New insights into the diagnosis of lymphomas

D. Soldini & E. Campo*

Hematopathology Section, Department of Anatomic Pathology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain

The current diagnosis of lymphoid neoplasias is based on the criteria of the World Health Organization (WHO) classification. This framework is built on two major principles: the stratification of neoplasms according to their derivation from precursor or mature cells and the definition of clinically relevant nonoverlapping diseases. The diagnosis is established by integrating the clinical, morphological, phenotypic, genetic and molecular characteristics of the tumors. This approach is reproducible, clinically relevant and scientifically sound. The elucidation of the human genome a decade ago and the development of high-throughput technologies have opened the possibility to search for comprehensive views of the genomic alterations of the tumors that are starting to influence our current approach to diagnosis. The new generation of sequencing technologies and their systematic application to human cancer and in particular to lymphoid neoplasms are revealing a landscape of somatic mutations of unprecedented complexity. These studies have already provided a number of important findings with functional and clinical implications. The translation of all this knowledge into the clinic is challenging and offers relevant perspectives.

Key words: gene expression profiling, lymphoma, microarray, next-generation sequencing

introduction

Lymphoid neoplasms are a heterogeneous group of malignancies for which a correct diagnosis requires the integration of morphology, immunophenotype, cytogenetic, molecular and clinical characteristics. The World Health Organization (WHO) classification represents an international agreement on this complex and rapidly evolving field to standardize the criteria necessary to establish more precise and reproducible diagnoses of distinct, mutually exclusive entities. In recent years, the development of very powerful technologies for genome analysis, including microarray platforms and next-generation sequencing (NGS), is revealing the genetic complexity of these tumors [1]. We are beginning to understand the potential implications of this information in the clinical practice including the diagnosis and discovery of new potential drug targets that may improve the therapeutic strategies for these patients in the near future. Hence, the purpose of this review is to present the current approach for the diagnosis of lymphoma and how it may be influenced by the new genomic perspectives.

diagnosis of lymphoid neoplasms: an integrative process

The diagnosis of lymphoid neoplasms requires the integration of different parameters of the tumor with the clinical manifestations of the patients. The specific value of different aspects may vary in different entities. Morphology represents the first step in the diagnostic process for most entities [2]. Nevertheless, as morphologic criteria entail a certain degree of subjectivity, the introduction of immunophenotyping studies, based on either immunohistochemistry (IHC) or flow cytometry (FCM) techniques, represents a major step toward achieving broader standardized criteria in the diagnosis. For cases, in which morphology, immunophenotype and clinical data might be inconclusive, genetic information such as the detection of translocations may be instrumental in establishing the correct diagnosis. The application of cytogenetic and molecular techniques has led to the discovery of genetic anomalies which have greatly advanced our understanding of the pathogenetic mechanisms of lymphomas and allowed for the refinement of their classification, and at the same time contributed to the improvement and standardization of routine diagnosis through the use of molecular diagnostics. Certain genetic anomalies are relatively specific of a given entity, such as the t(11;14) translocation leading to cyclinD1 overexpression in mantle cell lymphoma (MCL), whereas others, although characteristic of an entity, such as translocation of MYC in Burkitt lymphoma (BL) and of BCL2 in follicular lymphoma (FL), may occur in other diseases. The importance of genetic information is illustrated by the fact that in several entities, the presence of the translocation represents a key defining feature, which cannot be inferred on the sole basis of morphology or phenotype. For instance, in the case of anaplastic large cell lymphomas (ALCL), the ALK translocation status defines two clinically distinctive entities, ALK-positive and ALK-negative ALCL, the former being linked to younger age and better prognosis [3].

*Correspondence to: Dr E. Campo, Department of Anatomic Pathology, Hospital Clinic, Villarroel 170, 08013-Barcelona, Spain. Tel: +34-93-2275450; Fax: +34-93-2275572; E-mail: ecampo@clinic.ub.es

© The Author 2012. Published by Oxford University Press on behalf of the European Society for Medical Oncology. All rights reserved. For permissions, please email: journals.permissions@oup.com.
The routine use of this integrative approach in the study of lymphoid neoplasms has expanded our understanding of these tumors with the discovery of new clinical and pathological variants. In recent years, these studies have increased the detection in asymptomatic individuals of small populations of clonal lymphoid proliferations with the genetic and phenotypic features of certain neoplasias, such as monoclonal B-cell lymphocytosis, in situ FL or in situ MCL, which may correspond to early steps in the development of these tumors [4–6]. The identification of these lesions should not be interpreted as overt lymphomas and should be managed with caution.

In addition to morphology, phenotype and molecular genetics, clinical data are of major diagnostic and prognostic value in lymphoma diagnosis. Important clinical information for the diagnosis of certain entities is represented by age, immune status of the patient and topographic localization of the disease. Several entities in the current 2008 WHO classification include age as a defining feature, resulting in separate variants (e.g. pediatric or elderly) that occur mainly or exclusively in certain age groups and differ clinically and biologically from their counterparts in other age ranges. Examples of specific entities in children include pediatric FL and nodal marginal zone lymphoma (MZL), localized diseases with distinctive morphological features and very good prognosis [7, 8]. On the other hand, Epstein-Barr virus (EBV) positive diffuse large B-cell lymphoma (DLBCL) of the elderly was defined as an EBV-driven lymphoma occurring in patients >50 years, possibly related to the deterioration of the immune system associated with aging.

Information on the immunological status of the patient and viral infection status is often another prerequisite for a correct diagnosis. Several lymphoma entities have been associated with an immune-compromised status due to human immunodeficiency virus infection, post-transplant or immunosuppressive treatments. Lymphomas associated with immunosuppression and viral infections are frequently highly aggressive, show particular pathological features such as terminal B-cell differentiation and are commonly infected by EBV or HHV-8. Other lymphoma entities are almost invariably associated with EBV viral infection in the tumor T- or B-cell populations in immunocompetent patients, such as certain T-cell lymphoproliferative disorders of childhood, extranodal NK/T-cell lymphoma, nasal type (ENK/TCL, NT) and aggressive NK-cell leukemia or the infection by the retrovirus human T-cell leukemia virus type I in adult T-cell leukemia/lymphoma.

The importance of the topographic site of the tumor as a major diagnostic criterion in certain lymphomas is illustrated by the example of different DLBCL subtypes with a predominantly extra-nodal presentation. Site-specific extranodal DLBCL subtypes include primary mediastinal (thymic) large B-cell lymphoma, primary cutaneous DLBCL leg type and primary DLBCL of the central nervous system.

genomic studies in lymphoma diagnosis

The initial sequencing of the whole human genome more than a decade ago and, more recently, the introduction of very powerful technological developments in the field of genomic analysis have opened up unprecedented opportunities in the search for genomic alterations underlying cancer development and progression [9, 10]. Several high-throughput platforms have been designed for the study of either RNA gene expression profiling (GEP) or DNA changes, including chromosomal imbalances, single-nucleotide polymorphisms and epigenetic modifications.

microarrays for DNA studies

The use of DNA microarrays for the detection of gains or losses in specific chromosomal regions has contributed to the identification of tumor genomic profiles relatively characteristic of different diseases [11]. The number and distribution of genomic alterations in a given entity vary from patient to patient and this heterogeneity may account in part for the variable behavior of the tumor, with potential implications in terms of prognosis and response to therapy [12–14].

In conjunction with other techniques, the study of minimal chromosomal regions identified as recurrently deleted or amplified across different entities has proven useful in the discovery of putative oncogenes or tumor suppressor genes located in these regions, which may in some cases harbor several ‘driver’ genes with a potential cooperative effect in the biology of the tumor cells [15]. Two examples of putative tumor suppressors identified by array CGH are the inhibitor of nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NF-kB) activation TNFAIP3/A20, target of 6q23 deletions in MZL [16], and the PRDM1/BLIMP1 transcription factor, inactivated by 6q21 deletions in DLBCL [17]. DNA array studies have also identified a number of chromosomal alterations with prognostic implications. Some of the regions involved are well-known targets of molecular alterations with major pathogenetic significance, such as the inactivation of the tumor suppressor genes TP53 in 17p or CDKN2A/CDKN2B in 9p21 [18, 19].

gene expression profiling

Microarray technologies allow the study of the global GEP of tumors by measuring the quantity of RNA transcripts which are labeled and hybridized on the array [20, 21]. The role of GEP studies in refining the characterization of known entities and the identification of new subgroups is illustrated by the case of DLBCL, not otherwise specified (DLBCL-NOS). GEP studies identified two gene expression signatures which contributed to the definition of two major subgroups: the germinal center B cell-like (GCB DLBCL), characterized by the expression of genes related to germinal center cells, and the activated B cell-like (ABC DLBCL), with an expression pattern related to mitogenically activated B cells close to cells with a secretory function [17, 22–25]. Interestingly, the ABC and GCB subtypes show different clinical, pathological and biological features, supporting the idea that they may correspond to different entities [26].

GEP studies have been helpful for further characterizing the subtypes and new variants of known entities. For example, in the case of chronic lymphocytic leukaemia (CLL), GEP studies revealed a differential expression profile in the two major CLL
molecular subtypes with mutated and unmutated \( IGHV \) [27–29]. Further, the demonstration that cyclinD1-negative small B-cell lymphomas with morphologic and immunophenotypic characteristics of MCL shared the same GEP with conventional cyclinD1-positive MCL supported the idea that they may correspond to a peculiar variant of this disease [30, 31]. Two major studies have described the GEP of BL and have identified elements which proved instrumental to improve the differential diagnosis versus DLBCL [32, 33]. Major findings in both studies were the identification of aggressive B-NHL with morphologic features of DLBCL carrying a molecular GEP of BL (mBL), cases of mBL that lacked the rearrangement of MYC and the observation that molecular distinction between BL and DLBCL in some cases is not clear-cut. Interestingly, tumors with an intermediate signature between BL and DLBCL frequently showed the overexpression of BCL2, t(8;14) translocation and additional BCL2 or BCL6 rearrangements, more complex karyotypes, and presented in older patients with a worse outcome, when compared with cases with concordance between molecular and pathology diagnosis [32, 33]. These observations suggest that some aggressive B-NHL cases may have features intermediate between BL and DLBCL and support the proposal of the updated 2008 WHO classification to include a new category called B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and BL [6].

GEP studies have contributed to the discovery of biomarkers with diagnostic and/or prognostic significance, some of which have already found implementation in clinical practice. In CLL, GEP identified the expression of the ZAP70 kinase as a relative specific characteristic of the subtype with unmutated \( IGHV \). The detection of ZAP70 by FCM or IHC is now being used as a good prognostic factor in the disease [34–36]. The transcription factor \( SOX11 \) represents another interesting marker that can be detected by routine IHC in almost all MCL cases and it is particularly useful for the diagnosis of cyclinD1-negative variants [37].

Microarray studies in malignant lymphomas have provided new and robust prognostic models that improve the current schemes mainly based on clinical criteria, such as the International Prognostic Index. Thus, in DLBCL, the best predictors of survival in patients treated with immunotherapy include the signatures related to the cell of origin combined with signatures reflecting different cell populations of the tumor microenvironment [26, 38]. In particular, GCB origin and a signature related to extracellular matrix deposition and inflammatory cell infiltration are associated with improved patient outcome, whereas ABC subtype and a signature reflecting angiogenesis appear to be predictive of poorer survival [26].

Finally, the elucidation of oncogenic signaling pathways by GEP in different tumor subtypes may help to identify both novel drug targets and to define the patient population which is most susceptible to benefit from the different therapeutic strategies. Activation of the NF-κB pathway has been shown to be required for tumor cell survival in ABC DLBCL, which renders this pathway an attractive target for new therapies [39, 40]. The frequent platelet-derived growth factor receptor, alpha polypeptide overexpression in peripheral T-cell lymphoma may also open up opportunities for new targeted therapies in this disease [41]. The vast amounts of emerging information in the field of lymphoma genomics, together with the increasing availability of molecularly targeted drugs, emphasize the need to translate this growing body of knowledge into the clinical practice [42].

transferring genomic knowledge into the clinic

While GEP studies have identified robust diagnostic and prognostic signatures for several entities, the translation of this knowledge into clinical practice has been difficult. Microarray studies require costly equipment and relatively large amounts of high-quality RNA, preferably extracted from fresh-frozen material. In order to circumvent these practical limitations, protocols for nucleic acid extraction from routinely processed tissues have improved [43, 44] and microarray platforms specifically customized for some tumors by limiting the probe sets to those relevant for the signature of interest have been developed [45, 46]. Unfortunately, customized arrays for lymphomas are not yet commercially available. An alternative approach for the simultaneous quantitative assessment of the expression of a selected set of genes would be the use of other techniques, such as quantitative real-time PCR or RNAse protection assay, which might find easier implementation in routine diagnostic practice given the more widespread availability of the necessary equipment. Several studies have obtained promising results in the diagnosis of molecular subtypes of DLBCL or applying the MCL proliferation signature using a very limited number of genes with these techniques [47–49]. Some individual biomarkers identified by microarrays may be used routinely using practical methods such as FCM or IHC. ZAP70 and Annexin A1 are examples of individual biomarkers with a diagnostic or prognostic value for hairy cell leukemia and CLL, respectively, assessed by these techniques. Nevertheless, the translation of complex GEP signatures into clinics has been more difficult. The extensive application of a small set of selected IHC markers such as CD10, BCL6 and MUM1/IRF4 of the Hans classification to distinguish DLBCL, NOS into the subgroups of GCB and ABC DLBCL has been controversial. Some studies have shown a good correlation of the IHC evaluation and the GEP classification and the prognosis of the patients [50], whereas other studies have not confirmed these results [51, 52]. The reasons for these discordances are diverse and include the small number of antibodies available to capture the information of complex GEP signatures composed of a relative large number of genes and the difficulties in the interobserver evaluations, among others [51]. To this regard, recent computer-assisted approaches appear to overcome some of the limitations in terms of reproducibility and quantitative assessment [53]. Further studies are needed to implement all the information generated by microarrays technologies into the clinical practice.

next-generation sequencing

The new generation of sequencing technologies opens up new possibilities for the analysis of the mutational spectrum of
cancer due to their high speed, relative low cost and versatility to detect all types of genomic alterations. These methodologies are based on the fragmentation of DNA, subsequent amplification of the resulting DNA fragments and their simultaneous in-parallel sequencing. The millions of sequenced ‘reads’ are then aligned against and compared with the reference genome. The massive production of parallel sequences generates several reads for each given position of the genome. The number of reads per stretch of DNA is called ‘coverage’. By means of bioinformatic analyses based on complex algorithms that compare the sequences of tumor and constitutional DNA from the same individual the following kinds of genomic alterations can be detected: point mutations, small insertions or deletions (‘indels’), larger structural alterations, such as gains, amplifications, and hemi-/homozygous deletions (deviations from mean coverage per region) or translocations (mapping of a number of reads in two distant regions of the chromosome or in different chromosomes) [1].

These sequencing studies may be applied to the whole genome (WG), to specific regions of the genome, including all coding exons (exome), or to the whole transcriptome. WG sequencing provides a comprehensive view of all types of somatic mutations and indirect evidence regarding possible mechanisms involved in the mutational process [54]. Sequencing of specific targeted regions of the genome (or the exome) represents a cheaper and faster method, based on a selective capture of the genomic fragments of interest using tagged complementary oligonucleotides. Transcriptome or RNA sequencing (RNAseq) of complementary DNA fragments derived from total, messenger RNA or other types of RNA (e.g. microRNA) allows a quantitative assessment of gene expression. In contrast to microarray-based GEP platforms microRNA sequencing provides a comprehensive view of all types of sequences of tumor and constitutional DNA from the same organism the following kinds of genomic alterations can be detected: point mutations, small insertions or deletions (‘indels’), larger structural alterations, such as gains, amplifications, and hemi-/homozygous deletions (deviations from mean coverage per region) or translocations (mapping of reads in two distant regions of the chromosome or in different chromosomes) [1].

These sequencing studies may be applied to the whole genome (WG), to specific regions of the genome, including all coding exons (exome), or to the whole transcriptome. WG sequencing provides a comprehensive view of all types of somatic mutations and indirect evidence regarding possible mechanisms involved in the mutational process [54]. Sequencing of specific targeted regions of the genome (or the exome) represents a cheaper and faster method, based on a selective capture of the genomic fragments of interest using tagged complementary oligonucleotides. Transcriptome or RNA sequencing (RNAseq) of complementary DNA fragments derived from total, messenger RNA or other types of RNA (e.g. microRNA) allows a quantitative assessment of gene transcripts. In contrast to microarray-based GEP platforms discussed above, RNAseq does not require preselection of known reference genes and offers additional possibilities such as the identification of new fusion transcripts or alternative splice forms.

**panorama of somatic mutations in lymphoid neoplasms**

Very recently, the sequences of the WG, exome and transcriptome have been reported for several lymphoid neoplasms including CLL [54–57], hairy-cell leukemia (HCL) [58], FL [59], DLBCL [59, 60] and plasma cell myeloma (PCM) [61]. Across these studies, the number of somatic mutations identified in the WG varies from ~0.9/Mb in CLL to 2.9/Mb in PCM, and this variability is also observed at the exome studies (5–20 in CLL to ~35 in PCM) [54, 61]. Although the number of cases examined in most of these studies is relatively low, some patterns in the mutational profile seem to be common across different entities. In most cases, only a few genes are consistently mutated in a large proportion of the cases while the majority of mutations are only observed infrequently. In CLL, sequencing studies have revealed its marked genetic heterogeneity, with only few genes being recurrently mutated in 10–15% of the cases. These tend to cluster along signaling pathways such as NOTCH1 and WNT, RNA splicing and processing machinery, inflammatory response, DNA damage and cell cycle control [56, 57]. Although some mutations are distributed equally between the IGHV-mutated and unmutated CLL subtypes, other genes appear to be preferentially mutated in one of the two subtypes. Overall, the mutational pattern in DLBCL is similar to CLL in terms of relative distribution, with some mutations being more common in the ABC (e.g. BCR signaling and NF-κB pathway genes CD79b, MYD88 and A20) or GCB (e.g. BCL2 and EZH2) molecular subtypes of DLBCL.

**Table 1. Whole genome, exome and transcriptome studies in lymphomas**

<table>
<thead>
<tr>
<th>Study</th>
<th>Samples sequenced</th>
<th>Tumor</th>
<th>Average no. of mutationsa</th>
<th>Genes mutated in ≥10% of the samples screenedb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puente et al. [54]</td>
<td>4</td>
<td>CLL</td>
<td>20</td>
<td>NOTCH1 (31/255)</td>
</tr>
<tr>
<td>Wang et al. [57]</td>
<td>91</td>
<td>CLL</td>
<td>20</td>
<td>TP53 (15/9), SF3B1 (15/91), MYD88 (10/91)</td>
</tr>
<tr>
<td>Quesada et al. [56]</td>
<td>105</td>
<td>CLL</td>
<td>12</td>
<td>NOTCH1 (31/255)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>CLL, IGHV mutated</td>
<td>13</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>CLL, IGHV unmutated</td>
<td>11</td>
<td>NOTCH1 (7/69), SF3B1 (15/73), POT1 (5/45)</td>
</tr>
<tr>
<td>Fabbri et al. [55]</td>
<td>5</td>
<td>CLL</td>
<td>8</td>
<td>NOTCH1 (2/5)</td>
</tr>
<tr>
<td>Tiacci et al. [58]</td>
<td>1</td>
<td>HCL</td>
<td>5</td>
<td>BRAF (48/48)</td>
</tr>
<tr>
<td>Morin et al. [59]</td>
<td>1</td>
<td>FL</td>
<td>35</td>
<td>MLL2 (31/35), MEF2B 35/261)</td>
</tr>
<tr>
<td>Morin et al. [63]</td>
<td>13</td>
<td>DLBCL</td>
<td>22</td>
<td>EZH2 (31/320)</td>
</tr>
<tr>
<td>Morin et al. [59]</td>
<td>13</td>
<td>DLBCL</td>
<td>22</td>
<td>MLL2 (12/37), MEF2B 34/259)</td>
</tr>
<tr>
<td>Pasqualucci et al. [60]</td>
<td>6</td>
<td>DLBCL</td>
<td>14</td>
<td>SGK1 (18/106), GNA13 (18/106)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MLL2 (21/92), CREBBP (20/111), TP53 (19/111), MYOM2 (9/54), TNFAIP3 (17/111), PIM1 (16/111), CD36 (7/54), B2M (14/111), CD79B (13/111), PRDM1 (13/111), CARD11 (11/111)</td>
</tr>
<tr>
<td>Chapman et al. [61]</td>
<td>38</td>
<td>PCM</td>
<td>32</td>
<td>NRAS (9/31), KRAS (10/38), FAM46C (5/38), DUS3 (4/38)</td>
</tr>
</tbody>
</table>

*It stands for average number of nonsynonymous point mutations and indels in the protein coding region.

bAdditional samples were screened for the listed mutations; the number of samples with mutation over the total number is given in parentheses.

CLL, chronic lymphocytic leukaemia; HCL, hairy-cell leukaemia; FL, follicular lymphoma; PCM, plasma cell myeloma; DLBCL, diffuse large B-cell lymphoma.
In contrast, a pattern defined by one specific mutation being present in almost all cases and additional numerous mutations occurring at low frequency and clustering along common genetic pathways has been observed for BRAFV600E in HCL [58] and MYD88L265P in Waldenström’s macroglobulinemia (WM) and lymphoplasmacytic lymphoma [62] (Table 1).

For some of the genes found to be mutated in recent studies, extensive information on their functional role and clinical or diagnostic implications in the same or different entities is available. Nevertheless, other genes were previously unknown to be mutated in human cancer and could potentially constitute novel markers with a diagnostic and prognostic value and/or novel targets for drug discovery. For some of these genes, targeted drugs are already licensed for different indications (e.g. BRAFV600E inhibitor vemurafenib) or in clinical development. Moreover, some of these genes are mutated across different entities, e.g. BRAF in 100% of HCL, but also in 2% of CLL and 4% of PCM cases analyzed [56, 58, 61], which unveils a broad range of potential indications for clinical research of targeted agents.

In light of the immense potential of NGS techniques and other high-throughput technologies to improve our understanding of the pathogenesis of individual lymphoma entities, translating these findings into advances in clinical practice will probably represent the biggest challenge in the near future. Developing effective targeted drugs, either in monotherapy or in combinations, supported by a strong biological rationale tailored, e.g. to the mutational profile of the patient, and defining the population most likely to benefit from these novel therapeutic strategies, by means of very refined diagnostics involving biomarkers, could translate into major advances in the management of lymphomas.

**funding**

DS was supported by the Swiss National Science Foundation and by the Swiss Cancer League. EC research is supported by the CICYT SAF 08/SAF 2008-03630 and Instituto de Salud Carlos III, ‘Red Temática de Investigación Cooperativa de Cáncer’ (RTICC) (RD06/0020/0039).

**disclosure**

The authors have declared no conflicts of interest.

**references**


