The potential of copy number gains and losses, detected by array-based comparative genomic hybridization, for computational differential diagnosis of B-cell lymphomas and genetic regions involved in lymphomagenesis

Ichiro Takeuchi,[1] Hiroyuki Tagawa,[2] Akira Tsujikawa,[3] Masao Nakagawa,[2] Miyuki Katayama-Suguro,[3,2] Ying Guo,[2,4] and Masao Seto[2,5]

[1]Department of Scientific and Engineering Simulation, Nagoya Institute of Technology, Nagoya, Japan; [2]Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan; [3]Division of Information Engineering, Graduate School of Engineering, Mie University, Tsu, Japan; [4]Department of Pathology, State Key Laboratory of Cancer Biology, Xijing Hospital, Fourth Military Medical University, Shaanxi, P.R.China; [5]Department of Cancer Genetics, Nagoya University Graduate School of Medicine at Aichi Cancer Center, Nagoya, Japan

Manuscript received February 28, 2008. Revised version arrived August 19, 2008. Manuscript accepted September 8, 2008.

Correspondence:
Ichiro Takeuchi, Ph.D.
Department of Scientific and Engineering Simulation, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan.
E-mail: takeuchi.ichiro@nitech.ac.jp

The online version of this article contains a supplementary appendix.

ABSTRACT

Background

The differentiation of biologically and clinically different malignant lymphoma diseases or subtypes is crucial because it leads to better prognostication and therapeutic decision-making. Attempts have been made at subtype classification for diagnosing lymphomas on the basis of gene-expression profiling. Although array-based comparative genomic hybridization (array CGH) has identified a characteristic genomic alteration pattern for each disease entity, it has not been clear whether each patient with certain genomic alterations can be classified by array CGH data.

Design and Methods

Data on copy number gains and losses for 46 diffuse large B-cell lymphomas and 29 mantle cell lymphomas were used. The gene expressions of the diffuse large B-cell lymphomas cases were profiled and hierarchical clustering revealed that 28 of them were of the activated B-cell type and 18 were of the germinal center-B-cell type. Using these data, we developed a computer algorithm to classify lymphoma diseases or subtypes on the basis of copy number gains and losses.

Results

The method correctly classified 88% of the diffuse large B-cell lymphomas and mantle cell lymphomas, and 83% of the activated B-cell and germinal center-B-cell subtypes. These results demonstrate that copy number gains and losses detected by array CGH can be used for classifying lymphomas into biologically and clinically distinct diseases or subtypes.

Conclusions

Our computer algorithm based on array CGH data successfully classified diffuse large B-cell lymphomas and mantle cell lymphomas and activated B-cell and germinal center-B-cell subtypes with high accuracy. An important finding is that the regions automatically identified by the computer algorithm were located in the critical regions that are likely to be involved in the development of lymphoma.

Key words: diffuse large B-cell lymphoma, mantle cell lymphoma, array CGH, genome profile, lymphoma classification.

Citation: Takeuchi I, Tagawa H, Tsujikawa A, Nakagawa M, Katayama-Suguro M, Guo Y, and Seto M. The potential of copy number gains and losses, detected by array-based comparative genomic hybridization, for computational differential diagnosis of B-cell lymphomas and genetic regions involved in lymphomagenesis. Haematologica 2009; 94:61-69. doi: 10.3324/haematol.12986

Introduction

Malignant lymphomas comprise various disease entities

Malignant lymphomas are usually diagnosed on the basis of results of pathological and immunochemical investigations. The presence of disease-specific translocations and clinical features are also important for the diagnosis of lymphomas.[1] We have identified genomic copy number alterations for several malignant lymphomas including diffuse large B-cell lymphoma (DLBCL),[2,3] mantle cell lymphoma (MCL),[4] and T/NK cell leukemia/lymphoma[5] by means of array-based comparative genomic hybridization (array CGH). We also discovered subtype-specific genomic alterations in DLBCL[6] and adult T-cell leukemia/lymphoma.[7]

Several recent studies have shown the power of gene-expression analysis for the classification of malignant lymphoma diseases and subtypes.[8-12] In these studies, computer algorithms were developed to select differentially expressed genes and use them to construct the classifier. In the study presented here, we examined whether genomic copy number gains and losses detected by array CGH could also be used for the classification of malignant lymphomas and developed a computer algorithm for this purpose. This algorithm is similar to the ones used in gene expression-based classification,[11] but slightly modified to deal with array CGH data. We applied the algorithm to the classification of 75 cases of malignant lymphoma into 46 cases of DLBCL and 29 of MCL, as well as to further classify the 46 DLBCL cases into 28 of the activated-B-cell (ABC) subtype and 18 of the germinal center-B-cell (GCB) subtype.[4,6]

MCL is a single disease entity characterized by the translocation of (11;14)(q13;q32) accompanied by over-expression of CCND1.[1] DLBCL is known to be the most common tumor and accounts for 40% of all malignant lymphomas.[1] Gene expression analysis of DLBCL has demonstrated that these lymphomas comprise distinct tumor subtypes such as the ABC and GCB subtypes.[8] ABC DLBCL is an aggressive lymphoma and the overall survival rate of patients with this subtype is inferior to that of patients with the GCB subtype.[8,9] We recently demonstrated that ABC and GCB DLBCL have distinct patterns of genomic alterations.[6] However, although we demonstrated that each disease entity has a characteristic pattern of genomic alterations, it was not clear whether the array CGH data could be used for classification because patients with the same disease entity vary from case to case. In the current study, we investigated whether genomic copy number gains and losses detected by array CGH could reliably distinguish different lymphoma diseases (DLBCL and MCL) as well as different subtypes (ABC and GCB). We hypothesized that an analysis of genomic copy number gains and losses would provide useful information for accurate and reproducible diagnosis of malignant lymphomas.

Design and Methods

Array comparative genomic hybridization and gene expression profiling

The array consisted of 2304 BAC and PAC clones (ACC versions 3.0 and 4.0), covering the whole human genome with a resolution of roughly 1.3 Mb.[2-7] The array CGH data on 46 DLBCL and 29 MCL cases used for the present bioinformatics study were published previously (*Online Supplementary Appendix*).[4,6] All of the samples showed at least some genome copy number changes, indicating that the tumor percentage of the sample was over 20%, as previously described.[13] Expression profiles of all the DLBCL cases had been previously examined using the microarray glass slide of an Agilent oligonucleotide array custom-made for the Cancer Institute of the Japanese Foundation for Cancer Research (Tokyo, Japan), on which a total of 21,619 genes were spotted (Agilent Technologies, Palo Alto, CA, USA).[6] The DLBCL cases were classified into 28 ABC DLBCL and 18 GCB DLBCL cases by means of a hierarchical clustering algorithm, as described previously (*raw data: Online Supplementary Data 3*).[6] This subtype classification was also confirmed using the method described by Wright *et al.*[11] (*Online Supplementary Figure S1*). In order to identify the normal variations for the $\log_2$-ratio signals, we performed 16 normal versus normal hybridizations in the same array. Clones in the sex chromosome and those with average $\log_2$-ratios deviated from 0 by ±3.0 standard deviations (SD) were excluded. Clones that were not shared by array glass versions 3 and 4 were also excluded. This procedure resulted in the exclusion of 270 clones; the remaining 2035 clones were used for the analysis. The threshold of the $\log_2$-ratio for copy number gains and losses was determined to yield a false discovery rate of 10%. The $\log_2$-ratio signals with a difference between the duplicated $\log_2$-ratios deviating from 0 by ±3.0 SD were defined as no-copy-number alterations These preprocessing procedures are described in detail in *Online Supplementary Appendix*. All the data used in the present analysis can be obtained from the supplementary information page *http://www-nkn.ics.nitech.ac.jp/~takeuchi/ACGH* with username: *guest* and password: *acghclassifier*.

Array comparative genomic hybridization-based classifier

We developed a fully automatic computer algorithm for the array CGH-based classification of lymphoma subtypes. This algorithm is similar to those employed in the classification of malignant lymphomas using gene expression profiles.[11] Linear predictor scores were computed for each case on the basis of copy number gains and losses detected by the array CGH. The scaling factors (coefficients) of the linear predictor scores were selected as the (signed) negative log of the *p* values obtained with Fisher's exact test. Only those clones with the most significant differences determined with Fisher's exact test were used to produce the linear predictor scores, with the optimal number of clones deter-

mined empirically (*see below*). The distribution of the linear predictor scores for each of the two disease entities (DLBCL and MCL) was approximated by using the normal distribution. The means and variances of these normal distributions were estimated from the linear predictor score calculated for the cases with each disease entity. For a new case, we estimated the likelihood of it belonging to one of the disease entities and then classified it by applying Bayes' rule. The formal description of the array CGH-based linear compound Bayes' classifier is provided in *Online Supplementary Appendix*.

Validation

Leave-one-out cross-validation (LOOCV) was used to estimate the performance of the classifier. As discussed in recent publications,[14-16] LOOCV can produce a more reliable measure of classification accuracy than validating the performance with an independent validation set. We also used LOOCV to determine the optimal number of clones used to form linear predictor scores. For this purpose, we used nested-LOOCV with the outer loop to estimate the classification accuracy and the inner loop to determine the optimal number of clones. We also performed classification analyses by dividing the cases into training (60%) and validation (40%) sets. The classifier was then constructed with the training set and tested with the validation set. Results of the classification performances were not significantly different (p=0.05) from those of the LOOCV analyses for both the DLBCL-MCL and the ABC-GCB classifications.

Each clone's significance of the differences in copy number alterations was evaluated using Fisher's exact test, the false discovery rate and the family-wise error rate. The last two measures take into account multiple comparisons. We performed 10,000 label permutations to compute the false discovery rate and the family-wise error rate. The validation strategy and the computations for the significance measures are explained in detail in *Online Supplementary Appendix*.

Results

Classification of diffuse large B-cell lymphomas and mantle cell lymphomas

The classification accuracy estimated by LOOCV was 88% (95%CI: 0.822-0.938). The probabilities of DLBCL and MCL assignment obtained from the classifier are plotted in Figure 1A, and the classification results are summarized in Table 1. Without the cut-off threshold, eight DLBCL cases were mis-classified as MCL and three MCL cases were mis-classified as DLBCL. With

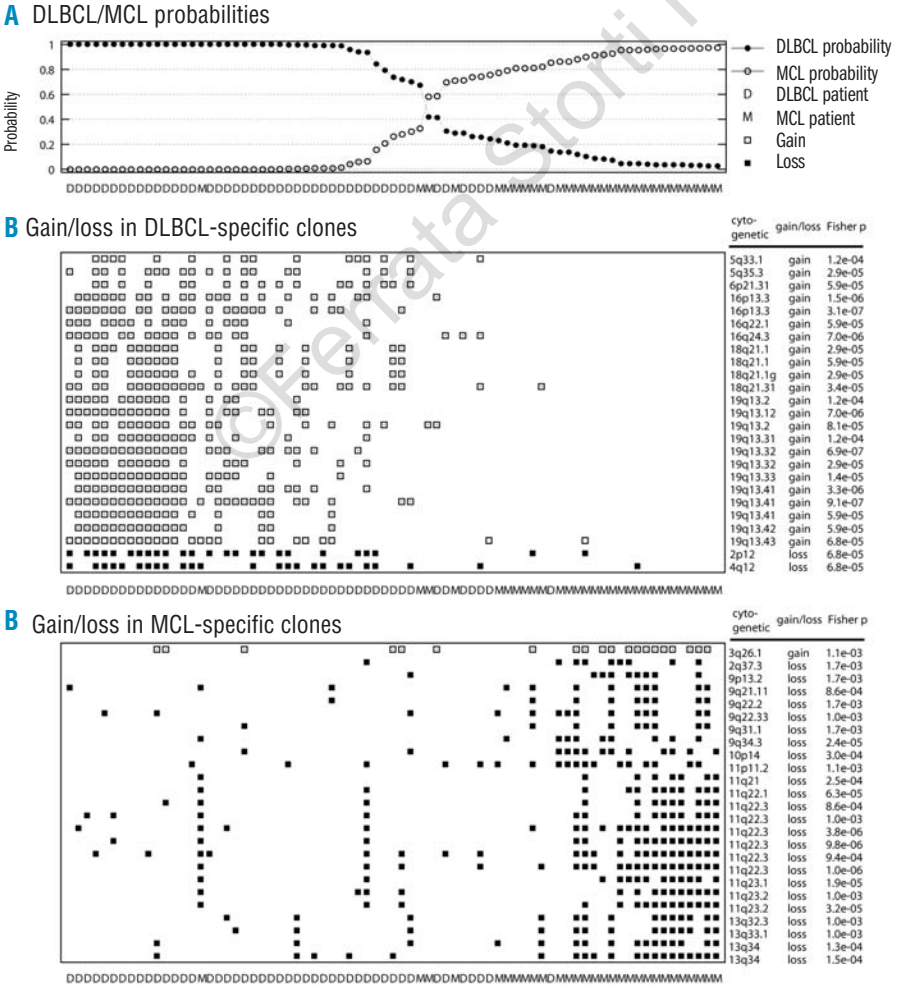

Figure 1. Performance of the array-CGH based classifier for the classification of diffuse large B-cell lymphomas and mantle cell lymphomas. (A) Probability of the 75 malignant lymphoma cases being diagnosed as diffuse large B-cell lymphoma or mantle cell lymphoma using the array CGH-based classifier. (B) Top 25 clones that showed gains or losses more frequently in diffuse large B-cell lymphoma than in mantle cell lymphoma. (C) Top 25 clones that showed gains or losses more frequently in mantle cell lymphoma than in diffuse large B-cell lymphoma.

the 80% cut-off level, only three cases were mis-diagnosed. For each of the LOOCV analyses, copy number gains and losses of an average of 49.7 clones (SD=13.5) were used for the classifier. We further tested the classifier's performance by dividing the cases into training and validation sets. As detailed in the Validation section, the classification accuracy was not significantly different (p=0.05) from that achieved using the LOOCV analyses.

Figure 1B shows the top 25 clones which showed gains and losses more frequently in DLBCL than in MCL, while Figure 1C shows the top 25 clones with the reverse difference in frequency. These differences in frequency were determined using the one-sided Fisher's exact test. Figures 1B and 1C also show gains and losses observed in 25×2 clones for all 75 patients. As can be seen from the detailed information on these 50 clones listed in Table 3, p values (from the one-sided Fisher's exact test) were below 1.7×10^{-3}, the false discovery rate was below 2.1×10^{-2}, and the family-wise error rate was below 7.1×10^{-1}. In the entire LOOCV analysis, only these 50 clones were selected for the classifications.

Table 1. Results of the classification of diffuse large B-cell lymphoma and mantle cell lymphoma using the array CGH-based classifier.

	Array CGH-based classifier				
	no cut-off[1]		cut-off=0.80[2]		
	DLBCL	MCL	DLBCL	MCL	unknown
DLBCL	39	7	35	1	10
MCL	2	27	1	23	5

The results obtained without using a cut-off threshold[1] and using a threshold of 80%[2] are shown.

Table 2. Results of the classification of activated-B-cell and germinal center-B-cell subtypes of diffuse large B-cell lymphoma using the array CGH-based classifier and a gene-expression based classifier.

	Array CGH-based classifier				
	no cut-off[1]		cut-off = 0.80[2]		
	ABC	GCB	ABC	GCB	unknown
ABC	27	1	21	1	6
GCB	7	11	5	11	2

The results obtained without using a cut-off threshold[1] and using a threshold of 80%[2] are shown.

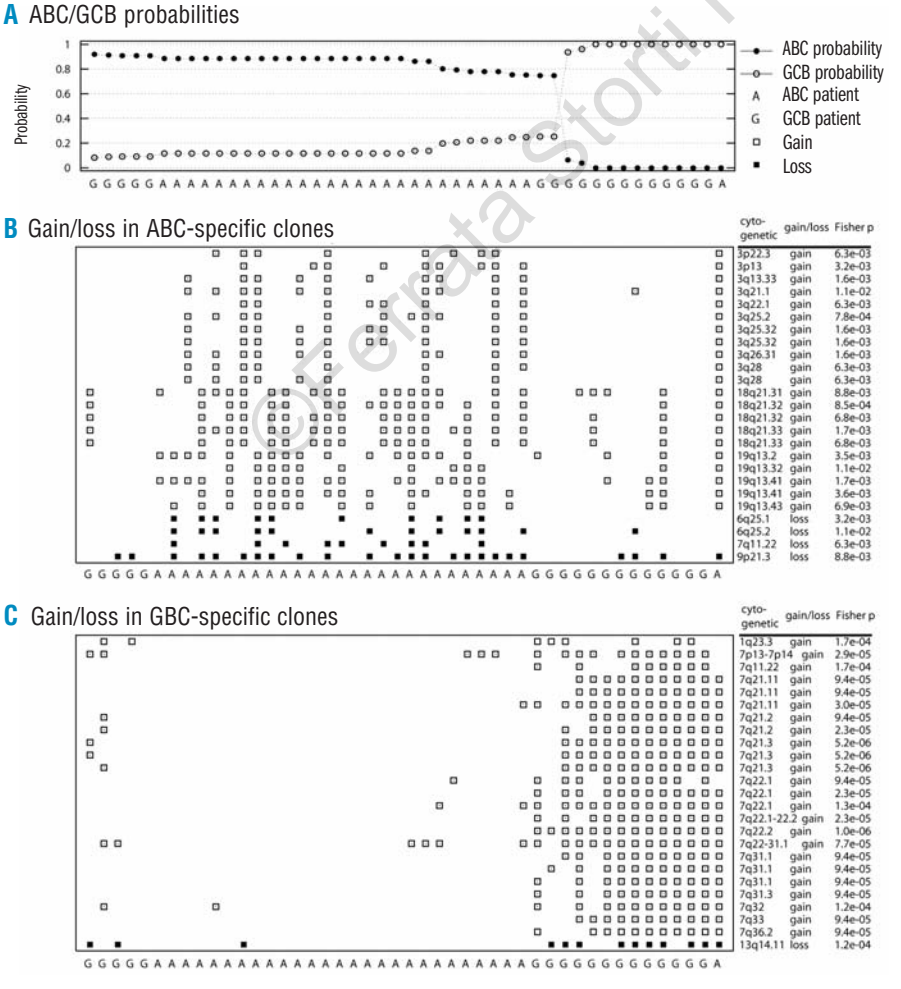

Figure 2. Performance of the array-CGH based classifier for the classification of activated-B-cell and germinal center-B-cell subtypes of diffuse large B-cell lymphoma. (A) Probability of the 46 malignant diffuse large B-cell lymphoma cases being diagnosed as activated-B-cell or germinal center-B-cell subtype using the array CGH-based classifier. (B) Top 25 clones which showed gains or losses more frequently in the activated-B-cell subtype than in the germinal center-B-cell subtype. (C) Top 25 clones which showed gains or losses more frequently in the germinal center-B-cell subtype than in the activated-B-cell subtype. Note that the subtypes activated-B-cell (represented as 'A') and germinal center-B-cell (represented as 'G') were defined using clustering analysis based on gene expression profiles.[6]

Activated-B-cell and germinal center-B-cell classification

The classification accuracy estimated by LOOCV was 82.6% (95%CI: 0.717-0.936). The probabilities of ABC and GCB assignment obtained from the classifier are plotted in Figure 2A, and the classification results are summarized in Table 2. One ABC case was mis-classified as GCB and seven GCB cases were mis-classified as ABC, and even with the 80% cut-off level, six cases were still mis-diagnosed. For each LOOCV analysis, copy number gains and losses of an average of 9.0 clones (SD=12.8) were used for the classifier. We further tested the classifier's performance by dividing the cases into training and validation sets as detailed previously in the section on Validation. The classification accuracy was slightly worse than that for the LOOCV analysis because of the small size of the sample used for constructing the classifier.

Figure 2B shows the top 25 clones which showed gains and losses more frequently in ABC than in GCB, while Figure 2C shows the top 25 clones with the reverse difference in frequency.

These differences in the frequency were determined using the one-sided Fisher's exact test. Figures 1B and 1C also show gains and losses observed in 25×2 clones for all 46 patients. As can be seen from the detailed information on these 50 clones listed in Table 4, p values (from the one-sided Fisher's exact test) were below 1.1×10^{-2}, and the false discovery rate was below 7.9×10^{-2}. In the entire LOOCV analysis, these 50 clones accounted for 92.5% of the clones used for the classifications.

Discussion

Genomic alterations including translocations and genomic copy number alterations are important events in lymphomagenesis. We previously showed that MCL, GCB DLBCL, and ABC DLBCL have characteristic genomic alteration patterns.[4,6] These findings led us to hypothesize that it might be possible to use array CGH

Table 3A. Clones with the most significant differences in copy number gains or losses between diffuse large B-cell lymphoma and mantle cell lymphoma. Top 25 clones with more gains and losses in diffuse large B-cell lymphoma than in mantle cell lymphoma (DLBCL-specific clones).

Clone ID	Clone name	Cytogenetic location	Genes	Fisher's p	FDR	FWER	Gain/loss
16-002	RP11-657D15	16p13.3	*E4F1, DSE1L2, DCI, RNPS1, ABCA3*	3.1E-07	2.0E-04	2.0E-04	Gain
19-033	RP11-384E6	19q13.32	*AKT2*	6.9E-07	2.5E-04	5.0E-04	Gain
19-042	RP11-10I11	19q13.41	*KLK7-14*	9.1E-07	1.7E-04	5.0E-04	Gain
16-001-01	RP11-31I10	16p13.3	*NME3, MRPS34, Q8TEP2, NM080861, NUBP2, IGFALS, HAGH*	1.5E-06	1.8E-04	8.0E-04	Gain
19-040	RP11-50I11	19q13.41	*BAX, CD37, BCL2L12*	3.3E-06	2.7E-04	1.5E-03	Gain
16-053-01	RP11-46C24	16q24.3		7.0E-06	4.8E-04	3.4E-03	Gain
19-027	RP11-38C1	19q13.12	*MLL4, U2AF1L3, PEN2, HSBX, NPHS1, KIRREL2*	7.0E-06	4.2E-04	3.4E-03	Gain
19-036	RP11-124P12	19q13.33		1.4E-05	8.5E-04	8.6E-03	Gain
05-119	RP11-265K23	5q35.3	*NSD1, RAB24, PXI9, MXD3, MAN2, RGS14*	2.9E-05	1.5E-03	1.7E-02	Gain
18-031	RP11-19L3	18q21.1		2.9E-05	1.4E-03	1.7E-02	Gain
18-034	RP11-43K24	18q21.1g		2.9E-05	1.3E-03	1.7E-02	Gain
19-034	RP11-317E13	19q13.32	*TGFB1*	2.9E-05	1.2E-03	1.7E-02	Gain
18-036-1	RP11-126O1	18q21.31	*MALT1*	3.4E-05	1.5E-03	2.4E-02	Gain
06-030	RP3-524E15	6p21.31	*BRPF3*	5.9E-05	2.3E-03	3.6E-02	Gain
16-040	RP11-529K1	16q22.1	*DDX19, SIAT4B, FUK*	5.9E-05	2.2E-03	3.6E-02	Gain
18-032	RP11-20H17	18q21.1		5.9E-05	2.1E-03	3.6E-02	Gain
19-043	RP11-256B9	19q13.41	*ZNF83, ZNF578, Q96LN7*	5.9E-05	2.0E-03	3.6E-02	Gain
19-044	RP11-158G19	19q13.42	*MYADM, PRKCG, CACNG6, 7, 8*	5.9E-05	1.9E-03	3.6E-02	Gain
19-046	RP11-45K21	19q13.43	*PEG3*	6.8E-05	2.5E-03	5.0E-02	Gain
02-061	RP11-113E4	2p12	*CTN2*	6.8E-05	2.4E-03	5.0E-02	Loss
04-047	RP11-586A2	4q12	*KIT*	6.8E-05	2.3E-03	5.0E-02	Loss
19-029	RP11-140E1	19q13.2		8.1E-05	3.0E-03	6.5E-02	Gain
05-102	RP11-54C4	5q33.1	*ATOX1, G3BP, GLRA1*	1.2E-04	3.7E-03	7.9E-02	Gain
19-025	RP11-32H17	19q13.2		1.2E-04	3.6E-03	7.9E-02	Gain
19-032	RP11-446K10	19q13.31	*PARK4, IL28*	1.2E-04	3.5E-03	7.9E-02	Gain

The table lists the clone ID, name, cytogenetic position and contained known genes. The differences in their copy number alteration properties were evaluated by Fisher's exact test. The p values and multiple comparison free measures, false discovery rate (FDR) and family-wise error rate (FWER) are shown.

data for the systematic diagnosis and classification of malignant lymphomas. Although each lymphoma entity has a characteristic genomic alteration pattern, patients with the same disease entity have heterogeneous genomic alteration patterns. It was, therefore, important to combine data on genomic alterations at several regions for accurate diagnosis. In this context, heterogeneity in DLBCL was speculated to be a problem, but we were able to develop a computer algorithm for the classification of lymphoma diseases and subtypes on the basis of copy number gains and losses detected by array CGH with high accuracy. Our array CGH-based classification algorithm is similar to one used in a previous study.[11] We slightly modified this algorithm to deal with array CGH data as shown in the *Online Supplementary Appendix*. Many other classification algorithms have been used for cancer classifications on the basis of gene-expression profiling.[17-19]

Several studies[8-12] have succeeded in demonstrating the power of gene expression profiling for the classification of lymphoma diseases and subtypes. In addition, genomic analysis has also been shown to be suitable for diagnostic purposes.[15,16] As demonstrated in our previous studies, smaller amounts of DNA can be used for analysis without amplification procedures.[2-7] Furthermore, greater stability and easier availability of DNA in comparison with RNA could be expected to make array CGH more reliable for diagnostic purposes. When we applied our method to the classification of different lymphoma entities (DLBCL and MCL) as well as different subtypes (ABC and GCB), the results showed that copy number gains and losses at a few dozen clones were effective for differentiating between disease entities as well as DLBCL subtypes. This study demonstrates that only a small subset of clones is required for a highly accurate classification.

The concordance between the ABC and GCB classification made by means of the hierarchical clustering method and classifier method described by Wright *et al.* was 91.3% (*Online Supplementary Figure S1*). The 83% accuracy achieved using array CGH data can, therefore, be assumed to be high. It remains to be determined which method of expression profiling classification is suitable for array CGH data classification.

The list of clones used for the classification of DLBCL and MCL diseases is provided in Table 3.The first 25 clones showed more frequent gains and losses in DLBCL than in MCL, and we designated them as *DLBCL-specific* clones. The other 25 clones showed more frequent gains and losses in MCL than in DLBCL,

Table 3B. Top 25 clones with more gains and losses in mantle cell lymphoma than in diffuse large B-cell lymphoma (MCL-specific clones).

Clone ID	Clone name	Cytogenetic location	Genes	Fisher's *p*	FDR	FWER	Gain / loss
11-068	RP11-758F15	11q22.3	*FDX1*	1.0E-06	1.8E-04	7.0E-04	Loss
11-065	RP11-144G7	11q22.3	*CUL5, ACAT, NPAT*	3.8E-06	2.7E-04	1.8E-03	Loss
11-066	RP11-241D13	11q22.3	*NPAT, ATM*	9.8E-06	6.7E-04	6.3E-03	Loss
11-073	RP11-307B17	11q23.1		1.9E-05	1.1E-03	1.1E-02	Loss
09-070	RP11-523A20	9q34.3	*TRAF2*	2.4E-05	1.3E-03	1.4E-02	Loss
11-072	RP11-667M19	11q23.2	*ZW10, USP28*	3.2E-05	1.3E-03	2.0E-02	Loss
11-060	RP11-864G5	11q22.1	*BIRC3, BIRC2*	6.3E-05	2.1E-03	4.0E-02	Loss
13-067	RP11-391H12	13q34	*CUL4A, LAMP, GRTP1*	1.3E-04	3.5E-03	8.9E-02	Loss
13-067-01	RP11-230F18	13q34	*C13orf17, C13orf11, TFDP1*	1.5E-04	4.3E-03	1.1E-01	Loss
11-055	RP11-775D16	11q21	*MAML2*	2.5E-04	5.8E-03	1.6E-01	Loss
10-008	RP11-401F24	10p14		3.0E-04	7.0E-03	2.1E-01	Loss
09-025	RP11-16N10	9q21.11	*TJP2*	8.6E-04	1.6E-02	5.1E-01	Loss
11-061	RP11-652L13	11q22.3		8.6E-04	1.6E-02	5.1E-01	Loss
11-067	RP11-402K1	11q22.3		9.4E-04	1.8E-02	5.5E-01	Loss
09-046-1	RP11-199C17	9q22.33	*TBC1D2, GPR51*	1.0E-03	1.8E-02	5.8E-01	Loss
11-065-1	RP11-95J9	11q22.3	*RAB39, CUL5*	1.0E-03	1.8E-02	5.8E-01	Loss
11-071	RP11-627G1	11q23.2	*TMPRSS5, ZW10*	1.0E-03	1.8E-02	5.8E-01	Loss
13-058-01	RP11-340C20	13q32.3	*PCCA*	1.0E-03	1.7E-02	5.8E-01	Loss
13-060	RP11-484I6	13q33.1	*Q8NDH2, Q8N800, NM138779, KDELC1, BIVM, ERCC5*	1.0E-03	1.7E-02	5.8E-01	Loss
03-126	RP11-85M11	3q26.1	*SLITRK3*	1.1E-03	1.9E-02	6.0E-01	Gain
11-029	RP11-193F22	11p11.2	*NM130783, TP53I11*	1.1E-03	1.9E-02	6.1E-01	Loss
02-167	RP11-546M8	2q37.3		1.7E-03	2.2E-02	7.1E-01	Loss
09-023	RP11-8N6	9p13.2	*MELK*	1.7E-03	2.2E-02	7.1E-01	Loss
09-041	RP11-563G12	9q22.2	*DIRAS2*	1.7E-03	2.1E-02	7.1E-01	Loss
09-047-2	RP11-467B11	9q31.1	*INVS*	1.7E-03	2.1E-02	7.1E-01	Loss

The table lists the clone ID, name, cytogenetic position and contained known genes. The differences in their copy number alteration properties were evaluated by Fisher's exact test. The p values and multiple comparison free measures, false discovery rate (FDR) and family-wise error rate (FWER) are shown.

Table 4A. Clones with the most significant differences in copy number gains or losses between activated-B-cell and germinal center-B-cell subtypes. A. Top 25 clones with more gains and losses in the activated-B-cell subtype than in the germinal center-B-cell subtype (ABC-specific clones).

Clone ID	Clone name	Cytogenetic location	Genes	Fisher's *p*	FDR	FWER	Gain/loss
03-111	RP11-554J1	3q25.2	*MBNL*	7.8E-04	1.4E-02	3.2E-01	Gain
18-036	RP11-4G8	18q21.32		8.5E-04	1.6E-02	3.6E-01	Gain
03-086	RP11-10G15	3q13.33	*Q9Y2K9, POLQ*	1.6E-03	2.2E-02	5.6E-01	Gain
03-117	RP11-113A11	3q25.32	*NM_016625*	1.6E-03	2.2E-02	5.6E-01	Gain
03-118	RP11-91L9	3q25.32		1.6E-03	2.1E-02	5.6E-01	Gain
03-135	RP11-816B4	3q26.31	*ECT2*	1.6E-03	2.1E-02	5.6E-01	Gain
18-038	RP11-299P2	18q21.33	*BCL2*	1.7E-03	2.4E-02	6.2E-01	Gain
19-039	RP11-699H21	19q13.41	*FUT2*	1.7E-03	2.4E-02	6.2E-01	Gain
03-055	RP11-56K23	3p13	*FOXP1*	3.2E-03	3.5E-02	7.7E-01	Gain
06-107	RP1-203A15	6q25.1	*LATS1*	3.2E-03	3.4E-02	7.7E-01	Loss
19-029	RP11-140E1	19q13.2		3.5E-03	4.1E-02	8.2E-01	Gain
19-043	RP11-256B9	19q13.41	*ZNF83, ZNF578, Q96LN7*	3.6E-03	4.6E-02	8.8E-01	Gain
03-028	RP11-129K12	3p22.3	*MLH1*	6.3E-03	6.4E-02	9.7E-01	Gain
03-094	RP11-65E22	3q22.1	*ACPP*	6.3E-03	6.4E-02	9.7E-01	Gain
03-157	RP11-67E18	3q28	*LPP*	6.3E-03	6.4E-02	9.7E-01	Gain
03-157-1	RP11-600G3	3q28	*TP73L*	6.3E-03	6.3E-02	9.7E-01	Gain
07-040	RP11-801B4	7q11.22	*AUTS2*	6.3E-03	6.3E-02	9.7E-01	Loss
18-037	RP11-40D15	18q21.32b-18q21.32c		6.8E-03	7.2E-02	9.8E-01	Gain
18-042-1	RP11-575O17	18q21.33	*SERPINB8,4,11,7*	6.8E-03	7.2E-02	9.8E-01	Gain
19-048	RP11-420P11	19q13.43		6.9E-03	7.2E-02	9.8E-01	Gain
18-036-1	RP11-126O1	18q21.31	*MALT1*	8.8E-03	7.4E-02	9.9E-01	Gain
09-014-1	RP11-149I2	9p21.3	*CDKN2A, B(P16-INK4), NSGX*	8.8E-03	7.4E-02	9.9E-01	Loss
03-087	RP11-67L2	3q21.1	*DIRC2, SEMA5B*	1.1E-02	8.0E-02	9.9E-01	Gain
19-035	RP11-46C6	19q13.32		1.1E-02	8.0E-02	9.9E-01	Gain
06-111	RP11-20H19	6q25.2		1.1E-02	7.9E-02	9.9E-01	Loss

The table lists the clone ID, name, cytogenetic position and contained known genes. The differences in their copy number alteration properties were evaluated by Fisher's exact test. The p values and multiple comparison free measures, false discovery rate (FDR) and family-wise error rate (FWER) are shown.

and are designated as *MCL-specific* clones. Among the top 25 MCL-specific clones, seven were in the 11q22 region, one of which was BAC RP11-241D13, which contains the *ATM* gene. It is known that the *ATM* gene is a tumor suppressor and that the inactivation of this gene does not activate DNA repair mechanisms properly.[20,21] Gene mutations and loss of heterogeneity have been identified in 56% of MCL.[21] However, neither loss of heterogeneity nor deletion of 11q22 was observed in DLBCL, according to a previous report.[6] The loss of 11q22 may, therefore, be strongly associated with the pathogenesis of MCL, while the presence or absence of this gene is also important for discriminating DLBCL and MCL.

The list of clones used for the classification of ABC and GCB subtypes is supplied in Table 4. The first 25 clones showed more frequent gains and losses in the ABC subtype than in the GCB subtype, and we designated them as *ABC-specific* clones. The other 25 clones showed more frequent gains and losses in the GCB subtype than in the ABC one, and we designated them as *GCB-specific* clones. The *BCL2* and *MALT1* genes were selected as ABC-specific clones. *MALT1* gene gain was previously suggested to play an important role in DLBCL.[22] Dierlamm *et al*. recently reported that the gain of 18q/*MALT1* is associated with the ABC subtype of DLBCL.[23] The fact that there are two ABC cases in the present study showing *MALT1* gains without any *BCL2* gain could indicate that *MALT1* may be the gene implicated in this region in the ABC subtype of DLBCL. Several clones at 3q25-qter were selected as ABC-specific in the present study. This is in accordance with the report by Bea *et al.,* who revealed that 65% of cases with 3q27 had 18q21-q22 gains among ABC subtype DLBCL.[24] These findings demonstrated that DLBCL subtyping by means of expression profiling is based on genomic alterations. The differential diagnosis of DLBCL and MCL by means of array CGH is less important because immunohistological markers for MCL, such as cyclin D1, already exist, although some cases of MCL can be misdiagnosed if the cyclin D1 does not stain clearly. More importantly, the clones selected with the algorithm used in our study are clearly associated with regions that are known to be characteristic to disease entities.

These include the 11q22 and 9q34.3 regions for MCL[21,25,26] and 18q21 and 19q13 for DLBCL.[6] Deletion of 9q34 has been reported to be a predictor of poor sur-

Table 4B. Top 25 clones with more gains and losses in the germinal center-B-cell subtype than in the activated-B-cell subtype (GCB-specific clones).

Clone ID	Clone name	Cytogenetic location	Genes	Fisher's *p*	FDR	FWER	Gain/loss
07-067	148A10	7q22.2	*LHFPL3*	1.0E-06	2.0E-04	2.0E-04	Gain
07-052	RP11-736C20	7q21.3		5.2E-06	6.5E-04	1.2E-03	Gain
07-053	648L18	7q21.3	*SGCE, PEG10*	5.2E-06	4.3E-04	1.2E-03	Gain
07-054	781H9	7q21.3	*DNCI1*	5.2E-06	3.3E-04	1.2E-03	Gain
07-051	RP5-911H5	7q21.2	*Q9P2G1*	2.3E-05	1.7E-03	7.3E-03	Gain
07-061	RP11-757A13	7q22.1	*CYP3A7, CYP3A4*	2.3E-05	1.4E-03	7.3E-03	Gain
07-066	RP11-451J3	7q22.1-q22.2	*ORC5L*	2.3E-05	1.2E-03	7.3E-03	Gain
07-024-1	RP11-141P12	7p13-7p14		2.9E-05	1.6E-03	9.4E-03	Gain
07-048	RP4-530J23	7q21.11	*GRM3*	3.0E-05	1.6E-03	1.1E-02	Gain
07-071	RP11-72J24	7q22-7q31.1		7.7E-05	3.8E-03	2.9E-02	Gain
07-045	RP11-448A3	7q21.11		9.4E-05	3.7E-03	3.0E-02	Gain
07-047	RP11-22M18	7q21.11		9.4E-05	3.4E-03	3.0E-02	Gain
07-050	RP5-1084H12	7q21.2		9.4E-05	3.1E-03	3.0E-02	Gain
07-059	RP4-808A1	7q22.1	*TRRAP, Q9P1H3, SUF1*	9.4E-05	2.9E-03	3.0E-02	Gain
07-069	RP11-77E2	7q31.1		9.4E-05	2.7E-03	3.0E-02	Gain
07-076	RP11-262F18	7q31.1	*PPP1R3A*	9.4E-05	2.5E-03	3.0E-02	Gain
07-077	RP11-78C11	7q31.1	*NM_199072*	9.4E-05	2.4E-03	3.0E-02	Gain
07-083-1	RP11-112P4	7q31.3		9.4E-05	2.2E-03	3.0E-02	Gain
07-091	RP11-233L24	7q33		9.4E-05	2.1E-03	3.0E-02	Gain
07-099	RP11-43L19	7q36.2		9.4E-05	2.0E-03	3.0E-02	Gain
07-086	RP11-35B6	7q32		1.2E-04	3.2E-03	4.4E-02	Gain
13-014	RP11-147D24	13q14.11	*NM017993*	1.2E-04	3.1E-03	4.4E-02	Loss
07-065	114K13	7q22.1	*PRO1598*	1.3E-04	3.6E-03	5.1E-02	Gain
01-096	RP11-331H2	1q23.3	*HSD17B7, NM_178550*	1.7E-04	4.9E-03	7.4E-02	Gain
07-037	RP11-535N23	7q11.22		1.7E-04	4.7E-03	7.4E-02	Gain

The table lists the clone ID, name, cytogenetic position and contained known genes. The differences in their copy number alteration properties were evaluated by Fisher's exact test. The p values and multiple comparison free measures, false discovery rate (FDR) and family-wise error rate (FWER) are shown.

vival in patients with MCL.[25,26] This seems to suggest that selected markers may play an important role in the pathogenesis and/or clinicopathological features of the various lymphoma entities. As some of the genetically altered areas have not yet been fully characterized at the molecular level, it is important to recognize that critical genes involved in disease development and progression still remain to be discovered.

Although it is important to identify such responsible genes, the identification of characteristic regions by means of a computer algorithm may be much more important than successful differential diagnosis based on array CGH data.

In summary, the results of our study show that genomic copy number gains and losses, detected by array CGH, can be used for the accurate diagnosis of different malignant lymphoma diseases and their subtypes. It was further demonstrated that copy number imbalances in only a few dozen clones differentiate different diseases and subtypes. Some clones used for the classification contained genes known to be strongly associated with tumor pathogenesis. This indicates that new target genes may be identified by using the classification procedure presented here.

Authorship and Disclosures

IT: designed and performed the data analysis and wrote the paper; HT: performed experiments on array CGH and wrote the paper; AT: contributed to application of the software for data analysis; MK-S: performed the gene-expression profiling experiments; YG: contributed to the pathological review and wrote the paper. MS: organized the research and wrote the paper. The authors reported no potential conflicts of interest.

References

1. Jaffe ES, Harris NL, Stein H, Vardiman JW. World Health Organization classification of tumors: pathology and genetics of tumors of hematopoietic and lymphoid tissues. IARC Press, Lyon, France, 2001.
2. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res 2004;64: 3087-95.
3. Tagawa H, Tsuzuki S, Suzuki R, Karnan S, Ota A, Kameoka Y, et al. Genome-wide array-based comparative genomic hybridization of diffuse large-B-cell lymphoma: comparison between CD5-positive and CD5-negative cases. Cancer Res 2004;64: 5948-55.
4. Tagawa H, Karnan S, Suzuki R, Matsuo K, Zhang X, Ota A, et al. Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. Oncogene 2005;24:1348-58.
5. Nakashima Y, Tagawa H, Suzuki R, Karnan S, Karube K, Ohshima K, et al. Genome-wide array-based comparative genomic hybridization of natural killer cell lymphoma/leukemia: different genomic alteration patterns of aggressive NK-cell leukemia and extranodal NK/T-cell lymphoma, nasal type. Genes Chromosomes Cancer 2005;44:247-55.
6. Tagawa H, Suguro M, Tsuzuki S, Matsuo K, Karnan S, Ohshima K, et al. Comparison of genome profiles for identification of distinct subgroups of diffuse large B-cell lymphoma. Blood 2005;106:1770-7.
7. Oshiro A, Tagawa H, Ohshima K, Karube K, Uike N, Tashiro Y, et al. Identification of subtype-specific genomic alterations in aggressive adult T-cell leukemia/lymphoma. Blood 2006;107:4500-7.
8. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.
9. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002;346:1937-47.
10. Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predics survival in mantle cell lymphoma. Cancer Cell 2003;3:185-97.
11. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B-cell lymphoma. Proc Natl Acad Sci USA 2003;100:9991-6.
12. Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma EJ, et al. Molecular diagnosis of Burkitt's lymphoma. N Engl J Med 2006;354:2431-42.
13. Fukuhara N, Nakamura T, Nakagawa M, Tagawa H, Takeuchi I, Yatabe Y, et al. Chromosomal imbalances are associated with outcome of Helicobacter pylori eradication in t(11;18)(q21;q21) negative gastric mucosa-associated lymphoid tissue lymphomas. Genes Chromosomes Cancer 2007;46:784-90.
14. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004; 4:309-14.
15. Wilhelm M, Veltman JA, Olshen AB, Jain AN, Moore DH, Presti JC Jr, et al. Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. Cancer Res 2002;62:957-60.
16. Veltman JA, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, et al. Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. Cancer Res 2003;63: 2872-80.
17. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-7.
18. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA 2001; 98:15149-54.
19. Dudoint S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc 2002;97:77-87.
20. Jares P, Colomer D, Campo E. Genetic and molecular pathogenesis of mantle cell lymphoma: perspectives for new targeted therapeutics. Nat Rev Cancer 2007;7:750-62.
21. Greiner TC, Dasgupta C, Ho VV, Weisenburger DD, Smith LM, Lynch JC, et al. Mutation and genomic deletion status of ataxia telangiectasia mutated (ATM) and p53 confer specific gene expression profiles in mantle cell lymphoma. Proc Natl Acad Sci USA 2006;103:2352-7.
22. Sanchez-Izquierdo D, Buchonnet G, Siebert R, Gascoyne RD, Climent J, Karran L, et al. MALT1 is deregulated by both chromosomal translocation and amplification in B-cell non-Hodgkin lymphoma. Blood 2003; 101:4539-46.
23. Dierlamm J, Murga Penas EM, Bentink S, Wessendorf S, Berger H, Hummel M, 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;93:688-96.
24. Bea S, Zettl A, Wright G, Salaverria I, Jehn P, Moreno V, et al. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. Blood 2005;106:3183-90.
25. Salaverria I, Zettl A, Beà S, Moreno V, Valls J, Hartmann E, et al. Specific secondary genetic alterations in mantle cell lymphoma provide prognostic information independent of the gene expression–based proliferation signature. J Clin Oncol 2007;25:1216-22.
26. Rubio-Moscardo F, Climent J, Siebert R, Piris MA, Martín-Subero JI, Nieländer I, et al. Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. Blood 2005; 105:4445-54.