang *et al.*, Evolution and impact of subclonal mutations in chronic lymphocytic leukemia, Cell 2013; 152:714–726.

[57] M. R. Green, A. J. Gentles, R. V. Nair, J. M. Irish, S. Kihira, C. L. Liu, I. Kela, E. S. Hopmans, J. H. Myklebust, H. Ji *et al.*, Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma, Blood 2013; 121:1604–1611.

[58] H.-Y. Chuang, L. Rassenti, M. Salcedo, K. Licon, A. Kohlmann, T. Haferlach, R. Foà, T. Ideker, and T. J. Kipps, Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression, Blood 2012; 120:2639–2649.

[59] R. Rahal, M. Frick, R. Romero, J. M. Korn, R. Kridel, F. C. Chan, B. Meissner, H.-e. Bhang, D. Ruddy, A. Kauffmann *et al.*, Pharmacological and genomic profiling identifies nf-κ b-targeted treatment strategies for mantle cell lymphoma, Nature medicine 2014; 20:87–92.

[60] D. W. Scott, G. W. Wright, P. M. Williams, C.-J. Lih, W. Walsh, E. S. Jaffe, A. Rosenwald, E. Campo, W. C. Chan, J. M. Connors *et al.*, Determining cell-of-origin subtypes of diffuse large b-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue, Blood 2014; 123:1214–1217.

[61] M. Brodtkorb, O. C. Lingjærde, K. Huse, G. Trøen, M. Hystad, V. I. Hilden, J. H. Myklebust, E. Leich, A. Rosenwald, J. Delabie *et al.*, Whole-genome integrative analysis reveals expression signatures predicting transformation in follicular lymphoma, Blood 2014; 123:1051–1054.

[62] S. Guo, J. K. Chan, J. Iqbal, T. McKeithan, K. Fu, B. Meng, Y. Pan, W. Cheuk, D. Luo, R. Wang *et al.*, Ezh2 mutations in follicular lymphoma from different ethnic groups and associated gene expression alterations, Clinical Cancer Research 2014; 20:3078–3086.

[63] P. P. Piccaluga, C. Agostinelli, A. Califano, M. Rossi, K. Basso, S. Zupo, P. Went, U. Klein, P. L. Zinzani, M. Baccarani *et al.*, Gene expression analysis of peripheral t cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets, Journal of Clinical Investigation 2007; 117:823.

[64] P. Ouillette, H. Erba, L. Kujawski, M. Kaminski, K. Shedden, and S. N. Malek, Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14, Cancer research 2008; 68:1012–1021.

[65] J. Isensee, H. Witt, R. Pregla, R. Hetzer, V. Regitz-Zagrosek, and P. R. Noppinger, Sexually dimorphic gene expression in the heart of mice and men, Journal of molecular medicine 2008; 86:61–74.