bjh research paper

Genomic lesions associated with a different clinical outcome in diffuse large B-Cell lymphoma treated with R-CHOP-21

Marta Scandurra,[1]* Michael Mian,[1,2]* Timothy C. Greiner,[3] Paola M. V. Rancoita,[1,4] Cassio P. De Campos,[1,4] Wing C. Chan,[3] Julie M. Vose,[3] Ekaterina Chigrinova,[1] Giorgio Inghirami,[5] Annalisa Chiappella,[6] Luca Baldini,[7] Maurilio Ponzoni,[8] Andres J.M. Ferreri,[8] Silvia Franceschetti,[9] Gianluca Gaidano,[9] Santiago Montes-Moreno,[10] Miguel A.Piris,[10] Fabio Facchetti,[11] Alessandra Tucci,[11] Josep Fr. Nomdedeu,[12] Thierry Lazure,[13] Olivier Lambotte,[13] Silvia Uccella,[14] Graziella Pinotti,[14] Giancarlo Pruneri,[15] Giovanni Martinelli,[15] Ken H. Young,[16] Maria Grazia Tibiletti,[14] Andrea Rinaldi,[1] Emanuele Zucca,[1] Ivo Kwee[1,4] and Francesco Bertoni[1]

[1]Laboratory of Experimental Oncology and Lymphoma Unit, Oncology Institute of Southern Switzerland (IOSI), Bellinzona, Switzerland, [2]Division of Haematology, Azienda Ospedaliera S. Maurizio, Bolzano/Bozen, Italy, [3]Department of Pathology and Microbiology, University of Nebraska, Omaha, NE, USA, [4]Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Manno, Switzerland, [5]Department of Pathology and Centre for Experimental Research and Medical Studies (CeRMS), University of Turin, Turin, Italy, [6]Department of Oncology and Haematology, Haematology 2 section, AOU San Giovanni Battista, Torino, [7]UO Ematologia 1/CTMO, Università degli Studi di Milano, Dipartimento di Scienze Mediche, Ospedale Maggiore Policlinico MaRe, IRCCS, Milano, [8]San Raffaele H Scientific Institute, Pathology Unit and Unit of Lymphoid Malignancies, Milan, [9]Division of Haematology, Department of Clinical and Experimental Medicine & BRMA, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy, [10]Programa de Patologia Molecular, Centro Nacional de Investigaciones Oncologicas, Madrid, Spain, [11]Department of Pathology, University of Brescia, I Servizio di Anatomia Patologica, and Division of Haematology, Spedali Civili di Brescia, Brescia, Italy, [12]Department of Haematology and

Summary

Despite recent therapeutic improvements, the clinical course of diffuse large B-cell lymphoma (DLBCL) still differs considerably among patients. We conducted this retrospective multi-centre study to evaluate the impact of genomic aberrations detected using a high-density genome wide-single nucleotide polymorphism-based array on clinical outcome in a population of DLBCL patients treated with R-CHOP-21 (rituximab, cyclophosphamide, doxorubicine, vincristine and prednisone repeated every 21 d). 166 DNA samples were analysed using the GeneChip Human Mapping 250K NspI. Genomic anomalies were analysed regarding their impact on the clinical course of 124 patients treated with R-CHOP-21. Unsupervised clustering was performed to identify genetically related subgroups of patients with different clinical outcomes. Twenty recurrent genetic lesions showed an impact on the clinical course. Loss of genomic material at 8p23.1 showed the strongest statistical significance and was associated with additional aberrations, such as 17p- and 15q-. Unsupervised clustering identified five DLBCL clusters with distinct genetic profiles, clinical characteristics and outcomes. Genetic features and clusters, associated with a different outcome in patients treated with R-CHOP, have been identified by arrayCGH.

Keywords: prognosis, lymphoma, comparative genomic hybridization, microarray, hepatitis C virus.

© 2010 Blackwell Publishing Ltd,

First published online 31 August 2010
doi:10.1111/j.1365-2141.2010.08326.x

Laboratori d'Hematologia, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain, [13]*Departments of Internal Medicine and of Pathology, University Hospital of Bicêtre, AP/HP, Le Kremlin Bicêtre, France,* [14]*Anatomic Pathology Unit and Department of Oncology, University of Insubria, Ospedale di Circolo, Varese,* [15]*Division of Pathology and Laboratory Medicine and Division of Haematology, European Institute of Oncology, Milan, Italy, and* [16]*Department of Pathology & Laboratory Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA*

Received 24 February 2010; accepted for publication 14 May 2010
Correspondence: Francesco Bertoni, MD, Laboratory of Experimental Oncology, Oncology Institute of Southern Switzerland (IOSI), via Vincenzo Vela 6, 6500 Bellinzona, Switzerland. E-mail: frbertoni@mac.com
*Equally contributed.

Non-Hodgkin lymphomas (NHLs) represent the fifth most frequent cancer (Jemal *et al*, 2009). Diffuse large B-cell lymphoma (DLBCL) constitutes the most common subtype, accounting for 25–30% of adult NHLs in Western countries, and is characterized by heterogeneous histopathological, clinical and biological features (Armitage, 2007; Lenz & Staudt, 2010). DLBCL represents a potentially curable disease when treated with R-CHOP-21 (rituximab, cyclophosphamide, doxorubicine, vincristine and prednisone repeated every 21 d) for 6–8 cycles. Nevertheless, the progression-free survival (PFS) in elderly patients is still below 55% and in younger patients does not exceed 85%.

Gene expression profiling (GEP) patterns and genomic aberrations have been identified to have an impact on the clinical course of DLBCL patients treated with CHOP (Lenz *et al*, 2008a; Lenz & Staudt, 2010). Differently from GEP data (Lenz *et al*, 2008b), no information is available on the prognostic impact of genomic aberrations detected using high resolution single nucleotide polymorphism (SNP) array in DLBCL patients treated with R-CHOP.

In this study, we analysed the impact of recurrent aberrations revealed by a high-density genome wide SNP-based array on clinical outcome in DLBCL patients treated with R-CHOP-21.

Materials and methods

Tumour panel

DNA was extracted from 166 frozen tumour biopsies of DLBCL, taken at diagnosis as previously described (Capello *et al*, 2010; Rinaldi *et al*, 2010). Cases were selected for the study based upon the availability of frozen material, and were required to have a fraction of malignant cells in the pathological specimen representing >70% of overall cellularity as determined by morphological and immunophenotypic studies. Cases of primary mediastinal large B-cell lymphoma, human immunodeficiency virus-related DLBCL and post-transplant DLBCL were excluded. Therapy was R-CHOP21 in 137 cases (13 with no follow-up data available) and other than R-CHOP21 in 14 (four, R-CHOP14; two, CHOP; four high-dose chemotherapy; four, other), and unknown in 15. As previously reported (Capello *et al*, 2010; Rinaldi *et al*, 2010), the distinction of DLBCL cases in germinal centre B-cell-like (GCB) or in non-GCB, was available in 110 cases: assessed in 52 cases by immunohistochemistry according to the algorithm of Hans *et al* (2004) (25 GCB, 27 non-GCB) and in 58 by GEP (Lenz *et al*, 2008b) (31 GCB, 27 non-GCB). All patients provided informed consent in accordance

 © 2010 Blackwell Publishing Ltd,

with the hospital's institutional review board and the Declaration of Helsinki. The study was approved by the Bellinzona Ethics Committee.

SNP array-CGH analysis and data mining

DNA samples were analysed using the GeneChip Human Mapping 250K NspI (Affymetrix, Santa Clara, CA, USA), as previously described (Capello *et al*, 2010; Rinaldi *et al*, 2010). Data acquisition was performed using the GeneChip operating software (GCOS, Affymetrix, Santa Clara, CA, USA) version 1.4 and genotyping analysis software (GTYPE, Affymetrix, Santa Clara, CA, USA) version 4.1. Genotype calls were calculated using the BRLMM algorithm (Bayesian Robust Linear Model with Mahalanobis distance classifier). Mapping data for probes were derived from the National Center for Biotechnology Information (NCBI) Human Genome Build 36, as provided by Affymetrix. The modified Bayesian Piecewise Regression (mBPCR) method (Rancoita *et al*, 2009) was used to estimate the copy number (CN) starting from raw CN values obtained with Affymetrix CNAT 4·01 using a Gaussian bandwidth of 0 Kb and 46 Caucasian normal female samples of the HapMap Project as reference samples. After normalization of each profile to a median log2-ratio of zero, thresholds for loss and gain defined as six times the median absolute deviation symmetrically around zero with an associated *P* value lower than <0·001 after Bonferroni multiple test correction. Loss of heterozygosity (LOH) profiles were obtained applying the method with haplotype correction for tumour-only LOH inference available in the dchip software (Beroukhim *et al*, 2006) using as reference the 60 CEPH (Centre d'Etude du Polymorphisme Humain) parents of the HapMap Project and computing the allele A frequency from the data. Figure S1 shows examples of the genomic lesions with raw and inferred CN, and LOH estimate. Frequency plots for DNA gains, DNA losses and LOH were calculated from mBPCR segmented CN data, thresholded for gain or loss or for LOH, without removing known copy number variations (CNV). All the 166 DLBCL profiles were used to identify the minimal recurrent alterations that are supposed to contain loci relevant for the tumour, following the algorithm developed by Lenz *et al* (2008a). Four distinct types of minimal common regions (MCR) were defined: short recurrent abnormality (SRA), long recurrent abnormality (LRA), abnormal chromosome arm (ACA) and abnormal whole chromosome (AWC). In MCR frequencies, patients carrying a lesion contained in both an SRA and in an LRA were counted for both types of MCR. For MCR occurring in at least 15% of cases, differences in frequencies between subgroups were evaluated using a Fisher's exact test followed by multiple test correction (MTC). MCRs were compared with the Database of Genomic Variants (http://projects.tcag.ca/variation/): regions showing an overlap above 80% between probes and known CNV were considered *bona fide CNV*.

In order to identify biological subgroups according to the genetic profile, unsupervised clustering was performed using the non-negative matrix factorization (NMF) algorithm on all 166 cases (Lee & Seung, 2001; Chigrinova *et al*, 2010). The best rank was chosen based upon the cophenetic correlation for the consensus matrix.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) 2008 web-accessible program (Huang da *et al*, 2009) was used to map genes contained within MCR, retrieved using the web-accessible Galaxy (http://galaxy.psu.edu/) (Giardine *et al*, 2005), to KEGG and Biocarta Pathways, and statistically examine the enrichment of gene members for each pathway.

Raw data will be available at the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) database.

Gene expression

Previously published gene expression data (Lenz *et al*, 2008b), obtained using the affymetrix genechip U133 plus 2.0 (Affymetrix, Santa Clara, CA, USA), were available for 54/166 cases. CEL files were imported in Partek Genomics Suite 6·4 (Partek, St Louis, MO, USA). Signal intensities were normalized by Partek RMA. Statistical differences were calculated by analysis of variance (anova) analysis and the genes with a *P* value <0·005 were analysed for enrichment of KEGG and Biocarta pathways using DAVID 2008 web-accessible program (Huang da *et al*, 2009).

To classify our samples according to the DLBCL consensus clusters described by Monti *et al* (2005), we trained a Naive Bayes classifier based on the 176 labelled samples from the original publication and 2117 probe-sets using the 'naiveBayes' function in the 'e1071' R-package. Then, this classifier was applied on our normalized GEP values to infer the best associated cluster: 'oxidative phosphorylation,' (OxPhos), 'B-cell receptor/proliferation,' (BCR) and 'host response' (HR) clusters.

Statistical analysis

Response to R-CHOP-21, overall survival (OS), PFS and disease-free survival (DFS), defined according to the criteria of Cheson *et al* (2007), were evaluated in the 124/166 cases treated with R-CHOP21 and with available follow-up data. Survival probabilities were plotted according to the Kaplan-Meier method. The impact of the MCR on the clinical outcome was determined with the log-rank test (Mantel & Haenszel, 1959). To control false positive rate the Benjamini-Hochberg-method (*Q*-value) was performed. Internal validation was quantified by calculating the Harrell concordance index (C-index) (Harrell, 2001; Kim *et al*, 2004; Hoster *et al*, 2008). A C-index of 0·5 indicates that outcomes are completely random, whereas a C-index of 1 indicates that the model is a perfect predictor. We used the bootstrap procedure as implemented in the 'validate' functions of the Design library in R (R Development Team 2009), which allowed for computation of an unbiased estimate of the C-index. We used 500 bootstrap samples. To identify MCR with an independent

impact on survival, we used a Cox proportional-hazard model with a stepwise selection of the significant variables. MCR occurring in at least 5% of the patients and with a significant impact in univariate analysis, and the single factors of the International Prognostic Index (IPI) for aggressive lymphomas were included in the model. A *P*-value below 0·05 was considered as statistically significant. Chi-square test was used to compare differences in clinical parameters among the different NMF subgroups previously identified. Statistical analyses on clinical data were performed with SPSS 17·0·1 (SPSS, Chicago, IL, USA) and R statistical package (R Development Team 2009).

Results

The frequencies of gains, losses and of LOH in 166 DLBCL samples are shown in Fig 1. The following MCRs were identified: 58 gains, 47 losses, 54 LOH, five recurrent amplifications and seven homozygous deletions (Table SI).

Amplifications occurred at 2p16·1 (7/166, 4%), 2p15 (3/166, 2%), 9p24·3–p21·3 (3/166, 2%), 13q32·1 (2/166, 1%) and 19q13·3 (2/166, 1%). Known oncogenes and lymphoma genes were present in 3/5 lesions: *BCL11A* and *REL* (2p16), *JAK2* (9p), *SPIB* (19q). The 2p15 included *C2orf86*, *MDH1*, *UGP2*. The 13q32·1 region contained *DNAJC3*, *HS6ST3*, *OXGR1*, *UGGT2* and mapped approximately 4 Mb telomeric to the transforming MIR17-92 cluster.

Homozygous deletions occurred at 3p14·2 (2/166, 1%), 6q23·3–q24·1 (2/166, 1%), 9p21·3 (5/166, 3%), 10q23·31 (2/166, 1%), 14q11·2 (3/166, 2%), 15q11·2 (2/166, 1%), and 19p13·3–p13·2 (2/166, 1%). Known tumour suppressor genes were included in 4/7 of these these MCRs: *FHIT* (3p14·2), *TNFAIP3* (*A20*) (6q23·3), *CDKN2A* (9p21·3) and *FAS* (10q23·31). The 14q and 15q might be considered *bona fide* CNV because they had a total overlap with known CNV, although the first lesion is located within the T-cell receptor alpha locus (*TRA@*), and could be due to infiltrating normal T-cells or a T-cell rearrangement that occurred in the lymphoma B-cell (Fuschiotti *et al*, 2007). The 15q deletion targeted *OR4N4* and *OR4M2*, G-protein-coupled receptors, members of the olfactory receptor gene family. The deletion of the 19p13 region has not been previously reported and comprised a series of genes, including *C3* (complement precursor 3), *TNFSF9* and *TNFSF14* (tumour necrosis factor ligand superfamily nine and 14).

Genomic regions and outcome in patients treated with R-CHOP-21

Table I shows the clinical features of 124/166 (75%) cases, treated with R-CHOP-21 and for whom follow-up data were available.

The rate of complete remissions (CR) was 81% (100/124), partial remissions (PR) 12% (15/124) and stable (SD) or progressive disease (PD) 7% (9/124). Among the complete responders, 27 (22%) eventually relapsed. After a median follow-up of 23 months (range, 1–123 months), the 3-year estimated OS was 88%, PFS 65% and DFS 71%.

Figure S2 shows the frequency of CN changes and of LOH for the 100 patients who obtained a CR after R-CHOP-21 and for the 24 who did not. The following lesions were more frequent in the non-CR patients: del(8p) (21% vs. 5%; $P = 0{\cdot}023$), del(8p21·2–p23·1) (21% vs. 6%; $P = 0{\cdot}036$), del(15q11·2) (8% vs. 0%; $P = 0{\cdot}036$) and LOH of 8p23·1

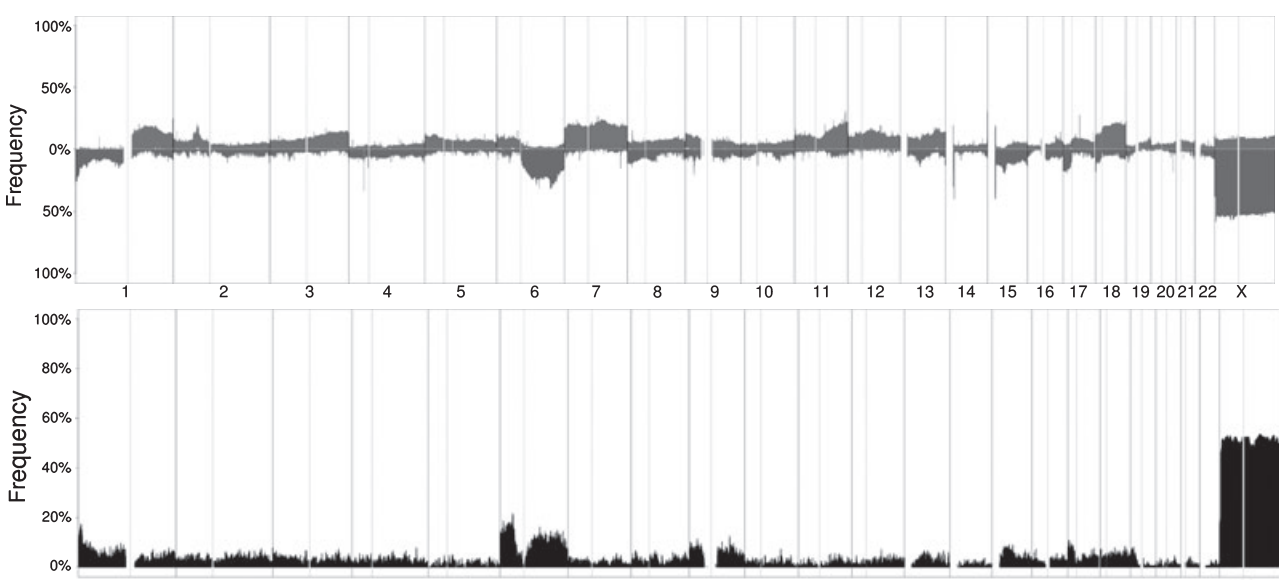

Fig 1. Frequency of DNA gains (red) and losses (blue) (upper panel) and LOH (lower panel), at observed in 166 DLBCL samples. *X*-axis, chromosome localization and physical mapping; *Y*-axis, percentage of cases showing the aberrations.

Table I. Clinical characteristics of 124 DLBCL patients treated with R-CHOP.

Clinical characteristics	*n*	Valid	Percent
Median age, years (range)		64 (18–86)	
Sex (m:f)	58:63	121	48:52
PS > 1	24	119	20
LDH > normal	62	110	57
Extranodal sites > 1	24	115	24
Stage III/IV	80	122	66
Bulky disease > 10 cm	24	109	22
IPI score > 1	77	103	75
B-Symptoms	40	119	34
Beta2-MG > normal	24	43	56
Germinal centre B-cell like*	48	78	55
HCV infection	10	76	13
Only extranodal disease	15	104	14
BM involvement	21	111	20

PS, performance status; LDH, lactate dehydrogenase IPI, International Prognostic Index; HCV, hepatitis C virus; BM, bone marrow.
*assessed by GEP or by immunohistochemistry. *assessed by GEP in 34 and by immunohistochemistry in 44 cases.

Table II. Twenty regions with a statistical significant impact on OS, PFS and/or DFS.

Region	OS *P*-value/ *Q*-value	PFS *P*-value/ *Q*-value	DFS *P*-value/ *Q*-value
1q42·2 LOH*	0·007/n.s.	<0·001/<0·001	0·034/n.s.
del(4p12–11)	n.s.	n.s.	0·028/n.s.
4q32·2 LOH*	n.s.	n.s.	0·045/n.s.
6q24·1–q24·2 LOH	n.s.	0·021/n.s.	n.s.
6q23·3–q24·1 hdel*	0·035/n.s.	n.s.	n.s.
6q22·32–q22·33 LOH	n.s.	n.s.	0·003/n.s.
6q25·1 LOH*	n.s.	n.s.	0·007/n.s.
8p23·1 LOH	0·015/n.s.	n.s.	n.s.
del(8p23·1)	0·002/0·1	n.s.	n.s.
del(8p23·1–21·2)	0·012/n.s.	n.s.	n.s.
del(8p)	0·01/n.s.	n.s.	n.s.
9p21·1 LOH	n.s.	0·047/n.s.	n.s.
del(9p24·2)	0·047/n.s.	n.s.	n.s.
9q31·1 LOH	0·04/n.s.	n.s.	n.s.
9p24·3–p21·3 ampl*	n.s.	n.s.	<0·001/0·001
10q23·31 hdel*	<0·001/<0·001	<0·001/<0·001	n.s.
14q11·2 hdel*	n.s.	n.s.	0·001/0·077
15q11·2 hdel*	<0·001/<0·001	<0·001/0·001	n.s.
del(15q12–21·2)	0·034/n.s.	n.s.	n.s.
del(18p11·32)	n.s.	0·027/n.s.	n.s.

*≤5 patients with the genomic alteration. hdel, homozygous deletion; ampl, amplification. n.s., >0·05 for *P*-value, >0·5 for *Q*-value.

(25% vs. 4%; $P = 0{\cdot}003$), 9p (12·5% vs. 2/100, 2%; $P = 0{\cdot}049$) and 15q21·3 (22% vs. 5%; $P = 0{\cdot}019$) (Table SI).

Twenty regions showed a statistically significant impact on OS, PFS and/or DFS (Table II).

Lesions with the strongest association with a worse outcome were deletions affecting the short arm of chromosome 8: del(8p23·1) ($P = 0{\cdot}002$, Fig 2), del(8p) ($P = 0{\cdot}01$), and del(8p23·1–21·2) ($P = 0{\cdot}012$). The prognostic relevance of these aberrations was underlined by an internal validation which calculated the C-index corrected for over-fitting by a bootstrap procedure. The C-index values were 0·64, and 0·584, 0·58.

Among the four homozygous deletions associated with a poor clinical outcome, two targeted known tumour suppressor genes: *TNFAIP3* at 6q23·3, and *FAS* at 10q23·31. Two, at 14q11·2 and 15q11·2, occurred at bona fide CNV: *TRA@* and *OR4N4*/*OR4M2* respectively). Although statistically significant, there were only a few cases bearing the four recurrent homozygous deletions, making it impossible to ascertain their real impact on outcome.

In order to identify possible pathways affected by these genomic regions, we mapped the genes contained within MCRs to known biological pathways. Seven pathways appeared to be significantly enriched: natural killer cell-mediated cytotoxicity, regulation of autophagy, antigen processing and presentation, Jak-STAT signalling pathway, Toll-like receptor signalling pathway, cytokine-cytokine receptor interaction and apoptosis (Table III).

In order to evaluate whether the identified MCRs were independent prognostic factors for OS, PFS and DFS, we performed a multivariate Cox regression. From the IPI factors, only lactate dehydrogenase (OS, $P = 0{\cdot}032$ and PFS, $P = 0{\cdot}002$) and stage III/IV (PFS, $P = 0{\cdot}017$) influenced survival independently in the multivariate analysis. None of the regions retained significance in the multivariate analysis.

Lesions associated with del(8p23·1)

We compared the pattern of genomic aberrations in patients with and without del(8p23·1) to elucidate the possible biological significance of genomic losses at this locus. Cases with del(8p23·1) presented several associated lesions (Table SII). The highest statistical significance was observed for losses occurring on the long arm of chromosome 15 (50% vs. 10%; $P = 0{\cdot}00004$), on 17p (41% vs. 10%; $P = 0{\cdot}0006$), on 10q23·33–q24·1 (32%, vs. 7%; $P = 0{\cdot}002$) and for 11q gains (41%, vs. 11%; $P = 0{\cdot}0009$).

Unsupervised clustering

In order to identify subgroups of patients with different genetic lesions and with possible distinct clinical outcome, we performed unsupervised clustering using the NMF algorithm. The latter represents a relatively new algorithm, showing advantages in the analysis of arrayCGH over other unsupervised approaches (Chigrinova *et al*, 2010). Five clusters were identified with distinct genetic profiles (Figs 3, and S3), without a significant bias on geographic provenance. Cluster 1 consisted of 77/166 patients (46%). Their genetic profile was heterogeneous and few recurrent aberrations could be

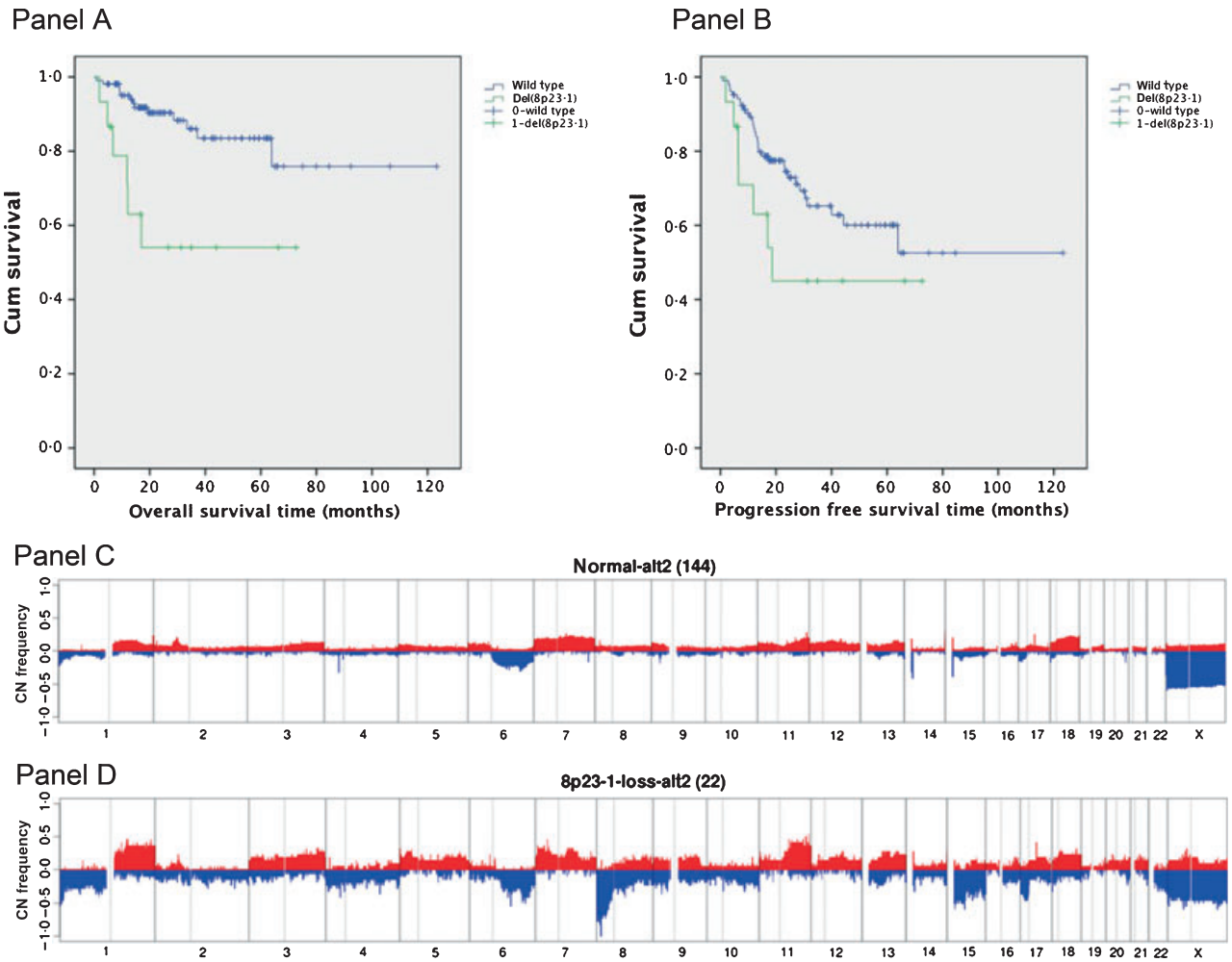

Fig 2. Del(8p23·1) in DLBCL: OS (P = 0·002; Panel A) and PFS (P = 0·119; Panel B) in patients without and with del(8p23·1); Frequency plots of DNA gains (red) and losses (blue) observed in patients with (22) and without (144) del(8p23·1) (Panels C and D).

observed. The second group (Cluster 2) was represented by 28/166 patients (17%) and characterized by gain of 1q and 3q, as well as loss of 1p, 6q and 17p. The third group included 8/166 patients (4%), with gains of 2p, 11p and 11q and losses of 1p, 6q, 9p, 15q, 17p and 18p. The fourth group consisted of 38/166 patients (23%). Their genetic profile showed gains of 1q, 2p, 6p, 7, 12q and losses of 6q, 8p, 15q and 17p. Finally, the fifth group was represented by 15/166 patients (9%). It was characterized by gains of 3q, 11q, 13q, 18q and losses of 1p, 6q, 8p, 9p and 15q.

Due to the distinct genomic profiles we compared Cluster 1 *versus* Clusters 2 and 4 and Cluster 2 *versus* Cluster 4. Cluster 1 showed a lower percentage of older patients (P = 0·068), a significantly higher rate of bone marrow (BM) involvement (16/57, 28% vs. 1/19, 5% and 2/24, 8%; P = 0·008) and hepatitis C virus (HCV) infection (9/38, 24% vs. 1/12, 8% and 0/20; P = 0·014). Clusters 2 and 4 showed differences in the number of patients with a high IPI-score (9/16, 56% vs. 7/26, 27%; P = 0·057), and relapse rate (5/17, 29% vs. 2/25, 8%; P = 0·068). We did not further analyse Clusters 3 and 5 (overall 23 patients).

Cluster 1 was not only characterized by a distinct genomic profile and different clinical features, such as high HCV infection and BM involvement, but, similar to Cluster 2, the outcome was relatively poor when compared to Cluster 4: high relapse rate (26%), 3-year estimated OS of 83%, 61% PFS and 66% DFS. Moreover, unlike Cluster 2 and 4, the PFS of Cluster 1 did not reach a plateau. The comparison of Cluster 2 and 4 showed a poorer DFS for Cluster two patients (P = 0·048). After 5 years both Clusters 2 and 4 reached a plateau in OS, PFS and DFS. In Cluster 2, the 3-year estimate of OS was 80%, PFS 62% and DFS 67%, whereas in Cluster 4, the 3-year estimate of OS was 92%, PFS 84% and DFS 83% (Fig S4).

The distribution of GCB or activated B-cell like (ABC) DLBCL, defined by GEP or by immunohistochemistry, did not present statistical differences among the five NMF clusters, although Cluster 4 consisted of mainly GCB (8/12, 67%), and Cluster 5 of ABC (7/10, 70%) (Fig S3). The distribution of the

© 2010 Blackwell Publishing Ltd,

Table III. Significantly enriched pathways among the MCRs with an impact on OS, PFS and DFS.

Pathway	Genes	*P*-value	Benjamini
Apoptosis	*CHP, TNFRSF10C, FAS, PPP3CC, TNFRSF10B, TNFRSF10A, TNFRSF10D,*	0·011	0·273
Natural killer cell mediated cytotoxicity	*CHP, IFNB1, IFNA6, IFNA8, SHC4, IFNA21, PPP3CC, TNFRSF10B, IFNA7, IFNA17, IFNA2, IFNA16, IFNA5, TNFRSF10C, IFNA4, FAS, IFNA1, IFNA10, TNFRSF10A, IFNA14, TNFRSF10D,*	$2{\cdot}98 \times 10^{-12}$	6×10^{-10}
Regulation of autophagy	*IFNA6, IFNA4, IFNA8, IFNA21, IFNA1, IFNA10, IFNA7, IFNA17, IFNA14, IFNA5, IFNA16, IFNA2*	$1{\cdot}05 \times 10^{-11}$	$1{\cdot}06 \times 10^{-9}$
Jak-STAT signalling pathway	*IFNA6, IFNB1, IFNA8, IFNA21, JAK2, IFNA17, IFNA7, IFNA2, IFNW1, IFNA16, IFNA5, SPRED1, IFNA4, IFNA1, IFNA10, IFNA14*	$1{\cdot}04 \times 10^{-6}$	$5{\cdot}24 \times 10^{-5}$
Toll-like receptor signalling pathway	*IFNA6, IFNB1, IFNA8, IFNA21, IFNA17, IFNA7, IFNA2, IFNA16, IFNA5, IFNA4, IFNA1, IFNA10, IFNA14*	$2{\cdot}17 \times 10^{-6}$	$8{\cdot}73 \times 10^{-5}$
Antigen processing and presentation	*IFNA6, IFNA8, IFNA21, PDIA3, CTSB, IFNA17, IFNA7, IFNA2, IFNA16, IFNA5, IFNA4, IFNA1, IFNA10, B2M, IFNA14*	$1{\cdot}48 \times 10^{-9}$	$9{\cdot}91 \times 10^{-8}$
Cytokine-cytokine receptor interaction	*IFNB1, IFNA6, IFNA8, IFNA21, TNFRSF10B, IFNA7, IFNA17, IFNA2, IFNW1, IFNA16, IFNA5, TNFRSF10C, IFNA4, FAS, IFNA1, IFNA10, TNFRSF10A, IFNA14, TNFRSF10D*	$1{\cdot}12 \times 10^{-5}$	$3{\cdot}76 \times 10^{-4}$

For each pathway, genes mapped in the MCRs, and the level of statistical enrichment (*P*-value and Benjamini-multiple test corrected) are given.

DLBCL-consensus clusters OxPhos, BCR, and HR presented a statistically significant distribution of HR cases, which were almost all in NMF Cluster 1 (12/16, 75%; $P = 0{\cdot}023$) (Fig S3).

A supervised analysis of the gene expression profile of Cluster 1 *versus* Clusters 2 and 4 identified 546 probes ($P < 0{\cdot}005$) (Table SIII). Genes of the T-cell receptor signalling pathway were the only significantly over-represented group of genes ($P = 2$ E−06; $Q = 4$ E−04). All 8/166 DLBCL cases described as T-cell rich were classified as Cluster 1.

Fig S5 shows the pattern of genomic lesions in 98 cases according to HCV-status. Although no statistical significance could be demonstrated, the 12 HCV+ DLBCL cases appeared to have less recurrent aberrations than HCV-negative cases, in accordance with that observed in NMF clustering.

Discussion

We conducted this retrospective multi-centre study to explore the impact of unbalanced genomic aberrations, detected using a high-density genome wide-SNP-based array, on OS, PFS and DFS in a population of DLBCL patients treated with the standard regimen R-CHOP-21. Deletions affecting the short arm of chromosome 8 had the most significant impact on survival. Five subgroups of patients with different aberrations and clinical features were identified by unsupervised clustering.

A series of 171 recurrent lesions were identified, largely reflecting what has already been reported in DLBCL (Tagawa *et al*, 2005a; Lenz *et al*, 2008a; Lenz & Staudt, 2010). The high resolution of the arrays allowed the detection of a discrete number of recurrent homozygous deletions, some targeting well known tumour suppressor genes, such as *CDKN2A* (9p21) or *FAS* (10q23·31), and the more recently reported *TNFAIP3* (Compagno *et al*, 2009). A homozygous deletion affected 19p13, possibly overlapping that reported in mantle cell lymphoma (Tagawa *et al*, 2005b). The lesion was still relatively large in size with several genes as possible targets, including *C3*, *TNFSF9 and TNFSF14*.

Twenty regions revealed a significant correlation with clinical outcome. In order to exclude a selection bias, we compared the observed clinical parameters to a previously published series of patients affected by DLBCL (Sehn *et al*, 2005), showing that they were comparable and thus reflected the general population of DLBCL patients. Loss of genomic material in the short arm of chromosome 8, in particular at 8p23·1, was associated with the most significant impact on OS, also confirmed by an internal validation by bootstrap. Moreover, the same region was more common among non responders. Del(8p) and loss of smaller regions of this chromosome arm were previously shown to be associated with poor characteristics and worse outcome in other lymphoid neoplasms (Martinez-Climent *et al*, 2001; Rubio-Moscardo *et al*, 2005a; Thelander *et al*, 2005; Forconi *et al*, 2008). Losses of 17p (*TP53, HIC1*) and 15q (*TP53BP1*) were observed at a statistically higher prevalence in association with loss of 8p, suggesting that the possible poor outcome after R-CHOP-21 could be due to the concomitant deregulation of a series of genes mapped on different chromosomes. Indeed, genes involved in DNA repair and apoptosis are mapped in these regions and could all be inactivated in patients bearing these deletions: *TNFRSF10A* (*TRAIL-R1*) and *TNFRSF10B* (*TRAIL-R2*) (8p), *TP53* and *HIC1* (17p) and *TP53BP1* (15q) (Rubio-Moscardo *et al*, 2005b; Young *et al*, 2007; Takeyama *et al*, 2008), which are enriched among patients with a poor CR-rate. Interestingly, in DLBCL, loss of *TP53* has been shown

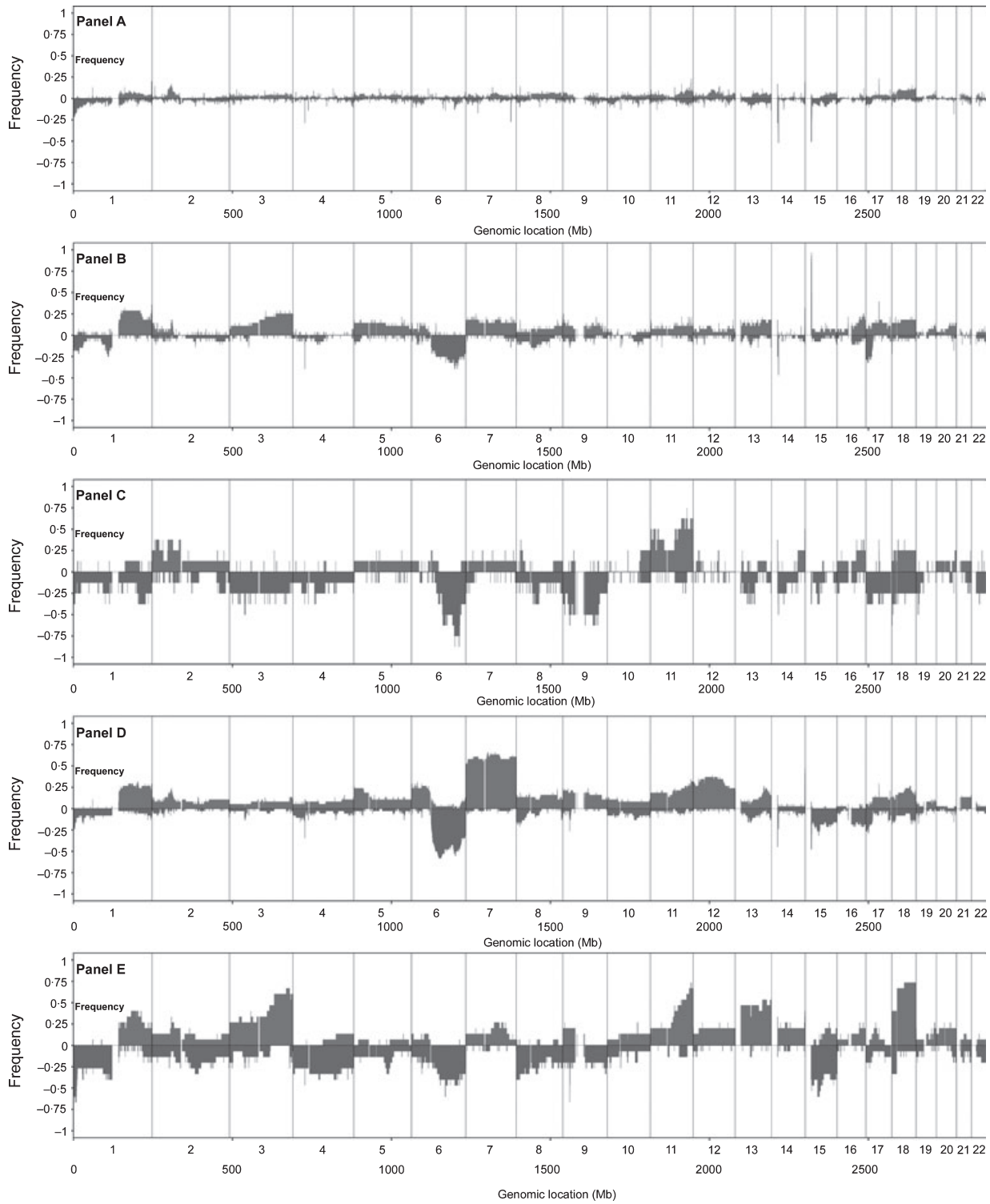

Fig 3. Frequency plot of 166 patients showing gains and losses according to the different NMF clusters. 3a: Cluster 1, 3b: Cluster 2, 3c: Cluster 3, 3d: Cluster 4, 3e: Cluster 5. *X*-axis, chromosome localization and physical mapping; *Y*-axis, percentage of cases showing the aberrations.

to correlate with low levels of TNFRSF10B (Young *et al*, 2007), and loss of 8p is a frequent phenomenon in patients with chronic lymphocytic leukaemia with 17p loss, possibly identifying a subgroup of patients with an extremely poor outcome (Forconi *et al*, 2008). Regarding *TP53*, in our series, loss of 17p13·1 did not show any impact on the clinical outcome,

similarly to that previously reported by Young *et al* (2007), who showed a negative influence of *TP53* mutations but not of the genomic loss on the clinical course in their series of DLBCL patients.

Del(9p21·3) had a negative impact on the response to treatment, albeit no apparent influence on survival was observed. These results are consistent with previous data reporting the unfavourable influence of this aberration on CHOP-treated DLBCL patients (Lenz *et al*, 2008a).

Functional annotation analysis and pathway mapping identified seven significantly enriched pathways among the genes of the regions with an impact on survival. Among the genes involved in these pathways were *TNFRST10A/TNFRSF10B*, *FAS* (natural killer mediated cytotoxicity and apoptosis) and *JAK2* (JAK-STAT pathway). Both, *TNFRST10A/TNFRSF10B* and *FAS* are genes coding for proteins with a crucial role in regulation of cell death and are known to be involved in mechanisms of tumour escape (Khong & Restifo, 2002). Therefore, loss of these genes in DLBCL strengthens the notion that both genes may play an important role in this disease. On the contrary, *JAK2*, a gene coding for a protein able to promote cell transformation by constitutive activation of the JAK-STAT pathway (Knoops *et al*, 2008), was amplified. The specific role of this aberration in DLBCL is yet unclear, whereas it has been shown that *JAK2* genomic gains are frequent in primary mediastinal large B-cell lymphoma and Hodgkin lymphoma, but they can occur also in DLBCL, more commonly among ABC-type (Lenz *et al*, 2008a).

The relatively small number of cases bearing the lesions and the median follow-up of only 2 years might partially explain the fact that none of the 20 genomic markers remained as an independent prognostic marker when analysed in a multivariate analysis alongside the IPI factors.

The unsupervised NMF-clustering identified five clusters with distinct genetic profiles, and some with different clinical characteristics and distinct courses.

Cluster one patients were characterized by a heterogeneous genomic profile, lacking most recurrent lesions and displaying the worst long term outcome amongst the different clusters. These patients had a high rate of HCV infection and BM involvement and a slightly younger age. GEP and morphology data suggested that this cluster could represent DLBCL cases with a high content of normal infiltrating T-cells, which would partially explain the low rate of recurrent genomic aberrations due to a lower content of neoplastic cells. These data could delineate a DLBCL subtype with distinct clinical features and a high rate of inflammatory cells. Indeed, when we applied the DLBCL-consensus cluster classifier (Monti *et al*, 2005) to our DLBCL cases with available GEP, there was a statistically significant enrichment of Cluster 1 for HR-DLBCL. The latter subgroup of DLBCL has been described as having increased expression of T-cell molecules, and, when compared to other DLBCL cases, HR patients were significantly younger, had a higher incidence of BM involvement and a lack of the recurrent chromosomal translocations involving BCL2 or BCL6 (Monti *et al*, 2005). Alternatively, we can speculate that, as we observed a high prevalence of HCV infection and because viruses can induce epigenetic transforming changes in human cells (Ferrari *et al*, 2009), at least part of Cluster 1 DLBCL, as well of the HR cluster, might have developed mainly through epigenetic changes rather than genomic aberrations. In support of this hypothesis, in our series, HCV-positive DLBCL cases showed a tendency for a lower rate of genomic lesions, but the number of cases was small and only studies on micro-dissected lymphoma cells will provide final results.

Similar to Cluster 1, the outcome of the Cluster two patients was relatively poor. This subgroup was characterized by 1q gain, 3q gain, 1p, 6q and 17p losses. Conversely, Cluster four patients, mainly characterized by gains of chromosome 7 and 12, presented a good outcome. Among the remaining two clusters, Cluster 3 was characterized by losses of 6q, 9, and 11q gains, and a trend for a good outcome, whilst Cluster 5 (gains of 3q, 11q and 18q; losses of 6q, 8p, 9p, 15q and 17p) for a poor outcome. The distribution of cases according to the cell-of-origin did not show statistical differences, probably due to the small number of cases with GEP, but, as expected based upon their genomic profiles, Clusters 4 and 5 presented more GCB and ABC cases, respectively. These data, together with our previous experience with NMF (Chigrinova *et al*, 2010), and the data reported using a model-based clustering (Shah *et al*, 2009), strongly suggest that the application of new unsupervised clustering algorithm on large DLBCL arrayCGH datasets will provide important insights for sub-classifying this heterogeneous disease.

In conclusion, genetic features associated with a different outcome were identified by genome-wide SNP-based array-CGH. When confirmed in other series, loss of 8p might highlight a group of patients with worse outcome when treated with R-CHOP-21 and characterized by a series of concomitant additional aberrations such as 17p and 15q losses. By applying unsupervised clustering, a distinct subgroup of patients characterized by a heterogeneous genomic profile, high prevalence of HCV infection, BM involvement and poor outcome with R-CHOP-21 could be observed. This subgroup deserves to be better characterized in order to design *ad hoc* therapeutic strategies.

Acknowledgements

Supported by Oncosuisse OCS-1939-8-2006; Swiss National Science Foundation (205321-112430, 205320-121886/1); Computational life science/Ticino in rete; Fondazione per la Ricerca e la Cura sui Linfomi (Lugano); PHS grant, UO1 CA 114778. M.S. is enrolled in the PhD program in Pharmaceutical Sciences, University of Geneva. M.M. is recipient of fellowship from Alto Adige Bolzano-AIL Onlus. E.C. is recipient of an European Society for Medical Oncology (ESMO) Fellowship Grant.

References

Armitage, J.O. (2007) How I treat patients with diffuse large B-cell lymphoma. *Blood*, **110,** 29–36.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R. & Li, C. (2006) Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS computational biology*, **2,** e41.

Capello, D., Scandurra, M., Poretti, G., Rancoita, P.M., Mian, M., Gloghini, A., Deambrogi, C., Martini, M., Rossi, D., Greiner, T.C., Chan, W.C., Ponzoni, M., Moreno, S.M., Piris, M.A., Canzonieri, V., Spina, M., Tirelli, U., Inghirami, G., Rinaldi, A., Zucca, E., Favera, R.D., Cavalli, F., Larocca, L.M., Kwee, I., Carbone, A., Gaidano, G. & Bertoni, F. (2010) Genome wide DNA-profiling of HIV-related B-cell lymphomas. *British Journal of Haematology*, **148,** 245–255.

Cheson, B.D., Pfistner, B., Juweid, M.E., Gascoyne, R.D., Specht, L., Horning, S.J., Coiffier, B., Fisher, R.I., Hagenbeek, A., Zucca, E., Rosen, S.T., Stroobants, S., Lister, T.A., Hoppe, R.T., Dreyling, M., Tobinai, K., Vose, J.M., Connors, J.M., Federico, M. & Diehl, V. (2007) Revised response criteria for malignant lymphoma. *Journal of Clinical Oncology*, **25,** 579–586.

Chigrinova, E., Kwee, I., Rinaldi, A., Poretti, G., Pruneri, G., Neri, A., Gaidano, G., Ponzoni, M., Zucca, E. & Bertoni, F. (2010) Non-negative matrix factorization to perform unsupervised clustering of genome wide DNA profiles in mature B cell lymphoid neoplasms. *British Journal of Haematology*, **150,** 229–232.

Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., Bhagat, G., Chadburn, A., Dalla Favera, R. & Pasqualucci, L. (2009) Mutations at multiple genes cause deregulation of the NF-kB pathway in diffuse large B-cell lymphoma. *Nature*, **459,** 717–721.

Ferrari, R., Berk, A.J. & Kurdistani, S.K. (2009) Viral manipulation of the host epigenome for oncogenic transformation. *Nature Reviews Genetics*, **10,** 290–294.

Forconi, F., Rinaldi, A., Kwee, I., Sozzi, E., Raspadori, D., Rancoita, P.M.V., Scandurra, M., Rossi, D., Deambrogi, C., Capello, D., Zucca, E., Marconi, D., Bomben, R., Gattei, V., Lauria, F., Gaidano, G. & Bertoni, F. (2008) Genome-wide DNA profiling identifies an unstable profile with recurrent imbalances predicting outcome in chronic lymphocytic leukemia with 17p deletion. *British Journal of Haematology*, **143,** 532–536.

Fuschiotti, P., Pasqual, N., Hierle, V., Borel, E., London, J., Marche, P.N. & Jouvin-Marche, E. (2007) Analysis of the TCR alpha-chain rearrangement profile in human T lymphocytes. *Molecular Immunology*, **44,** 3380–3388.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J. & Nekrutenko, A. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, **15,** 1451–1455.

Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Delabie, J., Ott, G., Muller-Hermelink, H.K., Campo, E., Braziel, R.M., Jaffe, E.S., Pan, Z., Farinha, P., Smith, L.M., Falini, B., Banham, A.H., Rosenwald, A., Staudt, L.M., Connors, J.M., Armitage, J.O. & Chan, W.C. (2004) Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*, **103,** 275–282.

Harrell, F.E. (2001) *Regression Modeling Strategies*. Springer-Verlag, Berlin.

Hoster, E., Dreyling, M., Klapper, W., Gisselbrecht, C., van Hoof, A., Kluin-Nelemans, H.C., Pfreundschuh, M., Reiser, M., Metzner, B., Einsele, H., Peter, N., Jung, W., Wormann, B., Ludwig, W.D., Duhrsen, U., Eimermacher, H., Wandt, H., Hasford, J., Hiddemann, W. & Unterhalt, M. (2008) A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood*, **111,** 558–565.

Huang da, W., Sherman, B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4,** 44–57.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J. & Thun, M.J. (2009) Cancer statistics, 2009. *CA Cancer Journal for Clinicians*, **59,** 225–249.

Khong, H.T. & Restifo, N.P. (2002) Natural selection of tumor variants in the generation of ''tumor escape'' phenotypes. *Nature Immunology*, **3,** 999–1005.

Kim, H.L., Seligson, D., Liu, X., Janzen, N., Bui, M.H., Yu, H., Shi, T., Figlin, R.A., Horvath, S. & Belldegrun, A.S. (2004) Using protein expressions to predict survival in clear cell renal carcinoma. *Clinical Cancer Research*, **10,** 5464–5471.

Knoops, L., Hornakova, T., Royer, Y., Constantinescu, S.N. & Renauld, J.C. (2008) JAK kinases overexpression promotes *in vitro* cell transformation. *Oncogene*, **27,** 1511–1519.

Lee, D.D. & Seung, S.H. (2001) Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, pp. 556–562. MIT Press, Cambridge, MA, USA.

Lenz, G. & Staudt, L.M. (2010) Aggressive lymphomas. *New England Journal of Medicine*, **362,** 1417–1429.

Lenz, G., Wright, G.W., Emre, N.C., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., Powell, J., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Rimsza, L.M., Fisher, R.I., Weisenburger, D.D., Chan, W.C. & Staudt, L.M. (2008a) Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **105,** 13520–13525.

Lenz, G., Wright, G., Dave, S.S., Xiao, W., Powell, J., Zhao, H., Xu, W., Tan, B., Goldschmidt, N., Iqbal, J., Vose, J., Bast, M., Fu, K., Weisenburger, D.D., Greiner, T.C., Armitage, J.O., Kyle, A., May, L., Gascoyne, R.D., Connors, J.M., Troen, G., Holte, H., Kvaloy, S., Dierickx, D., Verhoef, G., Delabie, J., Smeland, E.B., Jares, P., Martinez, A., Lopez-Guillermo, A., Montserrat, E., Campo, E., Braziel, R.M., Miller, T.P., Rimsza, L.M., Cook, J.R., Pohlman, B., Sweetenham, J., Tubbs, R.R., Fisher, R.I., Hartmann, E., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Wrench, D., Lister, T.A., Jaffe, E.S., Wilson, W.H., Chan, W.C. & Staudt, L.M. (2008b) Stromal gene signatures in large-B-cell lymphomas. *New England Journal of Medicine*, **359,** 2313–2323.

Mantel, N. & Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22,** 719–748.

Martinez-Climent, J.A., Vizcarra, E., Sanchez, D., Blesa, D., Marugan, I., Benet, I., Sole, F., Rubio-Moscardo, F., Terol, M.J., Climent, J., Sarsotti, E., Tormo, M., Andreu, E., Salido, M., Ruiz, M.A., Prosper, F., Siebert, R., Dyer, M.J. & Garcia-Conde, J. (2001) Loss of a novel

tumor suppressor gene locus at chromosome 8p is associated with leukemic mantle cell lymphoma. *Blood*, **98,** 3479–3482.

Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C., Dal Cin, P., Ladd, C., Pinkus, G.S., Salles, G., Harris, N.L., Dalla-Favera, R., Habermann, T.M., Aster, J.C., Golub, T.R. & Shipp, M.A. (2005) Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood*, **105,** 1851–1861.

R Development Team (2009) "*R: a Language and Environment for Statistical Computing*". R Foundation for Statiistical Computing, Vienna, Austria.

Rancoita, P.M.V., Hutter, M., Bertoni, F. & Kwee, I. (2009) Bayesian DNA copy number analysis. *BMC Bioinformatics*, **10,** 10.

Rinaldi, A., Capello, D., Scandurra, M., Greiner, T.C., Chan, W.C., Bhagat, G., Rossi, D., Morra, E., Paulli, M., Rambaldi, A., Rancoita, P.M.V., Inghirami, G., Ponzoni, M., Moreno, S.M., Piris, M.A., Mian, M., Chigrinova, E., Zucca, E., Dalla Favera, R., Gaidano, G., Kwee, I. & Bertoni, F. (2010) SNP-arrays provide new insights in the pathogenesis of post-transplant diffuse large B-cell lymphoma. *British Journal of Haematology*, **149,** 569–577.

Rubio-Moscardo, F., Climent, J., Siebert, R., Piris, M.A., Martin-Subero, J.I., Nielander, I., Garcia-Conde, J., Dyer, M.J., Terol, M.J., Pinkel, D. & Martinez-Climent, J.A. (2005a) Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood*, **105,** 4445–4454.

Rubio-Moscardo, F., Blesa, D., Mestre, C., Siebert, R., Balasas, T., Benito, A., Rosenwald, A., Climent, J., Martinez, J.I., Schilhabel, M., Karran, E.L., Gesk, S., Esteller, M., Deleeuw, R., Staudt, L.M., Fernandez-Luna, J.L., Pinkel, D., Dyer, M.J. & Martinez-Climent, J.A. (2005b) Characterization of 8p21.3 chromosomal deletions in B-cell lymphoma: TRAIL-R1 and TRAIL-R2 as candidate dosage-dependent tumor suppressor genes. *Blood*, **106,** 3214–3222.

Sehn, L.H., Donaldson, J., Chhanabhai, M., Fitzgerald, C., Gill, K., Klasa, R., MacPherson, N., O'Reilly, S., Spinelli, J.J., Sutherland, J., Wilson, K.S., Gascoyne, R.D. & Connors, J.M. (2005) Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia. *Journal of Clinical Oncology*, **23,** 5027–5033.

Shah, S.P., Cheung, Jr, K.J., Johnson, N.A., Alain, G., Gascoyne, R.D., Horsman, D.E., Ng, R.T. & Murphy, K.P. (2009) Model-based clustering of array CGH data. *Bioinformatics*, **25,** i30–i38.

Tagawa, H., Suguro, M., Tsuzuki, S., Matsuo, K., Karnan, S., Ohshima, K., Okamoto, M., Morishima, Y., Nakamura, S. & Seto, M. (2005a) Comparison of genome profiles for identification of distinct subgroups of diffuse large B-cell lymphoma. *Blood*, **106,** 1770–1777.

Tagawa, H., Karnan, S., Suzuki, R., Matsuo, K., Zhang, X., Ota, A., Morishima, Y., Nakamura, S. & Seto, M. (2005b) Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene*, **24,** 1348–1358.

Takeyama, K., Monti, S., Manis, J.P., Dal Cin, P., Getz, G., Beroukhim, R., Dutt, S., Aster, J.C., Alt, F.W., Golub, T.R. & Shipp, M.A. (2008) Integrative analysis reveals 53BP1 copy loss and decreased expression in a subset of human diffuse large B-cell lymphomas. *Oncogene*, **27,** 318–322.

Thelander, E.F., Walsh, S.H., Thorselius, M., Laurell, A., Landgren, O., Larsson, C., Rosenquist, R. & Lagercrantz, S. (2005) Mantle cell lymphomas with clonal immunoglobulin V(H)3-21 gene rearrangements exhibit fewer genomic imbalances than mantle cell lymphomas utilizing other immunoglobulin V(H) genes. *Modern Pathology*, **18,** 331–339.

Young, K.H., Weisenburger, D.D., Dave, B.J., Smith, L., Sanger, W., Iqbal, J., Campo, E., Delabie, J., Gascoyne, R.D., Ott, G., Rimsza, L., Muller-Hermelink, H.K., Jaffe, E.S., Rosenwald, A., Staudt, L.M., Chan, W.C. & Greiner, T.C. (2007) Mutations in the DNA-binding codons of TP53, which are associated with decreased expression of TRAILreceptor-2, predict for poor survival in diffuse large B-cell lymphoma. *Blood*, **110,** 4396–4405.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig S1. Examples of raw and inferred CN profiles. From top down, and indicated by red arrows: (i) del(8p) [with concomitant +8q, suggestive of i(8q)]; (ii), del(8p22-pter); (iii) del(8p23); (iv) del(6q) with homozygous deletion of *TNFAIP3*/*A20* locus; (v) del(10q23) with homozygous deletion of *FAS* locus.

Fig S2. Frequency of DNA gains (red) and losses (blue) observed in 124 patients according to complete remission after treatment.

Fig S3. NMF unsupervised clustering of 166 DLBCL cases.

Fig S4. Kaplan-Meier curves of OS (Panel A), PFS (Panel B) and DFS (Panel C) according to the different NMF clusters.

Fig S5. CN changes in 12 HCV-positive and in 78 HCV-negative DLBCL cases.

Table SI. Minimal Common Regions (MCR) identified in the whole population (166 cases).

Table SII. Fisher-test of gain and losses between patients carrying the 8p23.1 loss and not.

Table SIII. Differences in gene expression profile according to NMF Cluster 1.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.