A Web-Based Database for Differential Diagnosis of Hematologic Neoplasms Using Flow Cytometry

Condron, M; Nguyen, N.
Department of Pathology and Laboratory Medicine, University of Texas-Houston Medical School

ABSTRACT

Context
Interpretation of immunophenotyping results by flow cytometry involves recognition of patterns of expression of various immunologic markers by different hematologic neoplasms. Both the lack of consistency in marker expression for a particular neoplasm and the expression of similar patterns of markers by different neoplasms often hinder interpretation of marker results. A particular marker may be positive (or negative) for a certain neoplasm in most cases, but exceptions are often seen, and a definitive diagnostic pattern is usually not available for a given neoplasm [1-4]. Consequently, there is a need for decision-support tools to assist clinicians in diagnosing leukemia and lymphoma using flow cytometry data.

Design
We attempt to address this need by developing a web-enabled decision-making system that considers the incidence of positive and negative results of each marker for each disorder. This system enhances our previously described database for differential diagnosis of hematologic malignancies by including a more refined algorithm for ranking the disorders in the differential diagnosis for a particular marker pattern and by incorporating the latest World Health Organization (WHO) classification for hematologic neoplasms [4]. The current database includes 37 disorders and 44 markers, and it shows significant improvement in diagnostic accuracy over our previous system [5,6].

Results
Using validation data from 92 clinical cases from two medical centers, the present system ranked the actual diagnosis within the top three differential diagnoses in 89% of cases.

Conclusions
A computer-based decision-making system can be a useful aid in diagnosing hematologic malignancies with complex marker immunophenotypes.

REFERENCES

1. Keren, D.F., *Flow Cytometry in Clinical Diagnosis*. 2001, Chicago: American Society of Clinical Pathologists Press.
2. Sun, T., *Color Atlas and Text of Flow Cytometric Analysis of Hematologic Neoplasms*. 2002, New York: Igaku-Shoin.
3. Kjeldsberg, C., ed. *Practical Diagnosis of Hematologic Disorders*. 4th Edition, American Society of Clinical Pathologists Press: Chicago.
4. Jaffe, E.S., et al., eds. *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. WHO Classification of Tumours Series. 2001, IARC Press: Lyon.
5. Nguyen, A. and e. al, *A Web-based Database for Diagnosis of Haematologic Neoplasms Using Immunophenotyping by Flow Cytometry.* Medical Informatics & The Internet in Medicine, 2001. **26**(4): p. 309-323.
6. Nguyen, A.N.D. and e. al, *A Relational Database for Diagnosis of Hematologic Neoplasms using Immunophenotyping by Flow Cytometry.* American J. Clinical Pathology, 2000. **113**: p. 95-106.

BACKGROUND: CD-MarkerDX

This database prototype, the predecessor of the current database, represents an innovative application of medical informatics to laboratory diagnosis of leukemia and lymphoma. A total of 33 hematologic neoplasms and 42 immunologic markers were included in database CD-MarkerDX [5,6]. The diagnostic criteria for different neoplasms were based on the pattern of immunologic marker results. The marker result for a neoplasm was designated positive if, in the reviewed literature, more than 50% of the cases were found to be positive, or negative if less than 50%.

A list of differential diagnoses is provided by CD-MarkerDX with each set of input data. The differential diagnoses have an assigned value of matching factor (MF). The MF value for a neoplasm reflects how well its immunophenotyping pattern matches the marker data in a given case. This factor is defined as [5]:

MF = M / (M+N)

Where
MF = matching factor (0 • • • • • †
M = number of attributes matching the input data
N = number of attributes not matching the input

If several neoplasms in the database have the same MF for a given set of input results, the value of (M-N) is used as a secondary criterion in ranking differential diagnoses.

We tested this database using 92 clinical cases from two tertiary medical centers. The database ranked the actual diagnosis as one of the top three differential diagnoses in 80% of the cases tested (Table 1).

Note that in this database prototype, a marker is defined as either positive or negative for a certain disorder for simplicity. While CD-MarkerDX represents an improvement in designing a decision-support program for a wide audience via the Internet, it has not been designed to take into account the incidence of positive and negative results for each marker for each disorder. In the current project, this is a major feature of the designed database.

Table 1: Summary of the Validation Results in Our Preliminary Study [5, 6]

Rank	Number of Cases	Percentage	Accumulated Percentage
First	39	42	42
Second	23	25	67
Third	12	13	80
Fourth	10	11	91
Fifth	2	2	93
Lower	6	7	-
Total	92	100	-

IMPROVEMENTS IN THIS STUDY: CD-MarkerDX-PF

This study improves the previous work by introducing a different matching parameter, which we call the profile factor, PF.

PF = Sum (C_n) / N

where

C_n = profile coefficient for input data element n
n = 1 to N
N = number of non-empty attributes of a neoplasm with input data specified.

Calculation of the profile coefficient, C_n, is based on calculating the following ratios, derived from surveying published literature.

PosRatio(M,X) =
PosCases(M,X) / NumCases(M,X)

PosCases(M,X) = number of cases of disorder X in the literature which are positive for marker M

NumCases(M,X) = number of cases of disorder X in published literature which examined marker X

Also define:

NegRatio(M,X) = 1 – PosRatio(M,X)
Now C_n is assigned

PosRatio(M,X) if input data is +
NegRatio(M,X) if input data is -

and PF can be calculated.

PF is defined in a manner very similar to MF in the previous study, but since it incorporates retrospective data from a large number of cases, it is expected to be a better predictor of how well a given set of input marker data results matches a particular neoplasm's typical immunophenotype.

An additional improvement is that the importance of certain critical markers in diagnosing a disorder is also considered. If a certain marker is very specific for a disorder, its contribution is weighed twice as much as other markers in the disorder's profile.

For differential diagnosis of a given case, all the data that are available on marker results should be entered for the case under consideration. Lack of information in certain data fields does not prevent the database from processing the data. However, the accuracy of the suggested diagnosis would be compromised if results of important markers were left out.

The system calculates a value of PF for each neoplasm in the database and sorts them to generate a list of differential diagnoses that most closely match the marker data set that was entered.

Note that the PF values are not probabilities, but rather measures of how well a given immunophenotypic profile matches each neoplasm in the database. The difference between the PF and the previously defined MF is that the MF simply counted the number of markers which matched the data for a given neoplasm. PF weighs each count by the likelihood that the marker should have the state it does, based on published literature. Additionally, a subjective factor of 2 is applied for markers which are particularly specific for a neoplasm.

The system calculates a value of PF for each neoplasm in the database and sorts them to generate a list of differential diagnoses that most closely match the marker data set that was entered.

The results of this approach are summarized in Table 2. Comparison with Table 1 shows significant improvements in identifying the correct neoplasm and placing it in the top 3 differential diagnoses.

Table 2. Summary of Validation Results

Rank	Number of Cases	Percentage	Accumulated Percentage
First	55	60	60
Second	14	15	75
Third	13	14	89
Fourth	4	4	93
Fifth	2	2	95
Lower	4	4	98
Total	92	100	-

In 4% of the cases, the final diagnosis was ranked below the top 5 differential diagnoses for the following reasons:
1. Unusual immunophenotype: a case of CD5-positive diffuse large B cell lymphoma
2. Three cases of T-cell lymphoma: this deficiency is due to an intrinsic limitation of the database in diagnosing T-cell malignancies. Aberrant loss of random T-cell antigens is a characteristic finding in T-cell malignancies [1-4]. A suitable inference mechanism has not been successfully developed to detect such manifestation. Despite this shortcoming, a considerable number of T-cell cases are successfully ranked in the top five differential diagnoses by the database (11 cases out of 14, or 79% of the T-cell cases).
3. Incorrect final diagnosis: a case of lymphoplasmacytic lymphoma was not ranked by the database in the top five differential diagnoses. Review of microscopic slides and flow cytometry data for this case indicated that the diagnosis should be revised as CLL/SLL. In fact, CLL/SLL was ranked second by the database.

http://dpalm.med.uth.tmc.edu/faculty/bios/nguyen/Decision.html

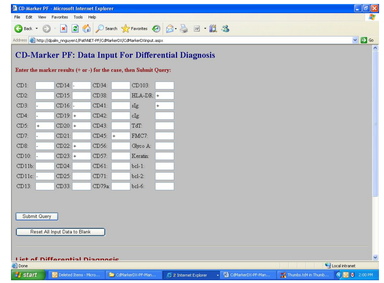

Fig. 1. Screen shot of data input for a consultation session
Marker data are entered into text boxes in the browser interface. A plus sign (+) or a minus sign (-) is entered if the marker result is positive or negative, respectively. If a marker result is not available for the disorder, the box is left blank. The user then clicks on the button "Submit Query" to obtain a list of differential diagnoses (see Fig. 2).

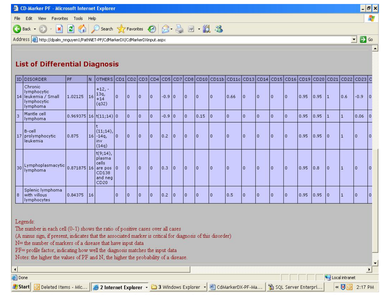

Fig.2. Screen shot of differential diagnoses generated from the user input
The input data are compared against all of the 37 neoplasms in the database and a list of differential diagnoses is displayed, ranked in decreasing order of match. The user can retrieve the marker profile for each neoplasm before making a final diagnosis (see Fig. 3).

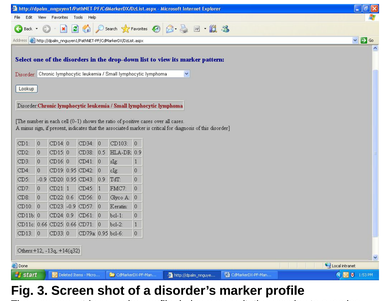

Fig. 3. Screen shot of a disorder's marker profile
The user can retrieve marker profile during a consultation session to examine the complete diagnostic criteria for each neoplasm.

Fig. 4. Screen shot of a marker description
The user can retrieve marker information during a consultation session to obtain more information on property of each marker.