Cross-Platform Assessment of Genomic Imbalance Confirms the Clinical Relevance of Genomic Complexity and Reveals Loci with Potential Pathogenic Roles in Diffuse Large B-Cell Lymphoma

Lizalynn M. Dias*,[1] Venkata Thodima*,[1] Julia Friedman,[1] Charles Ma,[1] Asha Guttapalli,[1] Geetu Mendiratta,[1] Imran N. Siddiqi,[2] Sergei Syrbu,[3] R. S. K. Chaganti,[4,5] Jane Houldsworth[1]

[1] Cancer Genetics, Inc., [2] Hematopathology Section, University of Southern California Keck School of Medicine, [3] Department of Pathology, Carver College of Medicine, University of Iowa, [4] Cell Biology Program, [5] Department of Medicine, Memorial Sloan-Kettering Cancer Center.

* Equal contribution

Running Title: Genomic imbalance large B-cell lymphoma

Keywords: Lymphoma, Molecular Genetics, Prognostication

Corresponding Author: Jane Houldsworth PhD, Cancer Genetics, Inc., 201 Route 17 North, Rutherford, NJ 07070. Email: jane.houldsworth@cgix.com. Phone: (201) 528-9171 FAX: (201) 528-9201

Overall word count: 3499

Abstract word count: 150

Number of figures: 3

Number of tables: 4

Number of references: 49

ABSTRACT

Genomic copy number alterations (CNAs) in diffuse large B-cell lymphoma (DLBCL) have roles in disease pathogenesis but overall clinical relevance remains unclear. Herein, an unbiased algorithm was uniformly applied across three genome profiling datasets comprising 392 newly-diagnosed DLBCL specimens that defined 32 overlapping CNAs, involving 36 minimal common regions (MCRs). Scoring criteria were established for 50 aberrations within the MCRs while considering peak gains/losses. Application of these criteria to independent datasets revealed novel candidate genes with coordinated expression, such as C*NOT2*, potentially with pathogenic roles. No one single aberration significantly associated with patient outcome across datasets, but genomic complexity, defined by imbalance in more than one MCR, significantly portended adverse outcome in two of three independent datasets. Thus, the standardized scoring of CNAs currently developed can be uniformly applied across platforms, affording robust validation of genomic imbalance and complexity in DLBCL and overall clinical utility as biomarkers of patient outcome.

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common aggressive non-Hodgkin's lymphoma (NHL) and despite the well-recognized clinical, pathologic, and genetic heterogeneity of DLBCL, almost all patients are treated with frontline rituximab-containing anthracycline-based chemotherapy regimen, generally cyclophosphamide, doxorubicin, vincristine and prednisone (RCHOP) or RCHOP-like [1,2]. Durable remission is achieved in approximately 60% of DLBCL patients where the highest likelihood of relapse is within the first two years [3-5]. Risk stratification of patients is primarily based on clinical features according to the revised International Prognostic Index (R-IPI) which, in general, is predictive of outcome [6]. Expression profiling studies of DLBCL specimens have revealed signatures with prognostic value, including cell-of-origin (germinal center B cell-like subtype [GCB] versus activated B cell-like subtype [ABC]) and functional pathway-related clusters [7-10]. In the former case, various immunohistochemistry-based algorithms have been developed and implemented in a clinical setting as surrogates for gene expression profiling but often with less than satisfactory concordance, thereby reducing the overall robustness of such prognostic biomarkers for clinical utility [11-13].

Various technologies have been employed to examine the DLBCL genome, leading in some cases to implementation in a routine clinical setting to guide DLBCL patient management. Such is the case for "double hit lymphomas", which bear genomic rearrangements at the *MYC* and *BCL2* or *BCL6* loci associated with overall poor outcome and routinely assessed by fluorescence in-situ hybridization [14,15]. More recent genomic profiling studies have involved the comprehensive assessment of structural variants such as genomic imbalance (gain and loss) mostly by array-comparative genomic hybridization (aCGH) and of somatic mutations by massively-parallel sequencing (MPS) [16-20]. In the former case, genomic copy number alterations (CNAs) in DLBCL have been described to have roles in disease pathogenesis and clinical relevance. For instance, gains of 7q, 8q (*MYC*), and 18q (*MALT* and *BCL2*), and loss of

8p, 9p (*CDKN2A*), and 17p (*TP53*) have variously been reported to be associated with shorter overall survival (OS) with others enriched in cell-of-origin subtypes [21-26]. In another study, genomic gains/losses involving the *CDKN2A-TP53-RB-E2F* axis reportedly defined DLBCL cases with overall higher genomic complexity and adverse outcome, independent of IPI [24]. Unfortunately, the clinical relevance of these various CNAs and potential for implementation in a clinical setting remain unclear due to few or no substantiating studies in independent cohorts, the use of different diagnostic platforms and analytical approaches across studies, and lack of uniformity of scoring specimens for CNAs.

Herein, a common analytical approach was applied across three independent publicly available genomic profiling datasets of almost 400 newly-diagnosed DLBCL to identify recurrent loci of genomic gain/loss and determine their prognostic value in clinically well-characterized DLBCL cohorts. Using this approach to delineate minimal common regions (MCR) and peaks of gain/loss across datasets, robust criteria were developed in order to score specimens as positive or negative for a specific aberration. Such scoring permitted correlative assessment of the clinical utility of individual CNAs as well as overall genomic complexity as biomarkers of outcome in a standardized manner that could be applied to other datasets utilizing different platforms for CNA detection.

MATERIALS AND METHODS

DLBCL aCGH/SNP Datasets and Specimens

Four aCGH or SNP public datasets of fresh frozen biopsy specimens of newly diagnosed DLBCL were available for download and had associated patient outcome data available for all or a subset of the patients: DS-172 comprising 172 CHOP-treated patients and for which matching expression data were available for 162 (GSE10846, U133 plus 2.0, Affymetrix Inc., Santa Clara, CA), DS-169 comprising cases from 99 CHOP- and 70 RCHOP-treated patients, DS-51HR comprising 51 high-IPI RCHOP-treated patients, and DS-124

comprising 124 RCHOP-treated patients. Platform and clinical characteristics of all datasets are provided in Supplementary Table S1.

For two additional *de novo* DLBCL patient cohorts, genomic DNA was extracted essentially following the protocol by EH van Beers *et al*. [27], from five 10 micron sections of diagnostic formalin-fixed paraffin-embedded (FFPE) biopsies, confirmed to exhibit more than 70% tumor burden or 2-3 tumor-enriched FFPE cores. DS-41 comprised biopsy material collected from 41 patients during routine care at the University of Iowa Hospitals and Clinics and DS-66 from 66 patients at the Los Angeles County/University of Southern California Medical Center. For both specimen cohorts, patients had received RCHOP or RCHOP-like regimens and all studies were performed with the respective IRB approval.

Genomic Identification of Significant Targets In Cancer (GISTIC) Analysis

For DS-172 and DS-51HR, duplicate probes were averaged, and the circular binary segmentation method was used to define segments ($P < 0.01$) within the "DNA copy" package in Bioconductor (R) (Version 2.10). Sites of known normal copy number variants derived from the Database of Genomic Variants (http://projects.tcag.ca/variation) were excluded from further analyses (Supplementary Methods). GISTIC (Version 0.9.2) was applied with a minimum acceptable segment of eight contiguous probes and log ratios ≥ 0.2 for gain and ≤ -0.2 for loss with an acceptable significant false discovery rate (FDR) q-value of 0.25. These log-ratio thresholds do not take into account variable tumor burden and heterogeneity across specimens. The published GISTIC data for DS-169 were used directly, where the same segmentation criteria had been applied [24].

Targeted aCGH and CNA Detection

For DNAs from DS-41 and DS-66, where the bulk of the DNA was greater than 800bp in size, heat fragmentation was performed prior to labeling at 95°C until the bulk DNA was 400-800bp. An equimixture of normal male and female DNA (Promega, Madison, WI) (MF) was similarly heat-fragmented to serve as reference DNA. Test and reference DNAs (1µg) were

differentially labeled using the CGH Labeling Kit for Oligo Arrays, (Enzo Lifesciences, Farmingdale, NY) and hybridized to a targeted oligonucleotide array representing genomic regions commonly altered in mature B-cell neoplasms (Agilent Technologies Inc., Santa Clara, CA) (Supplementary Table S2) [28]. Data were extracted using Feature Extraction Version 10.7.3.1 (Agilent).

For correlative analyses, CNAs were identified for all specimens using the Rank segmentation algorithm within the Nexus Copy Number Analysis Software (Version 6.1, Biodiscovery Inc., Hawthorne, CA). For all publicly available datasets, intra array replicates were combined and an aberration filter was applied of a minimum of eight continuous probes with an acceptable average value of $\log_2$ ratio change of ±0.2. For DS-41 and DS-66, an average value of $\log_2$ ratio change of ±0.3 was considered acceptable for a minimum of eight continuous probes. Raw data files for DS-41 and DS-66 have been deposited in GEO (GSEXXXX). All genomic coordinates are given according to the NCBI37/hg19 assembly.

Integrated Copy Number and Expression Analysis, and Pathway Enrichment

The raw expression data for 162 specimens of DS-172 were normalized using the Robust Multichip Average method and positively correlated expressed genes with respect to copy number were identified by the univariate t-test within the Bioconductor (R) package (Supplementary Methods). Those genes exhibiting at least a 1.2 fold change and $P \leq 0.05$ were considered significant and after FDR correction (Benjamini and Hochberg) were entered for pathway and functional enrichment using Ingenuity Pathway Analysis (Qiagen Inc., Gaithersburg, MD). Pathways with corrected *P*-values ≤ 0.05 were considered significant.

Clinical Correlative Analyses

Recurrent aberrations in RCHOP or RCHOP-like treated patients were tested for correlation with OS using the log-rank statistic and Kaplan-Meier methods. The Bonferoni method was used for FDR correction. Co-occurrence of aberrations in specimens was tested using the Pearson Chi-square test. A *P*-value of ≤ 0.05 was considered significant.

RESULTS

Recurrent Overlapping CNAs and Peaks in 392 Newly-Diagnosed DLBCL

In order to identify recurrent loci of genomic imbalance in newly diagnosed DLBCL specimens agnostic of the platform, three publicly available aCGH/SNP datasets (DS-172, DS-169, DS-51HR) were used. These were selected for discovery purposes based on their lack of suitability for evaluation of outcome association. Genomic regions of significant gain and loss were identified by application of GISTIC separately to DS-172 and DS-51HR and together with published GISTIC analysis of DS-169 were examined for overlapping significant regions of gain and loss [24]. Sixteen regions of genomic imbalance displayed overlap between all three datasets and another 16 in at least two (Figure 1A, Supplementary Table S3). The frequencies of each of the 32 overlapping CNAs varied across datasets (Figure 1B). DS-172 displayed overall lower frequencies and DS-51HR the highest, perhaps reflecting platform and dataset differences. Within the 32 overlapping region CNAs defined by the most distal and proximal boundaries of the overlapping GISTIC-designated CNAs, 36 minimal common regions (MCR) of genomic gain/loss were identified (Table 1, Figure 1C). Two MCR CNAs were found each for the gains mapped to 1p12-q44 and 12p11.21-q21.33, and losses to 2q22.2-q24.3 and 9p24.3-p21.1.

GISTIC also identified peak(s) of each significant gain/loss. Overlapping peaks were evident in at least two of the datasets for 18 of the 32 region CNAs, and included two each for the gain of chr7, and losses of 9p24.3-p21.1 and 15q12-q24.1. MCRs of the overlapping peaks (peak-MCRs) were determined and found to range in size from 2.9Kbp to 1.8Mbp (Table 2, Figure 1C). Seven peak-MCRs involved singleton genes, three of which were previously reported to have roles in DLBCL including *CD58*, *TNFAIP3*, and *CD70* [29-31]. Another peak-MCR at 9p24.1 targeted *JAK2* whose activation has been well documented in hematological

malignancies, predominantly by mutation [32]. The remaining three other peak-MCRs involving single genes included *HLA-DQA2* (A6p22.1-p12.1), *CNOT2* (A12p11.21-q21.33), and *CSGALNACT1* (D8p23.3-p11.23), not previously implicated in lymphoma biology. Of the remaining peak-MCRs, most had five or less mapped genes or miRs and included *REL*, *CDKN2A*, *FAS*, and *B2M* with known roles in DLBCL [21,29,33,34]. Peaks within the region CNAs were also detected by GISTIC that did not overlap across the datasets (data not shown) and included several mapped genes with previously reported roles in DLBCL: *BCL6*, *BCL2*, *TCF4*, *MUM1*, *PRDM1*, *PTEN*, *RB1*, *DLEU1*, *DLEU2*, and the *MIR16-1* locus [10,35-40]. These non-overlapping peaks were considered further when scoring for the respective CNAs.

Establishment of Criteria for Scoring Specimens

For subsequent expression and clinical correlative analyses, standardized criteria were established across all 392 specimens, to permit scoring for the presence of an aberration within MCRs that could be applied to any platform used to evaluate genomic imbalance. For the 32 MCR CNAs ≤ 10Mbp (Table 1), a specimen was scored positive if at least 50% of the MCR CNA exhibited the respective genomic gain/loss and if applicable, also included the overlapping peak-MCR (Table 2), with few exceptions (Figure 2). For 2p16.1-p15 gain, the requirement was reduced to 40% to more robustly accommodate the peak and for six ≤ 2Mbp, specimens with at least 90% of the MCR CNA altered (and if applicable, the peak-MCR), were scored positive to eliminate false positives on account of a known normal CNV. For the four remaining aberrations > 10Mbp, at least 3Mbp within the MCR CNA must have exhibited the gain/loss to score positive, and encompassed either a peak-MCR or a non-overlapping peak (Figure 2). This reduced requirement took into account the often multiple sites of involvement known for such large regions exemplified by 6q loss [41] . In total, the scoring criteria for 50 aberrations within the MCR CNAs were established (Supplementary Table S3, Figure 2). As a form of validation, aberrations detected in DS-172 were correlated with GCB and ABC cell-of-origin subtypes where the relative imbalance of six aberrations occurring at higher frequency in the GCB

subtype and five higher in ABC subtype were re-capitulated using the established criteria (Supplementary Table S4) [10].

These scoring criteria for the 50 aberrations were then applied to DS-124, DS-66, and a subset of 70 patients from DS-169 for correlative analyses described below. For one additional dataset (DS-41), minor modifications to the scoring criteria were required for 15 aberrations, to account for regions represented on the targeted array (Supplementary Table S3) but another eight were considered not scorable.

Integrated Genomic Imbalance, Expression, and Pathway Analysis

For 162 samples of DS-172, gene expression data were also available and permitted correlation between expression of probe sets in regions CNAs and the respective presence/absence of the 50 aberrations as scored according to the criteria established above. Significant correlated expression was found for 38 of the 50 aberrations representing 20 regions and 17 MCRs, comprising a total of 796 unique RefSeqs (723 genes) (Supplementary Table S5). Of these, 19 RefSeqs (17 genes) mapped to six peak-MCRs (Table 3). Notably, *CNOT2* (A12q14.2-q21.1) and *TNFAIP3* (D6q11.1-q27) were previously identified as singleton genes at peak-MCRs (Table 2). Ingenuity Pathway Analysis was performed on the significantly correlated 723 gene set which revealed significant enrichment in five canonical pathways (corrected *P*-value ≤ 0.05): p53 signalling ($P = 0.01$), PKCθ signalling in T-lymphocytes ($P = 0.025$), geranlygeranyldiphosphate biosynthesis ($P = 0.029$), B cell receptor signaling ($P = 0.039$), and RANK signaling in osteoclasts ($P = 0.039$).

Clinical Relevance of Genomic Imbalance

The clinical relevance of individual loci of genomic gain/loss in newly-diagnosed DLBCL was tested in datasets for which clinical outcome information was available after RCHOP or RCHOP-like immunochemotherapy: 70 patients from DS-169, DS-124, and two additional datasets (DS-41, DS-66) that utilized customized arrays. Based on the standardized presence/absence scores for each of the 50 aberrations, associations with outcome in each

dataset were tested using the log rank statistic. Across three datasets, ten aberrations significantly associated with shorter OS ($P \leq 0.05$) in at least one dataset, including two gains on 12q and two deletions each mapped to MCR CNAs involving 6q21 and 17p13 (Table 4). After FDR correction, gain of 19q13.33-q13.43 in DS-169 remained significant ($P < 0.001$) and the two losses at 6q21 and of 9p21.3 in DS-41 retained a trend ($0.05 < P < 0.1$). Apart from the gain of 19q13 and loss of 15q15-q21 which also showed a trend in an additional dataset, no aberration was significantly associated with adverse outcome in more than one dataset. Five of the clinically relevant aberrations detected in DS-169 did not significantly co-occur in specimens, apart from those mapped to the same chromosome arms. In DS-124, gain of 19q13 and loss of 8p21 significantly co-occurred, as did the losses of 6q21 and 9p21 in DS-41. In DS-66, with the shortest median follow-up amongst survivors (28 months), none of the recurrent 50 genomic aberrations significantly associated with outcome.

Assessment of Genomic Complexity and Clinical Relevance

It has been reported that DLBCL (DS-169) with genomic gain/loss affecting the *CDKN2A-TP53-RB-E2F* axis exhibit increased genomic complexity and shorter OS [24]. To verify this observation, for each of the nine CNAs marking a complex genome in that study, a comparable aberration or MCR CNA identified in the current study was assigned, and specimens scored “complex” if any one of the nine CNAs was evident. The specimens in DS-41 could not be scored due to the absence of few scorable aberrations. Across the 362 newly-diagnosed DLBCL specimens (DS-124, DS-172, and DS-66) that could be scored for genomic complexity in this manner, the average number of MCR and region CNAs in the complex group was 4.13 (±2.64) and 3.80 (±2.47) per specimen respectively (138 specimens), while it was 0.56 (±1.01) per specimen for both MCRs and region CNAs in the remainder 224 specimens. Thus, the previous observation of increased genomic instability in cases bearing genomic gain/loss affecting the *CDKN2A-TP53-RB-E2F* axis was confirmed. Association of genomic complexity with OS was tested in RCHOP and RCHOP-like treated patients of DS-169 (subset of 70

cases), DS-124, and DS-66 (Figure 3). The association between genomic complexity and shorter OS was recapitulated in DS-169 using the current comparable aberrations (Figure 3A) [24]. However, in the other two datasets, genomic complexity assessed in this manner did not significantly associate with reduced OS (Figure 3B and 3C).

As an alternative method for assessment of genomic complexity, the absolute number of CNAs involving the 36 MCRS was evaluated. Across all specimens (excluding DS-51HR and DS-41), the median count of aberrations in MCRs was one, so specimens with aberrations involving more than one MCR CNA were considered “complex”. Testing of association of genomic complexity scored in this manner with OS in RCHOP and RCHOP-like treated DLBCL patients revealed that patients with biopsy specimens with complex genomes in two of the three datasets evaluated, had significantly poorer outcome (Figure 3D and E). No association was found in the third dataset (DS-66) (Figure 3F, Supplementary Table S1). Overall, genomic complexity assessed by the absolute number of aberrations involving the 36 MCRS exhibited stronger association across datasets with adverse outcome than did aberrations affecting the *CDKN2A-TP53-RB-E2F* axis.

DISCUSSION

In this study, a cross-platform approach was undertaken to identify common regions of genomic gain and loss in almost 400 newly-diagnosed DLBCL patients in three independent aCGH/SNP datasets. Using this approach, 32 common regions of genomic imbalance were identified, comprising 36 MCRs, for which scoring criteria were standardized for a total of 50 aberrations taking into account loci of peak gains and losses. Application of these scoring criteria to clinically annotated RCHOP or RCHOP-like treated DLBCL patients with aCGH/SNP data allowed for robust inter-institutional assessment of the potential of these genomic aberrations and in general, overall genomic complexity, to serve as biomarkers of patient outcome. Such cross validation is an important consideration in establishing the overall clinical

utility of biomarkers, which was underscored in the current study where no one single site of genomic gain/loss was found to be consistently associated with patient outcome across all four independent datasets. Such is the case for loss of 9p21 (*CDKN2A*), previously observed to have prognostic value as a single abnormality in DLBCL but validated in only one dataset in the current study [10,21]. Inconsistency of clinical relevance of single abnormalities could partly be explained by differences in clinical feature distributions, such as IPI, across datasets. Indeed in DS-66 where no one specific aberration was associated with OS, A19q13.33-q13.43 and D9p21.3 were significantly enriched in patients with high IPI (data not shown), thereby supporting the overall association of these aberrations with adverse patient outcome [10,21,24]. Importantly, the scoring criteria developed in this study can be applied to other emerging datasets, such as within The Cancer Genome Atlas consortium to confirm these findings. With the promise of genomic imbalance detection from whole exome or genome MPS efforts, the same scoring criteria for genomic imbalance could also be similarly applied.

Application of GISTIC to the three discovery datasets yielded overlapping regional CNAs that, for the most part, have previously been reported in DLBCL [21-24]. Genes mapping to these common loci of genomic gain/loss have additionally been reported to harbor recurrent somatic mutations, such as *B2M*, *BCL6*, *CARD11*, *CD38*, *CD70*, *EZH2*, *FOXO1*, *PIM1*, *TNFAIP3*, *TNFRSF14*, and *TP53* [16-20]. Further examination of peaks of overlapping genomic imbalance in this study, not only identified several genes with previously reported roles in DLBCL, which confirmed the validity of the approach, but also few novel genes and miRNAs. For three aberrations, novel singleton genes were implicated: gains of *HLA-DQA2* (A6p22.1-p12.1) and *CNOT2* (A12p11.21-q21.33), and loss of *CSGALNACT1* (D8p23.3-p11.23). Interestingly, the expression of *CNOT2* also significantly correlated with the genomic status where gain was detected in up to 25% of DLBCL. CNOT2 is part of the CCR4-NOT deadenylase complex, a master regulator of RNA processing and degradation, where it heterodimerizes with CNOT3, and along with CNOT1 recruits the catalytic subunits of the

complex to target mRNAs [42]. Expression of *CNOT3* likewise exhibited significant positive correlation with gain of 19q13, frequently observed in DLBCL and most often in cases not overlapping with those exhibiting gain of 12q15 (data not shown). Follow-up studies are required to support functional roles of these genes in lymphomagenesis. Several other genes were also found in the current study whose altered expression coordinately correlated with genomic gain or loss that also mapped to peaks, including *TMEM30A* (CDC50A) that was coordinately de-regulated with deletion of 6q13 and recently reported to be recurrently mutated in DLBCL and follicular lymphoma [16,43]. The current cross-platform approach identified several novel genes with potential functional roles in the pathogenesis of this disease that, in combination with MPS data, will further focus efforts on the relevance of candidate genes [44].

It has been suggested in hematopoietic neoplasms, including DLBCL, that genomic complexity is associated with overall adverse outcome [24,28,45-48]. This has been reported not only in response to standard frontline therapies, but also recently for targeted therapies in these diseases [49]. In the current study, using the standardized scoring criteria it was found that those patients with specimens bearing genomic gain/loss in more than one of 36 involved MCRs indeed exhibited shorter OS following RCHOP or RCHOP-like therapy across datasets. Given the few reports of the mutation spectrum of this disease, it remains to be determined if mutation load also exhibits clinical relevance. Overall, the standardized scoring of genomic imbalance developed herein is well-suited for adoption in a clinical diagnostic setting and therefore will enable robust and reproducible validation of the clinical relevance of genomic complexity in DLBCL.

ACKNOWLEDGEMENTS AND DECLARATIONS OF INTERESTS

Conflicts of Interest: LD, VT, JF, CM, AG, GM, and JH were employees of, RSKC was a paid consultant of, and LD, VT, JF, CM, AG, GM, RSKC, and JH were stock/stock option holders of

Cancer Genetics, Inc (CGI). RSKC was on the Board of Directors of CGI. INS and SS had no conflicts of interest.

REFERENCES

1. Armitage JO. How I treat patients with diffuse large B-cell lymphoma. Blood 2007;110:29-36.
2. Kubuschok B, Held G, Pfreundschuh M. Management of Diffuse Large B-Cell Lymphoma (DLBCL). Cancer Treat Res 2015;165:271-288.
3. Coiffier B, Lepage E, Briere J, et al. . CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. The New England journal of medicine 2002;346:235-242.
4. Lee L, Crump M, Khor S, et al. . Impact of rituximab on treatment outcomes of patients with diffuse large b-cell lymphoma: a population-based analysis. Br J Haematol 2012;158:481-488.
5. Pfreundschuh M, Trumper L, Osterborg A, et al. . CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. The lancet oncology 2006;7:379-391.
6. Sehn LH, Berry B, Chhanabhai M, et al. . The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood 2007;109:1857-1861.
7. Monti S, Savage KJ, Kutok JL, 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.
8. Alizadeh AA, Eisen MB, Davis RE, et al. . Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-511.
9. Shipp MA, Ross KN, Tamayo P, et al. . Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nature medicine 2002;8:68-74.

10. Lenz G, Wright GW, Emre NC, et al. . 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 2008;105:13520-13525.
11. Barton S, Hawkes EA, Wotherspoon A, Cunningham D. Are we ready to stratify treatment for diffuse large B-cell lymphoma using molecular hallmarks? Oncologist 2012;17:1562-1573.
12. Coutinho R, Clear AJ, Owen A, et al. . Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for diffuse large B-cell lymphoma: implications for therapeutic strategies. Clin Cancer Res 2013;19:6686-6695.
13. Culpin RE, Sieniawski M, Angus B, et al. . Prognostic significance of immunohistochemistry-based markers and algorithms in immunochemotherapy-treated diffuse large B cell lymphoma patients. Histopathology 2013;63:788-801.
14. Swerdlow SH. Diagnosis of 'double hit' diffuse large B-cell lymphoma and B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and Burkitt lymphoma: when and how, FISH versus IHC. Hematology Am Soc Hematol Educ Program 2014;2014:90-99.
15. Green TM, Young KH, Visco C, et al. . Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2012;30:3460-3467.
16. Pasqualucci L, Trifonov V, Fabbri G, et al. . Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 2011;43:830-837.
17. Morin RD, Mungall K, Pleasance E, et al. . Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 2013;122:1256-1265.

18. Zhang J, Grubor V, Love CL, et al. . Genetic heterogeneity of diffuse large B-cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America 2013;110:1398-1403.

19. Lohr JG, Stojanov P, Lawrence MS, et al. . Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proceedings of the National Academy of Sciences of the United States of America 2012;109:3879-3884.

20. Morin RD, Mendez-Lago M, Mungall AJ, et al. . Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298-303.

21. Jardin F, Jais JP, Molina TJ, et al. . Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment: a GELA study. Blood 2010;116:1092-1104.

22. Scandurra M, Mian M, Greiner TC, et al. . Genomic lesions associated with a different clinical outcome in diffuse large B-Cell lymphoma treated with R-CHOP-21. British journal of haematology 2010;151:221-231.

23. Testoni M, Kwee I, Greiner TC, et al. . Gains of MYC locus and outcome in patients with diffuse large B-cell lymphoma treated with R-CHOP. British journal of haematology 2011;155:274-277.

24. Monti S, Chapuy B, Takeyama K, 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.

25. Chigrinova E, Mian M, Shen Y, et al. . Integrated profiling of diffuse large B-cell lymphoma with 7q gain. Br J Haematol 2011;153:499-503.

26. Iqbal J, Neppalli VT, Wright G, et al. . BCL2 expression is a prognostic marker for the activated B-cell-like type of diffuse large B-cell lymphoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2006;24:961-968.

27. van Beers EH, Joosse SA, Ligtenberg MJ, et al. . A multiplex PCR predictor for aCGH success of FFPE samples. Br J Cancer 2006;94:333-337.

28. Houldsworth J, Guttapalli A, Thodima V, et al. . Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia. Leuk Lymphoma 2014;55:920-928.

29. Challa-Malladi M, Lieu YK, Califano O, et al. . Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. Cancer cell 2011;20:728-740.

30. Bertrand P, Maingonnat C, Penther D, et al. . The costimulatory molecule CD70 is regulated by distinct molecular mechanisms and is associated with overall survival in diffuse large B-cell lymphoma. Genes Chromosomes Cancer 2013;52:764-774.

31. Honma K, Tsuzuki S, Nakagawa M, et al. . TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas. Blood 2009;114:2467-2475.

32. Vainchenker W, Constantinescu SN. JAK/STAT signaling in hematological malignancies. Oncogene 2013;32:2601-2613.

33. Bea S, Zettl A, Wright G, 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.

34. Takahashi H, Feuerhake F, Kutok JL, et al. . FAS death domain deletions and cellular FADD-like interleukin 1beta converting enzyme inhibitory protein (long) overexpression: alternative mechanisms for deregulating the extrinsic apoptotic pathway in diffuse large B-cell lymphoma subtypes. Clin Cancer Res 2006;12:3265-3271.

35. Mian M, Scandurra M, Chigrinova E, et al. . Clinical and molecular characterization of diffuse large B-cell lymphomas with 13q14.3 deletion. Ann Oncol 2012;23:729-735.

36. Pasqualucci L, Compagno M, Houldsworth J, et al. . Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. The Journal of experimental medicine 2006;203:311-317.

37. Dierlamm J, Murga Penas EM, Bentink S, et al. . Gain of chromosome region 18q21 including the MALT1 gene is associated with the activated B-cell-like gene expression subtype and increased BCL2 gene dosage and protein expression in diffuse large B-cell lymphoma. Haematologica 2008;93:688-696.

38. Akyurek N, Uner A, Benekli M, Barista I. Prognostic significance of MYC, BCL2, and BCL6 rearrangements in patients with diffuse large B-cell lymphoma treated with cyclophosphamide, doxorubicin, vincristine, and prednisone plus rituximab. Cancer 2012;118:4173-4183.

39. Care MA, Barrans S, Worrillow L, Jack A, Westhead DR, Tooze RM. A microarray platform-independent classification tool for cell of origin class allows comparative analysis of gene expression in diffuse large B-cell lymphoma. PLoS One 2013;8:e55895.

40. Colomo L, Lopez-Guillermo A, Perales M, et al. . Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B-cell lymphoma. Blood 2003;101:78-84.

41. Taborelli M, Tibiletti MG, Martin V, Pozzi B, Bertoni F, Capella C. Chromosome band 6q deletion pattern in malignant lymphomas. Cancer Genet Cytogenet 2006;165:106-113.

42. Boland A, Chen Y, Raisch T, et al. . Structure and assembly of the NOT module of the human CCR4-NOT complex. Nat Struct Mol Biol 2013;20:1289-1297.

43. Segawa K, Kurata S, Yanagihashi Y, Brummelkamp TR, Matsuda F, Nagata S. Caspase-mediated cleavage of phospholipid flippase for apoptotic phosphatidylserine exposure. Science 2014;344:1164-1168.

44. Yang Y, Staudt LM. Protein ubiquitination in lymphoid malignancies. Immunol Rev 2015;263:240-256.

45. Kay NE, Eckel-Passow JE, Braggio E, et al. . Progressive but previously untreated CLL patients with greater array CGH complexity exhibit a less durable response to chemoimmunotherapy. Cancer Genet Cytogenet 2010;203:161-168.

46. Delgado J, Salaverria I, Baumann T, et al. . Genomic complexity and IGHV mutational status are key predictors of outcome of chronic lymphocytic leukemia patients with TP53 disruption. Haematologica 2014;99:e231-234.

47. Klapper W, Kreuz M, Kohler CW, et al. . Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. Blood 2012;119:1882-1887.

48. Offit K, Wong G, Filippa DA, Tao Y, Chaganti RS. Cytogenetic analysis of 434 consecutively ascertained specimens of non-Hodgkin's lymphoma: clinical correlations. Blood 1991;77:1508-1515.

49. Knight SJ, Yau C, Clifford R, et al. . Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. Leukemia 2012;26:1564-1575.

TABLES

Table 1 36 MCRs of genomic gain and loss in 392 newly-diagnosed DLBCL specimens.

Genomic Gains			**Genomic Losses**		
MCR CNA	**Start (Mbp)**	**End (Mbp)**	**MCR CNA**	**Start (Mbp)**	**End (Mbp)**
A1q22.1-q25.1	145,390,170	174,990,852	D1p36.32-36.31	2,436,018	6,342,694
A1q31.3-q32.1	196,256,151	206,325,234	D1p13.1	116,368,309	117,144,417
A2p16.1-p15	57,798,324	64,087,027	D2q22.3	144,608,496	146,483,034
A3q27.3-q29	187,651,865	196,853,350	D2q24.2	161,836,066	161,963,707
A5p15.33	304,462	1,944,706	D3p21.31-p21.2	48,849,803	50,713,840
A6p21.32-p21.2	32,192,560	37,798,373	D3p14.2	60,408,342	60,690,266
A7p22.2	2,804,495	3,921,089	D6p21.33	30,919,878	31,536,224
A8q24.3	144,439,451	146,138,827	D6q11.1-q27	61,962,715	170,373,079
A9q34.3	138,543,735	140,878,804	D8p22-21.3	13,241,832	23,112,533
A11p15.4	3,225,556	3,808,818	D9p24.1	4,971,499	5,596,997
A11q23.3	120,599,996	120,820,140	D9p21.3	21,448,157	22,622,538
A12q13.11-q13.12	46,766,301	51,194,543	D10q23.31	90,567,045	90,986,391
A12q14.2-q21.1	64,762,188	73,252,451	D13q14.13-q14.3	46,917,541	53,172,272
A13q31.3	91,878,316	92,275,215	D15q15.1-q21.1	40,295,857	46,224,648
A16q24.3	89,644,837	89,662,874	D16q12.1-q12.2	50,674,936	52,984,751
A18p11.21-18q23	15,381,304	77,861,995	D17p13.3-p11.2	1,000,000	16,936,602
A19p13.3	260,999	2,183,174	D19p13.3	6,427,364	6,851,788
A19q13.33-q13.43	50,132,339	58,866,674			
A21q22.3	43,014,315	47,970,581			

Table 2 Minimal common regions (MCRs) of overlapping GISTIC-defined peaks and mapped genes/refseqs and miRs.

Region CNA	Overlapping Peak-MCRs			
	Start (Mbp)	**End (Mbp)**	**n**	**Candidate Genes/MiRs**
A2p16.3-p14	60,993,696	61,147,244	2	*REL*, *PAPOLG*
A3q22.1-q29	195,262,873	196,054,003	15	*TRFC* (*CD71*), *PCYTIA*, *MUC20*, *MUC4*, *MIR570*
A6p22.1-p12.3	32,690,831	32,709,466	1	*HLA-DQA2*
Achr7	1,017,099	1,081,180	3	*CYP2W1*, *C7orf50*, *MIR339*
	101,059,950	101,481,721	3	*COL26A1*, *MYL10*, *CUX1*
A9q34.11-q34.3	138,543,735	140,343,466	81	*CAD9*, *NOTCH1*, *EGLF7*, *MIR4673*, *MIR4674*
A12p11.21-q21.33	70,345,927	70,719,344	1	*CNOT2*
A13q31.3-q32.1	92,020,763	92,032,734	0	
A21q22.13-q22.3	43,959,706	44,358,680	5	*SLC37A1*, *PDE9A*, *WDR4*, *NDUVF3*, *HERV-Fb1*
D1p36.33-p36.22	2,926,556	3,762,888	12	*TP73*, *PRDM16*, *MIR551A*
D1p13.2-p12	117,070,416	117,111,959	1	*CD58*
D6p22.1-p21.32	31,167,499	31,326,959	3	*HCG27*, *HLA-B*, *HLA-C*
D6p12.3-q27	138,049,196	138,241,766	1	*TNFAIP3*
D8p23.3-p11.23	19,362,769	19,667,874	1	*CSGALNACT1*
D9p24.3-p21.1	5,106,681	5,109,576	1	*JAK2*
	21,978,347	21,984,369	2	*CDKN2A*, *MTAP*
D10q23.2-23.32	90,597,629	90,775,628	5	*FAS*, *FAS-AS1*, *ANKRD22*, *STAMBPL1*, *ACTA2*
D15q12-q24.1	43,133,828	43,370,177	2	*TTBK2*, *UBR1*
	44,939,261	45,026,122	4	*SPG11*, *PATL2*, *B2M*, *TRIM69*
D17p13.3-q11.1	7,841,724	7,920,934	2	*CNTROB*, *GUCY2D*
D19p13.3-p13.2	6,535,980	6,620,909	1	*CD70*

Table 3 Genes/refseqs mapped to peaks and exhibiting coordinately-regulated expression with genomic gain/loss.

MCR CNA	Band (Aberration No.)	Coordinately Expressed Genes/RefSeqs
A1q21.1-q25.1	1q23.3 (Ab3)	LOC100505773
A2p16.1-p15	2p16.1 (Ab5)	*PAPOLG*
A3q27.3-q29	3q29 (Ab6)	*SDHA/SDHAP1/SDHAP2*, SDHAP2, *MUC20*, *TFRC*, *PCYT1A* , LOC440993
A12q14.2-q21.1	12q15 (Ab15)	*CNOT2*
A18p11.21-q23	18q21.2 (Ab21)	*MEX3C*
	18q21.31 (Ab22)	*FECH*, *NARS*
D6q11.1-q27	6q13 (Ab35)	*TMEM30A*
	6q23.3 (Ab39)	*TNFAIP3*
D15q15.1-q21.1	15q15.2 (Ab45)	*UBR1*
	15q21.1 (Ab46)	*TRIM69*
D17p13.3-q11.2	17p13.1 (Ab48)	*POLR2A*, *SENP3*, *FXR2*

Table 4 Association of genomic gain/loss with overall survival in four independent datasets of DLBCL patients treated with RCHOP or RCHOP-like regimens.

MCR CNA (Aberration Number, Cytoband)	DS-169 (n=70)		DS-124		DS-41		DS-66	
	Freq	*P*-value	Freq	*P*-value	Freq	*P*-value	Freq	*P*-value
A12q13.11-q13.12 (Ab14, 12q13)	11.4%	0.042	3.2%	0.293	31.7%	0.511	21.2%	0.98
A12q14.2-q21.1 (Ab15, 12q15)	14.3%	0.024	4.8%	0.549	26.8%	0.508	22.7%	0.91
A19q13.33-q13.43 (Ab26, 19q13)	12.9%	<0.001	0.8%	ND	9.8%	0.066	9.1%	0.82
D6q11.1-q27 (Ab36, 6q21)	27.1%	0.709	13.7%	0.709	19.5%	0.005	33.3%	0.98
D6q11.1-q27 (Ab37, 6q21)	25.7%	0.83	14.5%	0.604	14.6%	0.002	30.3%	0.82
D8p22-p21.3 (Ab40, 8p21.3)	7.1%	0.437	5.6%	0.023	2.4%	ND	4.5%	0.48
D9p21.3 (Ab42, 9p21.3)	20.0%	0.447	3.2%	0.345	9.8%	0.008	13.6%	0.58
D15q15.1-q21.1 (Ab45, 15q15.2)	8.6%	0.098	8.9%	0.039	19.5%	0.128*	19.7%	0.37
D17p13.3-p11.2 (Ab48, 17p13.1)	10.0%	0.021	12.9%	0.344	24.4%	0.972	18.2%	0.2
D17p13.3-p11.2 (Ab49, 17p13.1)	10.0%	0.021	12.9%	0.344	24.4%	0.972	16.7%	0.13

Freq, Frequency; ND, not done

* Trending with longer OS

FIGURE LEGENDS

Figure 1 Overlapping significant loci of gain and deletion in three publicly available DLBCL datasets. A; Venn diagram showing the number of significant gains and deletions as defined by GISTIC in each of three datasets (DS-169, DS-172 and DS-51HR) altogether comprising 392 DLBCL specimens. A total of 32 significant gains/deletions were found to overlap in at least two of the three datasets. B; Frequencies of 32 overlapping significant gains/deletions in each of the three discovery datasets based on the respective GISTIC-derived CNA. The 16 gains/deletions common to all three datasets are underlined. A: gain, D: deletion/loss. C; Schematic of CNA regions (distal and proximal boundaries of overlapping GISTIC-defined regions of gain/loss)), MCRs (minimal common region of overlapping gain/loss), peaks (GISTIC-defined peaks of gain/loss, hatched box), and peak-MCRs (minimal common region of overlapping peak gain/loss, solid box).

Figure 2 Scoring criteria for 50 genomic aberrations. The criteria utilized to score each specimen for each aberration are shown, grouped according to the respective rules (Supplementary Table S3), with an example depicted. Overlapping peak-MCR (solid box), non-overlapping peak (hatched box).

Figure 3 Effect of genomic complexity on RCHOP and RCHOP-like treated DLCBL patient outcome. Kaplan-Meier plots for each of three datasets of RCHOP- and RCHOP-like DLBCL patients: DS-169 (70 patients) (A,D), DS-124 (B,E), and DS-66 (C,F). Cases were stratified for genomic complexity as evaluated by two methods. In the first (A-C), specimens were scored based on the presence of a CNA along the *CDKN2A-TP53-RB-E2F* axis (complex, Co) or absence (not complex, NCo). For the second (D-F), specimens bearing CNAs involving more than one MCR were scored as complex (Co) and one or none as noncomplex (NCo). The log-rank statistic was used to test for significance.