J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Research Article

A pathologist-in-the-loop IHC antibody test selection using the entropy-based probabilistic method

Dmitriy Shin[1,4], Gerald Arthur[1,2,4], Charles Caldwell[1,2,4], Mihail Popescu[3,4], Marius Petruc[4], Alberto Diaz-Arias[1], Chi-Ren Shyu[4,5]

[1]Departments of Pathology and Anatomical Sciences, [2]Ellis Fischel Cancer Center, [3]Health Management and Informatics, [4]MU Informatics Institute, [5]Computer Science, University of Missouri

E-mail: *Dmitriy Shin - shindm@health.missouri.edu
*Corresponding author

Received: 19 September 11 Accepted: 21 November 11 Published: 29 February 12

This article may be cited as:
Shin D, Arthur G, Caldwell C, Popescu M, Petruc M, Diaz-Arias A, *et al.* A pathologist-in-the-loop IHC antibody test selection using the entropy-based probabilistic method. J Pathol Inform 2012;3:1.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2012/3/1/1/93393

Abstract

Background: Immunohistochemistry (IHC) is an important tool to identify and quantify expression of certain proteins (antigens) to gain insights into the molecular processes in a diseased tissue. However, it is a challenge for pathologists to remember the discriminative characteristics of the growing number of such antigens across multiple diseases. The complexity of their expression patterns, fueled by continuous discoveries in molecular pathology, gives rise to a combinatorial explosion that places an unprecedented burden on a practicing pathologist and therefore increases cost and variability of IHC studies. **Materials and Methods:** To tackle these issues, we have developed antibody test optimized selection method, a novel informatics tool to help pathologists in improving the IHC antibody selection process. The method uses extensions of Shannon's information entropies and Bayesian probabilities to dynamically build an efficient diagnostic tree. **Results:** A comparative analysis of our method with the expert and World Health Organization classification guidelines showed that the proposed method brings threefold reduction in number of antibody tests required to reach a diagnostic conclusion. **Conclusion:** The developed method can significantly streamline the antibody test selection process, decrease associated costs and reduce inter- and intrapathologist variability in IHC decision-making.

Key words: Antibody test selection, clinical decision support, computer--aided pathology diagnosis, entropy maximization, knowledge representation

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.93393

Quick Response Code:

INTRODUCTION

An extensive corpus of fundamental medical knowledge has been gained through decades of the scientific examination of tissues stained with the basic hematoxylin and eosin stains. Countless studies have identified precise diagnostic entities based on fine visual and logical distinctions evident from alterations of cellular morphology.[1] Even more precise diagnostic entities are being discovered based on genomic and proteomic signatures of the diseases and their subclasses. Because of that, it becomes increasingly important to utilize methods of diagnostic molecular pathology to provide precise diagnosis as well at the genomic/proteomic level, which can lead to improved treatment and therefore better clinical outcomes. However, there are several

challenges associated with genomic/proteomic studies.

Figure 1 depicts a typical pathology diagnostic workflow involving extensive iterative processes. A diagnostic process usually starts with some preliminary hypothesis (step 1) after receiving initial information and gathering initial evidence. If further morphological observations and clinical findings (step 2) reject that hypothesis (step 3), a new hypothesis is generated (step 4). However, if the evidence supports the hypothesis (step 3), the pathologist may or may not immediately report it as a diagnosis. Instead, he or she might be interested in knowing if the diagnosis can be refined to reveal possible *disease subclasses* and/or additional important genomic and proteomic characteristics (step 5). If so, a refined hypothesis is generated (step 7). New evidence is then collected (step 8) and analyzed to see if the refined hypothesis can be accepted (step 3). Steps 3,5,7,8 can be repeated in a loop several times until after all possible refinement attempts, which can provide precise evidence for treatment, are exhausted. This diagnostic loop is a technique that pathologists use to gradually drill down to the most precise diagnosis. The technique can be especially important in solving difficult cases and/or to tailor treatment to individual patient genomic makeup in a personalized medicine setting. Among the most important types of evidence being collected during the diagnosis refinement step are the results of application of antibody-linked stains, a.k.a. immunohistochemistry (IHC). IHC tests have permitted the highly specific identification of a diverse variety of cellular proteins that play an essential role in the molecular pathology of a wide range of diseases. These diagnostic tools have also allowed the emergence of new fields of morphoproteomics and morphogenomics,[2] which have great potential to provide more specific and accurate diagnoses and, consequently, a more reliable estimation of prognosis. As a consequence of this transformative technology, great demands have been imposed upon the integrative intellectual skills of even the most experienced and well-trained surgical pathologists. The mental retention and recall of the large number of facts generated during diagnosis refinement in the interpretation of specific protein patterns in biopsies from even a single organ system can be problematic. For example, just for one family of diseases, lymphomas, which can have more than 60 variations,[3] there could be more than 80 different antigens,[4,5] that may need to be analyzed before an interpretation of the underlying biological processes of a given case can be obtained.[6] Currently, the cluster designation system includes more than 400 antigens[7] and the number of proteins functioning as biomarkers keeps increasing with continuous developments in molecular biology.[8-10] Since pathologists can consider more than just two states of protein expression (e.g., weakly expressed, strongly expressed, etc.), the number of all possible combinations

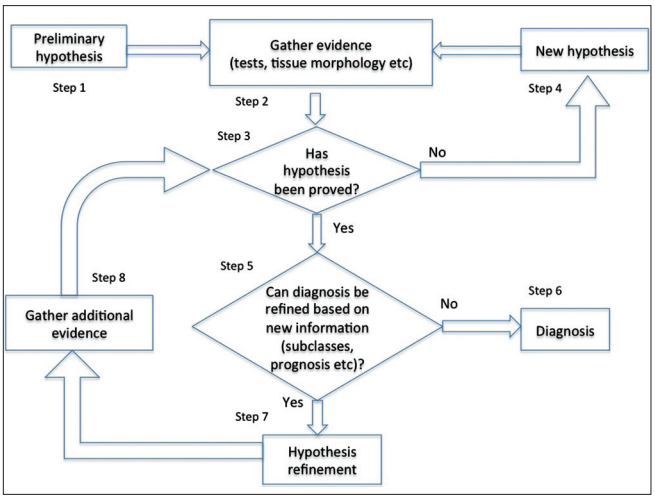

Figure 1: A typical pathology diagnostic workflow

of antigen expression creates a combinatorial explosion that a practicing pathologist can face during the diagnostic process.

Typically, IHC is performed using panels of antibodies ranging from five to a dozen or more as judged appropriate for the tissue and tumor being examined. The composition of a panel is selected to both confirm and rule out diseases for further diagnostic consideration. The exact content of antibody test panels and the sequence in which tests are performed can sometimes vary significantly from pathologist to pathologist (interpathologist variability). Moreover, the same pathologist may order different tests for very similar cases on different days (intrapathologist variability). Such variability results in high inconsistency rates in pathology disease diagnosis.[11] It is very easy for a pathologist to overlook the inclusion of a certain antigen into the panel and miss an opportunity to obtain a crucial clue about the disease in question. On the other hand, a pathologist might include unnecessary tests in the panels with discriminatory power that may not be required. Therefore, it appears to be inefficient to design panels of fixed size to examine predetermined disease groups rather than to permit a dynamic selection of the antibodies required to evaluate a more realistic set of diseases with which the pathologist may be confronted.

Given the central diagnostic importance of IHC and predicted rapid increase in its utilization, it is timely to question whether there may be more effective and efficient schema for the implementation of IHC than those that have been developed empirically over decades. It is becoming increasingly important for modern healthcare to deliver cost-effective service in general and reduce cost of pathology and laboratory medicine in particular.[12] Hence, a method that can save both in pathologists' time and cost of reagents can find widespread application in the Digital Pathology era,[13,14]

when pathology diagnosis is more and more reliant on the assistance of information technology and computational tools.[15]

Previously, there have been attempts with certain degree of success to develop computerized methods to assist pathologists in their decision-making. Early studies include Pathfinder expert system created by Heckerman, *et al*.[16] While this method helps to organize the diagnostic process using etiological and morphological features, it is not suitable for the new challenges and complexity posed by increasing usage of IHC in the pathology workflow. A number of other works had used probabilistic methods to improve decision-making in pathology.[17,18] More recent studies, using a noncomputational approach, on solving problems due to the complexity of IHC tests include works by Taylor.[4,5]

We have identified a few distinct features of the diagnostic workflow [Figure 1] discussed earlier that, in our view, make it hard to directly apply most of the abovementioned traditional methods or human heuristic approaches. First of all, due to its iterative nature, the disease diagnosis refinement process [Figure 1] cannot be simply viewed as a classification problem mainly because there is not a clearly defined final classification goal for such type of process. A pathologist keeps diagnosis refinement (into disease class, subclass, grade, therapeutic target, etc.) as long as he or she can obtain new information. Second of all, in such a diagnostic workflow, there is not a fixed feature space. In this paper, we use feature, fact, factor, and test interchangeably. Instead, the process is based on a dynamic set of facts at hand. Depending on the diagnostic path, the process may require a different set of tests to be conducted. Thirdly, pathologists often use different tests to reveal the same biological mechanisms (e.g., proliferation of B-cells can be identified by a test for CD20 *or* CD23 antigen). Therefore, the IHC panel design philosophy is a complex mental process aimed at constructing a comprehensive but limited set of IHC tests that can generate valuable information at each diagnostic refinement step.

The pathology community needs a robust computational tool that can propose a preliminary diagnostic path and dynamically suggest a set of IHC tests to be performed. To answer this need, we have developed a novel method that helps to streamline IHC studies by dynamically suggesting an efficient set of IHC tests at each step of the pathology workflow. Our approach, while not entirely new computationally, is a novel application of information entropies and Bayesian probabilities in IHC that can significantly improve speed of IHC diagnostic workflow and reduce intra- and interpathologist variability. Our tool also helps to visualize possible diagnostic paths, which brings a great deal of clarity into the diagnostic process.

MATERIALS AND METHODS

Data Set

The data in our work are represented by a matrix of antigen expression profiles (MAEP) across multiple diseases. In general, this type of matrix can be constructed from existing medical literature for specific diseases. Figure 2 shows a small subset of the matrix used in this study.

Rows and columns of matrix A represent correspondingly disease immunotypic profiles across multiple antigens and antigen expression profiles across multiple diseases. Each row starts with the name of a disease followed, after a vertical bar, by that disease's incidence rate. Disease incidence rates for each malignancy were obtained from the Cancer Statistics Review data provided by the Surveillance Epidemiology and End Results (SEER) website, maintained by the National Cancer Institute.[19] We have to mention here that to tune ATOS to geographically local disease incidents rates, the data from regional cancer registries can be used. An element $A_{i,j}$ of the matrix gives information about expression of the antigen j in disease i and probability of getting such expression value. In the tests reported here, the antigen expression scores and their probabilities were derived using primarily studies by Higgins *et al*.[6] and the WHO disease classification,[3] for several families of non-Hodgkin and Hodgkin lymphoma. For simplicity and clarity purposes, we consider only two discrete levels for antigens expression in this paper: expressed ("+") and not expressed ("-"). We use *antigen* interchangeable with diagnostic factor or just simply factor; and antigen test with the antibody test.

	Antigens (ant expr|probability)			
Disease|Inc Rate	**CD11c**	**CD15**	**CD19**	**CD20**
MCL|0.51	-|0.95	-|0.95	+|0.95	+|0.95
CLL/SLL|5.17	+|0.95	ND	+|0.95	+|0.95
FL|3.18	-|0.95	-|0.95	+|0.95	+|0.95
MALT|2.35	+|0.75	-|0.95	+|0.95	-|0.95
MZL nodal|0.66	-|0.95	-|0.95	+|0.95	+|0.95
MZL splenic|0.31	-|0.95	-|0.95	+|0.95	+|0.95
LPL/WM|0.62	ND	-|0.95	+|0.95	+|0.95
HCL|0.33	+|0.95	-|0.95	ND	+|0.95
PCM|5.86	-|0.95	-|0.95	-|0.95	+|0.75
Burkitt|0.3	-|0.95	-|0.95	+|0.95	+|0.95
DLBCL GC|3.21	-|0.95	-|0.95	+|0.95	+|0.95
DLBCL Non-GC|3.93	-|0.95	-|0.95	+|0.95	+|0.95
ALCL ALK +|0.25	ND	-|0.95	ND	-|0.95
MF|0.52	ND	-|0.95	-|0.95	-|0.95
B-ALL/LBL|0.76	ND	-|0.95	+|0.95	-|0.75
T-ALL/LBL|0.22	ND	-|0.95	ND	-|0.95
PTCL NOS|0.3	ND	-|0.95	-|0.95	-|0.95
AILT|0.05	ND	-|0.95	-|0.95	+|0.95
CHL|2.59	ND	+|0.95	ND	-|0.75
NLPHL|0.08	ND	-|0.95	ND	+|0.95

Figure 2: Excerpt from a matrix of antigen expression scores across multiple diseases

Antigen Tree Induction Algorithm

Our method extends Shannon's information entropies[20] and Bayesian probabilities to iteratively build an efficient antigen selection tree (EAS-Tree) using a matrix of antigen expression profiles (*Data Set* section). EAS-Tree can then be used by pathologists to streamline the antibody testing process. The main difference between a traditional decision tree (DT) approach,[21] and our EAS-Tree is that a DT is usually used to confirm a single path of antigen expressions while an EAS-Tree is to cover a comprehensive set of diseases and then rule out irrelevant ones. Moreover, a DT utilizing the entropy minimization concept likely will result in a skewed and deep tree. Such a tree likely will require a large number of antibody tests to provide conclusive recommendations to the pathologists. Unlike the DT approach, the EAS-Tree seeks a balanced tree structure, which suggests to pathologists a small number of antibody tests using an entropy maximization approach. We believe that our method strives to closely mimic the reasoning process of pathologists. To explain the ideas behind this, let us use the following MAEP corresponding to differential diagnosis of which the listed five diseases have equal possibilities for diagnosis (i.e. equal incidence rates).

Antigen A is expressed in only Disease 1 [Table 1]. Intuitively, a pathologist would want to test for an expression of Antigen A, because if the test comes out positive the diagnostic process would quickly end (with Disease 1 as diagnosis in his/her mind). However, if we assume that no other factors can cause an expression of Antigen A, the probability of getting positive expression of Antigen A is 1/5, while the probability of getting negative expression results is 4/5. It means that, no matter what the test outcome of Antigen A will be, it will suggest the following: (1) 20% of chances to confirm Disease 1 ("+" expression) or (2) 80% of chances to rule out only Disease 1 leaving four diseases ("-" expressions) to be checked. From the same example if we chose another antigen, Antigen B, which is expressed in two out of five diseases, the test for that antigen in the most likely case would rule out two diseases. As mentioned previously, an efficient diagnostic process should start with a comprehensive list with diseases and quickly rule out as many irrelevant ones as possible. For this reason, the choice of testing Antigen B is over Antigen A. We

Table 1: An example MAEP: "+" expressed, "-" not expressed

	Antigen A	Antigen B
Disease 1	+	+
Disease 2	-	+
Disease 3	-	-
Disease 4	-	-
Disease 5	-	-

select an antigen Ag_j with maximal entropy,[20] which can be computed by the following formula:

$$H(Ag_j) = -\sum_{l=1}^{Mj} P_{Agj}(l) log_2 P_{Agj}(l)$$

where $P_{Agj}(l)$ is a fraction of diseases in *MAEP* for which antigen Ag_j has lth expression level.

The tree induction algorithm builds an EAS-Tree by splitting the current *MAEP* into partitions, one for every level of expression of a certain antigen. The antigen is then represented by a node with child nodes for each expression level. At each step of the tree induction, the algorithm picks an antigen that splits diseases in current $MAEP_j^l$ as evenly as possible. To represent the matrix of antigen expression profiles, we use *MAEP* with super/subscripts to identify parts of the matrix that were generated during a split by the antigen tree induction step. For example, $MAEP_j^l$ is a submatrix that is resulted from splitting the matrix by jth antigen Ag_j and consisting of diseases in which Ag_j has expression value l.

Figure 3 lists the pseudo code for the selection of antigen and the procedure of node splitting for EAS-Tree building.

Every call of this procedure finds an antigen in *MAEP* with the maximum entropy[20] [Figure 3, line 2], splits current *MAEP* using that antigen [Figure 3, line 4] and create new child nodes in EAS-Tree [Figure 3, line 5]. The antigen with maximum entropy splits current *MAEP* into l partitions, one per expression level of the

Procedure EntropySplit(*MAEP*, *Parent*)

Input:
MAEP - current matrix of antigen expressions
Parent - parent node in EAS-Tree
$\boldsymbol{A} = \{Ag_1, Ag_2, \ldots, Ag_n\}$ – set of n antigens in current ***MAEP***
$\boldsymbol{D} = \{D_1, D_2, \ldots, D_m\}$ – set of m diseases in current ***MAEP***

Output: EAS-Tree

1. **WHILE** ($A <> \emptyset$ *AND* $Size(D) > 1$) {
2. $j_{max} = argmax_j\, Entropy(Ag_j), \forall\, Ag_j \in A$
3. assign $Ag_{j_{max}}$ to *Parent* node
4. form $MAEP^+_{j_{max}} = (D^+_{j_{max}}, A^{new})$,
   $MAEP^-_{j_{max}} = (D^-_{j_{max}}, A^{new})$, and
   $MAEP^m_{j_{max}} = (D^m_{j_{max}}, A^{new})$,
   where $D^+_{j_{max}} \subset D$, for which $Ag_{j_{max}}$ is expressed,
   $D^-_{j_{max}} \subset D$, for which $Ag_{j_{max}}$ is not expressed,
   $D^m_{j_{max}} \subset D$, for which expression of $Ag_{j_{max}}$ is missing and
   $A^{new} = A - Ag_{j_{max}}$
5. add EAS-Tree child nodes $Child^+$ and $Child^-$ to *Parent*
6. call EntropySplit($MAEP^+_{j_{max}} \cup MAEP^m_{j_{max}}, Child^+$)
7. call EntropySplit($MAEP^-_{j_{max}} \cup MAEP^m_{j_{max}}, Child^-$)
8. }

Figure 3: Pseudo code of ATOS's EntropySplit() procedure

splitting antigen (l=2 in this paper). For every resulting *MAEP* partition, the *EntropySplit()* procedure calls itself recursively [Figure 3, lines 6 and 7] and the process repeats. The algorithm stops when there are no available antigens or there is only one disease profile left in the database [Figure 3, line 1].

Since antigen expression information for some diseases might not be available, *MAEP* may have missing values. In our implementation of ATOS we do not impute missing expression values. After splitting *MAEP* into partitions, diseases for which expression value of the splitting antigen is not available are grouped together in $MAEP^{m}_{jmax}$ [Figure 3, line 4] and passed to the next recursive calls of *EntropySplit()* [Figure 3, lines 6 and 7].

In Figure 4a the initial *MAEP*, with 17 diseases, is used to explain the logic of the tree induction process in a recursive way.

During the first call, the algorithm selects antigen Ag_3 as having the largest entropy value of:

$$(Ag_3) = \left(\frac{8}{17}\right) log_2 \left(\frac{17}{8}\right) + \left(\frac{9}{17}\right) log_2 \left(\frac{17}{9}\right) = 0.997 \quad (1)$$

Indeed, the expression values of Ag_3 almost evenly splits *MAEP* into two partitions $MAEP_3^+$ and $MAEP_3^-$, one having diseases with "+" expression values of Ag_3 and the other one with "-" expression values. This is illustrated by the leftmost column in Figure 4a where disease rows are grouped by expression values to form partitions. Ag_3 becomes the root of the antigen tree [Figure 5]. For each resulting partition, the tree induction algorithm recursively calls itself by passing $MAEP_3^+$ and $MAEP_3^-$. At next level of this recursion, the selected antigens are Ag_{11} and Ag_7 for partitions $MAEP_3^+$ and $MAEP_3^-$ respectively [Figure 4a]. Ag_{11} and Ag_7 are added to the tree as child nodes for Ag_3 [Figure 5]. At this step there are four new partitions $MAEP_{11}^+$ and $MAEP_{11}^-$ Figure 4b, $MAEP_7^+$ and $MAEP_7^-$ [Figure 4c]. For each of these partitions, the induction procedure calls itself recursively again and then adds new antigen nodes Ag_4 Ag_4 and Ag_2 to their parent node Ag_{11} and new nodes Ag_{10} and Ag_5 to Ag_7 [Figure 5].

When the algorithm looks for an antigen that best splits a current partition, it chooses from all antigens with the exception of those that appear in the path to the root antigen node. For example, when the algorithm processes partitions generated by Ag_4, it can choose any available antigen except Ag_4, Ag_{11}, or Ag_3, since any of these antigens will have zero entropy for partitions $MAEP_4^+$ and $MAEP_4^-$ [Figure 4a]. However, some antigens may appear in different branches of the antigen tree [see double circled nodes on Figure 5].

To reflect a more realistic picture, instead of simply counting occurrences of expression levels over diseases for each antigen in *MAEP*, there are other important parameters to be considered for entropy calculations.

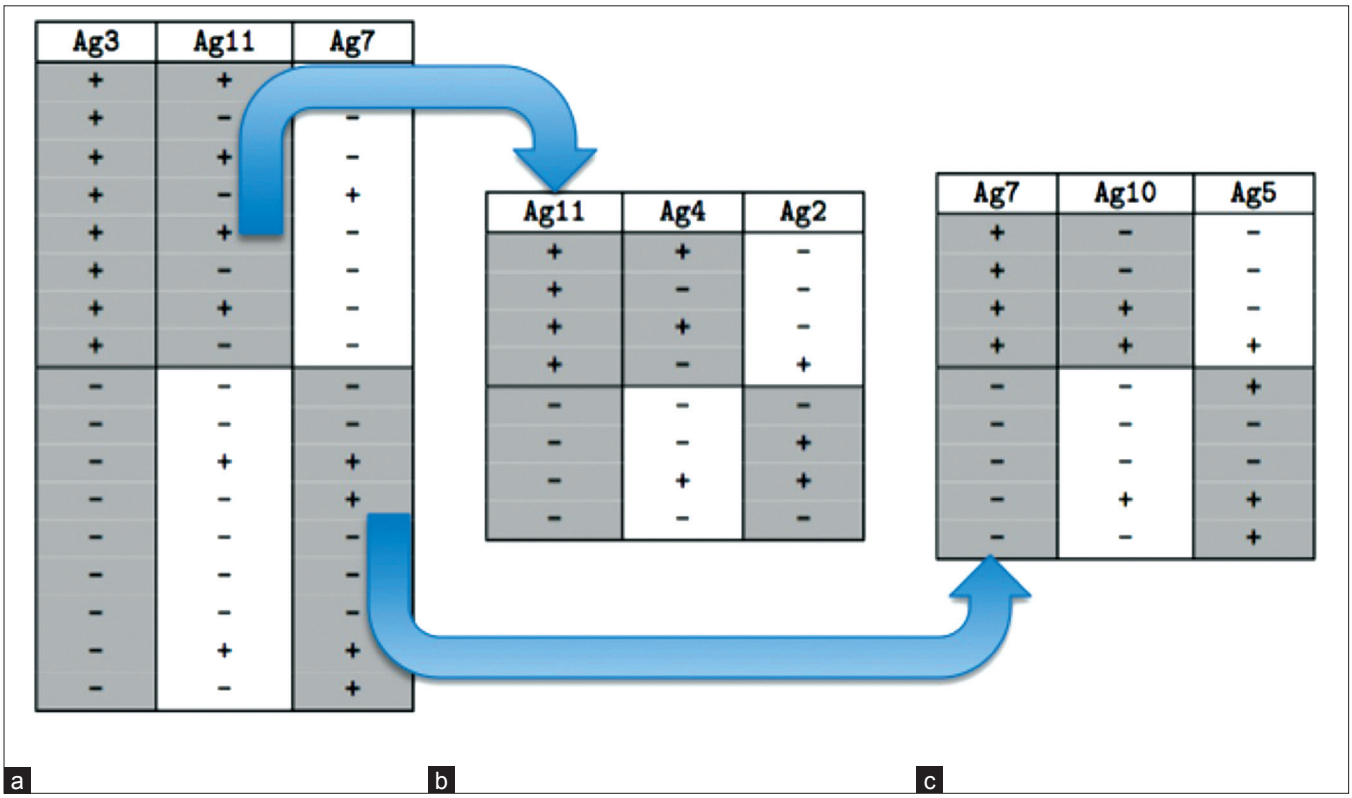

Figure 4: (a-c) An example of splitting of MAEP

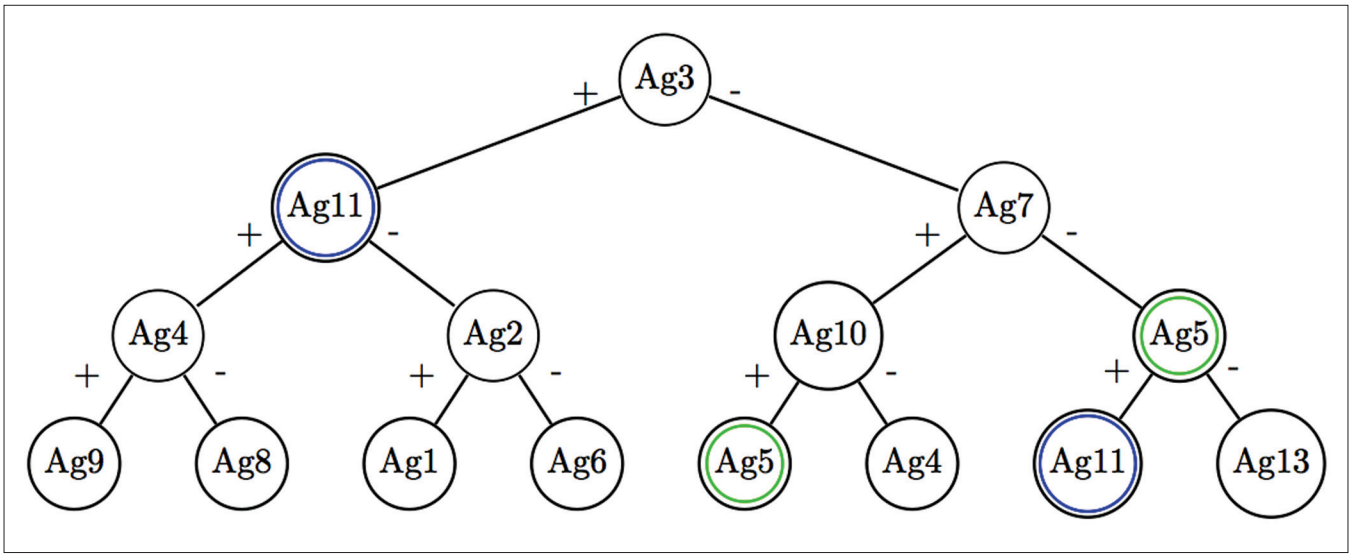

Figure 5: An efficient antigen selection tree (EAS-Tree) based on MAEP from Figure 4

During disease diagnosis pathologists usually take into account disease incidence rates. The rationale behind this is that some diseases are prevalent while others are not. A specific antigen expression level is less likely to manifest the presence of a disease with a lower incidence rate than a disease with a high incidence rate. To address this fact we modify entropy computation and include disease incidence statistics in the following way. First of all, the "raw" disease incidence rates, as discussed in the *Data Set* section, have to be normalized. Since the differential diagnosis assumes that the pathology case in question should be diagnosed only with a disease from *MAEP*, we need to make sure that the incidence rates will sum to unity, i.e. $\sum_{i=1}^{N} P_{in}(D_i)=1$, where $P_{in}(D_i)$ is a normalized incidence rate for disease D_i. We compute $P_{in}(D_i)$ using the following formula: $P_{in}(D_i)=\frac{P_{in}^{raw}(D_i)}{\sum_{k=1}^{N} P_{in}^{raw}(D_k)}$. Now the normalized incidence rates can be used as probabilities to compute joint probability $P^*(l_{Agj}, D_i)$ for a certain expression value l for jth antigen and ith disease:

$$P^*(l_{Agj}, D_i) = P(l_{Agj}|D_i)\, P_{in}(D_i)$$

By summing up all joint probabilities $P^*(l_{Agj}, D_i)$ for each level and then normalizing the sums, we obtain marginal probabilities for each expression level l of jth antigen and use them to compute information entropies with the formula

$$H(Ag_j) = -\sum_{l=1}^{M} P^*(l_{Agj}, D_i) log_2 P^*(l_{Agj}, D_i)$$

where the probability takes into account disease incidence rates and the probabilities for antigen expression levels of certain diseases as listed in Figure 2.

Using an EAST to guide immunohistochemical studies.

After an EAS-Tree is built, a pathologist can interactively use it as a guidance to conduct actual antibody tests. A pathologist-in-the-loop interface provides a graphical interface for pathologists to select antibody tests and an indicator about the effectiveness of the selection. The outcomes of these tests will then be fed back to the system. If a test for a certain antigen has already been completed, this test will be excluded for future testing. Since ideally each test under any outcome should rule out as many disease diagnoses as possible, the whole testing process should result in an efficient testing workflow. In other words, the testing process using this method, on average, should result in faster turnaround.

Another way, which resembles the traditional approach[6] is to aggregate nodes of EAS-Tree into test panels. One of the ways of doing that is to aggregate adjacent levels and construct and antigen *panel* tree [Figure 6]. If the number of distinct expression levels l is constant for all antigens in the database (as in our example), the number of antigens in a panel for a given number of aggregated antigen tree levels K equals to $\sum_{i=1}^{K-1} l^i$. In the general case, the number of the antigens in panels might be different for each node in the antigen panel tree depending on values of l. It has to be noted that while this method does not fully eliminate the redundancy of the expert-based IHC classification schemas,[6] it can result in substantial decrease in the number of unnecessary tests and significantly reduce inter- and intrapathologist variability by making the antibody test selection process uniform.

RESULTS AND DISCUSSION

Method Evaluation

A typical model of the conventional, recommended IHC classification schema for lymphoid tumors was constructed in accordance with text and tables presented in an expert review article[6] as well as WHO guidelines.[3] This model

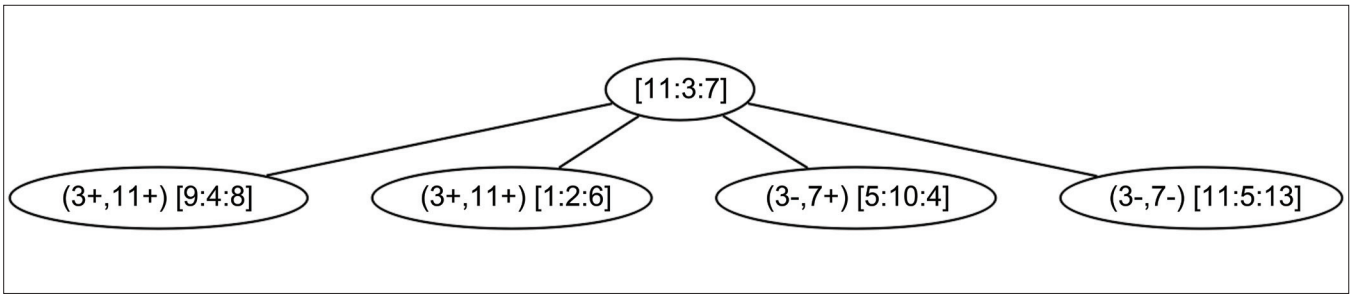

Figure 6: An example of antigen panel tree. The numbers in square brackets identify antigens. The numbers with signs in parentheses indicate outcomes of the testing for corresponding antigens

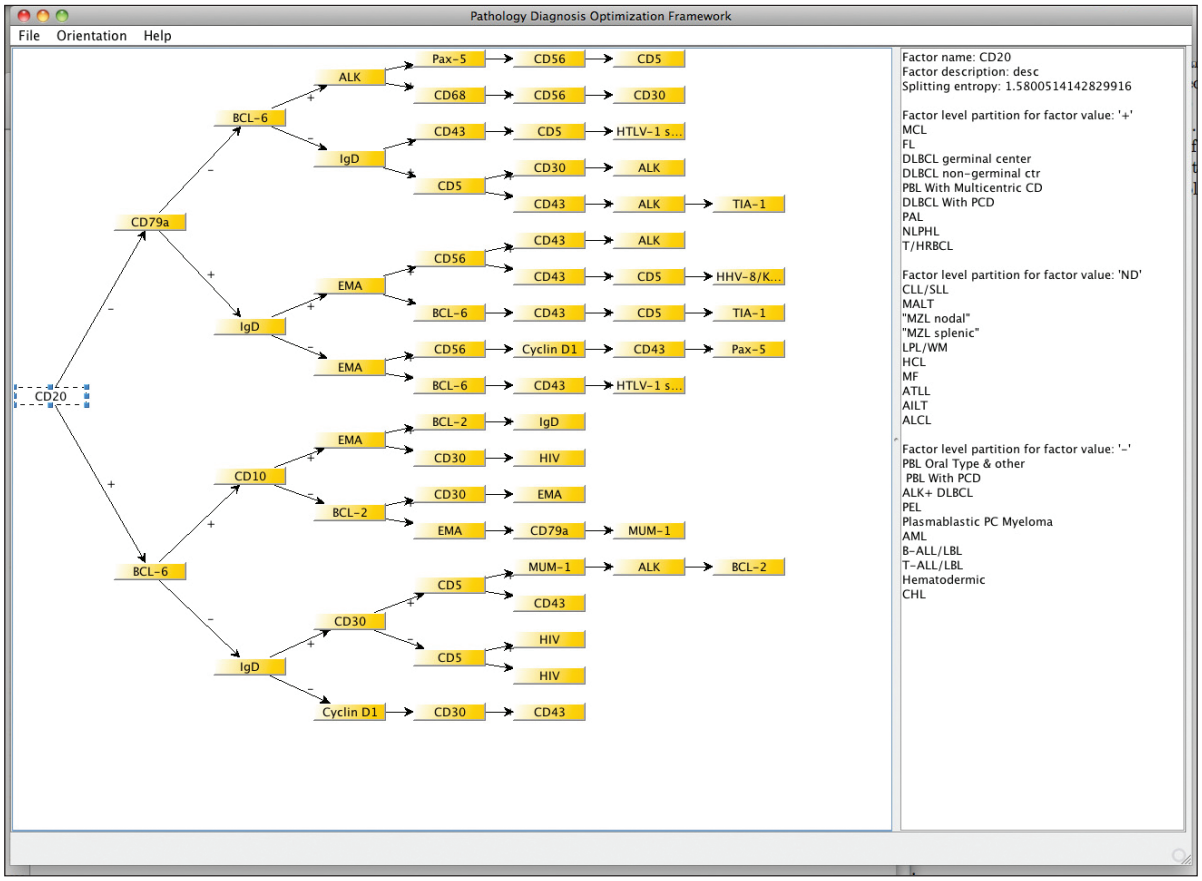

Figure 7: ATOS user interface for diagnostic tree building using an automatic entropy-based probabilistic algorithm.

employs two to three distinct panels performed in series as determined in conjunction with morphologic features within the tumor. To compare our method with this model, a *MAEP* for hematopoietic tumors was compiled from the same sources (see the Methods section). This comprises 20 lymphoma/leukemia cases and 37 antibodies, which represent a relatively broad range of neoplasms and a sizable selection of antibodies. Our fist assessment of the quality of the resulting EAS-Tree is to validate nodes from the first few levels with recommendations from the published literature or guidelines. In Figure 7, the three antigen nodes from the first two levels of the tree are CD20, CD79a, and BCL-6 antigens, which are B-cell markers and suggested to be analyzed first by guidelines in.[6] This tree would be valuable for the pathologists to determine if the case is B-cell or T-cell-based lymphoma/leukemia. Since the tree is balanced, the method finds the shortest path to the smallest subset of disease diagnosis candidates. In the worst-case scenario, a pathologist would need to test eight antibodies out of total 37, and in the best test scenario only six.

Comparison of the Conventional Model (WHO, Experts Guidelines) and Entropy-based ATOS

No diagnostic discrepancies were identified using either the conventional model or our method; however, in general, our method's pathways were distinctly shorter for the majority of disease diagnoses. On average, our

Table 2: Comparison of immunotypic findings for 14 representative lymphomas/leukemias between human expert model (Conv.) and the entropy-based computer algorithm (ATOS) ("+": positive, "- ": negative outcomes)

Disease	Conv. ABs+	Conv. ABs-	ATOS ABs+	ATOS ABs-
ALCL	7	13	4	2
MF	4	16	1	5
MCL	5	10	3	3
CLL	5	10	4	2
FL	5	10	2	4
MALT	3	13	3	4
HCL	7	9	4	1
MZL nodal	5	10	3	3
B-ALL/LBL	7	11	2	4
T-ALL/LBL	8	13	3	3
DLBCL GC	8	10	5	1
DLBCL NGC	9	9	4	3
CHL	7	8	3	3
PCM	2	13	3	3
MEAN	5.86	11.07	3.14	2.93

ALCL: Anaplastic large cell lymphoma, MF: Mycosis fungoides, MCL: Mantle cell lymphoma, CLL: Chronic lymphocytic leukemia, FL: Follicular lymphoma, MALT: Mucosa-associated lymphoid tissue lymphoma, HCL: Hairy cell leukemia, MZL: Marginal zone lymphoma, B-ALL/LBL: Acute B lymphoblastic leukemia, T-ALL/LBL: Acute T lymphoblastic leukemia, DLBCL: Diffuse large B-cell lymphoma, CHL: Classic Hodgkin's lymphoma, PCM: Plasma cell myeloma

method requires slightly less than half of the positive markers and less than a third of the negative markers to reach a definitive diagnosis. This brevity is inherent in the algorithm, which is designed to produce the *shallowest* diagnostic tree. The method stops when the diagnostic branch contains only a single entity. However, if desired, the algorithm could be readily extended to identify any specified number of additional positive or negative confirmatory antigen tests. Table 2 shows a comparison of number of antibody (AB) tests required to reach a diagnostic conclusion using our method (ATOS) and conventional (Conv.) methods.[3,6] A two-sample T-test, conducted with the assumption of unequal variances, was significant for the difference between the mean number of tests in conventional and ATOS panels. For positive, negative, and total tests the *P*-values were correspondingly 2.619e-04, 1.214e-10, and 3.394e-11.

Comparison of Manual Antibody Selection Using ATOS GUI and Entropy-based ATOS

For the next series of experiments we have developed a simulator program that uses the *MAEP* and ATOS GUI and allows a pathologist to build antigen selection tree manually. The goal was to capture the way each pathologist approaches the task of reducing the number of suspected diagnoses at every step. The program includes a graphical user interface.

Interface (GUI) makes the process user friendly and helps

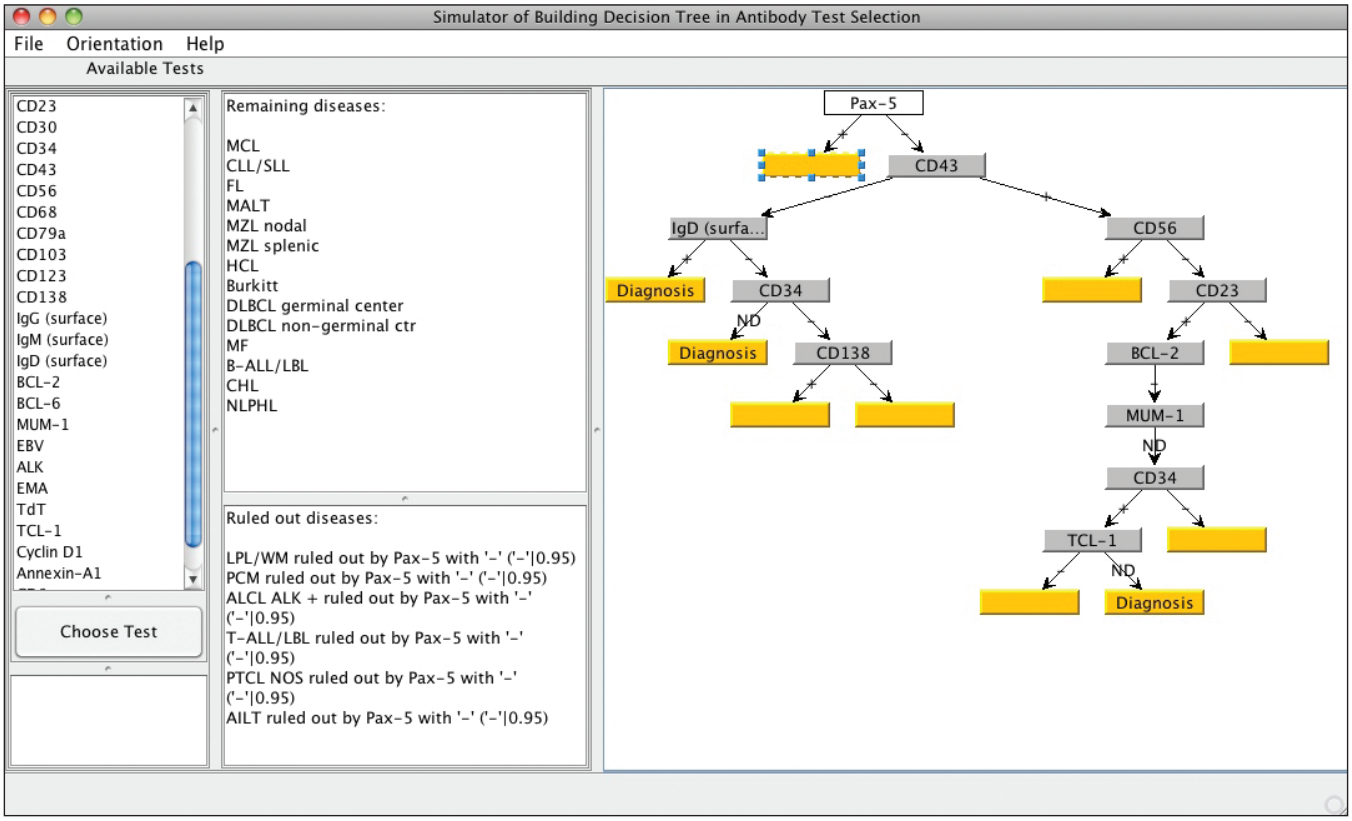

Figure 8: ATOS GUI in the simulator program during manual antibody test selection

to retain the logical order of the diagnostic process rather than rely on memory. This interface is shown in Figure 8. Moreover, the program provides linkages for the antigen descriptions and their genomic/proteomic functions derived from well-known online knowledge bases such as UniProtKB/Swiss-Prot[22] and Gene Wiki.[23]

Five experienced pathologists participated in the experiments, each with overall experience in practicing anatomic pathology ranging from 7 to 30 years. They were asked to build an antigen selection tree manually. The requirements for an experiment session were to select an antibody test for a given set of remaining (suspected) diseases at each step. The pathologists were not told to use any specific selection criteria; they had to use their own methods. Similar to our entropy-based algorithm, once an antibody test has been used, it could not be used again down in the corresponding decision branch. However, it could be used in a different branch of the tree.

To perform comparative analysis of the outcomes of manual test selection using the simulator program with experts,[6] WHO[3] IHC guidelines, and our entropy-based probabilistic algorithm, we plot a bar chart in Figure 9 that shows the number of tests required to diagnose each disease (for manual test selection averages across pathologists are shown). Just by using the ATOS-GUI (the simulator program) and the knowledge base of genomic/proteomic functions, the pathologists could significantly improve the antibody test selection process in terms of reducing the overall number of IHC tests required to reach a diagnostic conclusion. On average, the simulator program helps to reduce the number of tests up to twofold, compared to the expert IHC classification schema and WHO guidelines. Furthermore, by using our entropy-based probabilistic algorithm, we obtain an additional 35-40% decrease in the number of antibody tests.

Interpathologist Variability

To measure an interpathology variability, we calculated an average difference between numbers of antibody tests required to diagnose the same disease by different pathologists participated in our research. We then constructed histograms that show the frequency for each such difference [Figure 10]. For example, from the top panel of Figure 10 we can see that there were three diseases for which pathologists used the same number of tests, five diseases when the number of tests differed on average by one, one disease when the number of tests differed by 3, six diseases when the number of tests differed by four tests, and so on. The bars appearing on the left side of the histogram correspond to lower variability, while the bars on the right side to the higher variability. We repeated the experiment and plotted another histogram, which is shown in the bottom panel of Figure 10. It can be seen from the chart that the difference in the number of tests in some cases can reach 10 tests.

The overall level of interpathologist variability is significant, showing that the current antibody selection

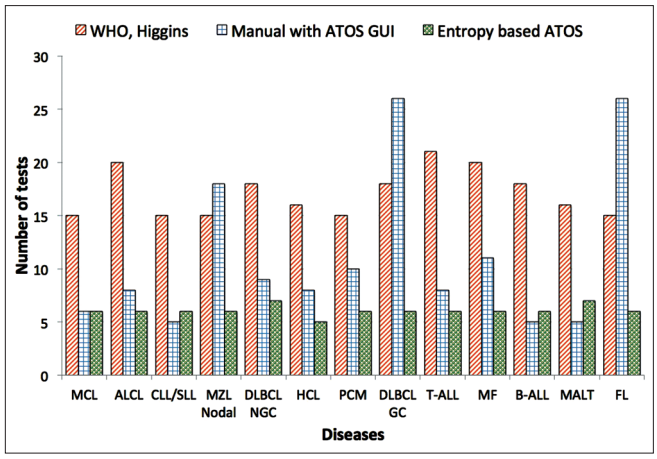

Figure 9: Comparison of conventional antibody test selection (WHO, Higgins), manual test selection using ATOS GUI in the simulator program and the automatic entropy-based probabilistic algorithm (ATOS method)

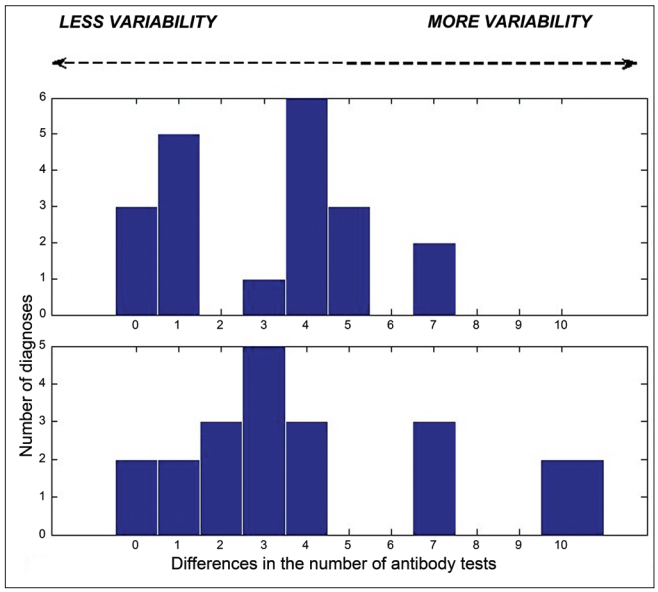

Figure 10: Interpathologist variability (two experiments)

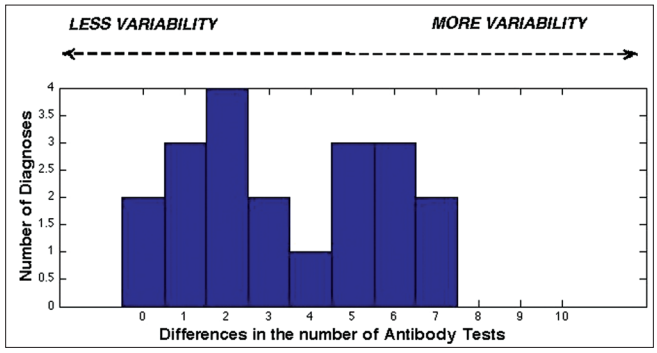

Figure 11: Intrapathologist variability

process is not well standardized. Even though the pathologists were aware of the general guidelines for using IHC for malignancies of lymphoid tissue WHO,[3] each of them used his or her own heuristics to assemble the diagnostic tree. The use of such tacit knowledge by pathologists during decision-making has been previously discussed in the literature.[24,25] Due to consistent tree building by our algorithm based on *MAEP* obtained from the literature, variability using our method is zero in all runs to emphasize the benefit of utilizing our entropy-based probabilistic algorithm in light of maintaining low interpathologist variability.

Intrapathologist Variability

Contrary to our expectation that intrapathologist variability would be smaller than interpathologist variability, interestingly we did not observe that. The overall level of intrapathologist variability had not been found significantly different from interpathologist variability, based on the analysis of the corresponding histograms [Figures 10 and 11]. Studies on other types of cancers have reported similar findings.[11] After further analyses of the EAS-Trees, we concluded that the differences in diagnostic paths generated by the same pathologist can be attributed to the fact that some antigens carry similar functions. The pathologist could have picked a different antigen in place of the one that was no longer available.

Experimental Clinical Study

To preliminary assess clinical utility of ATOS, we have conducted a small retrospective study. A total of 20 cases of hematopoietic tumors have been selected. The diagnoses of these cases were performed in the past by five different pathologists that used traditional methods to choose IHC tests. The cases were then processed again by three expert hematopathologists in the context of this study. They analyzed hematoxylin and eosin (H and E) slides and came up with differential diagnoses, which then were fed into ATOS. For each case, the number of tests suggested by ATOS that lead to original diagnosis was noted. We then set to answer the following two questions: (1) Is the number of ATOS tests that lead to original diagnosis statistically lower than the number of tests in traditional antibody panels? (2) Can ATOS panel be used by a pathologist to arrive to the proper diagnosis?

To answer the first question we have conducted statistical tests using SAS and R tools. Since each case was diagnosed again by three pathologists we had four-sample data setup: one traditional and three ATOS. Statistically the question was to compare means of three ATOS samples with the traditional one. Null hypothesis Bartlett's, Flinger-Killeen's, and Levene's tests showed that the data fail to satisfy homoscedasticity requirement with corresponding *P*-values 6.474e-07, 4.133e-04, and 2.000e-04. The normality assumption violation was demonstrated by Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests (*P* values: 0.0001, <0.0100, <0.0050, <0.0050). We have therefore decided to use nonparametric tests. A Kruskal-Wallis test was significant (*P* value 6.678e-06), which showed overall high level of difference in mean number of tests across all samples. Pairwise differences between each ATOS sample and traditional one were measured by Behrens--Fisher and Steel tests. The results were significant with *P*-values for the Behrens-Fisher test for all pairs being less than 5.84e-06 and for the Steel test being less than 3.08e-04. We have, therefore, come to the conclusion that ATOS suggests a statistically significantly lower number of tests than the number of tests in traditional IHC panels in real clinical cases.

To answer the second question the ATOS panels of several cases from the study were manually analyzed. This instance of an ATOS algorithm is based on immunohistochemical data derived from a limited literature review[6,3] and appears to accurately derive the proper diagnosis. Given the clinical setting and H and E findings, the analysis demonstrated the use of redundant or noncontributory stains during original processing of the cases by conventional panels. Several cases illustrated the type of diagnostic situations in which the ATOS might be of value in the evaluation of lymphomas. For instance, it would (1) exclude examination for widely positive antigens; (2) eliminate the implementation of multiple B-cell antibodies simultaneously; (3) avoid the use of stains which are listed as +/- for multiple tumors in the differential; (4) avoid the use of antibodies to evaluate large cell lymphomas when no large cells are present; (5) exclude the application of epithelial-associated antibodies for tumors with a purely lymphoid appearance. Of course, these would be presented as guidelines, not absolute rules for all cases.

The conducted experiments have shown that the pathology diagnostic process can greatly benefit from our new improved methodology for streamlining antibody test selection in IHC studies. One of the most important advantages of our computerized approach is that it is more quantitative than qualitative. For instance, substitution of antibody tests may lead to significantly different results, because of the subtle differences in expression patterns of corresponding antigens. For example, different B-cell surface expressed proteins can produce slightly different staining in the same disease. A pathologist may not recall and account for that at the time of diagnosis, which can, therefore, complicate the diagnostic process and increase turnaround time. Another problem is related to the fact that many antibodies have been poorly clinically validated[26] and are often selected based on the sensitivity only. Due to the low specificity some of the tests do not have sufficient discriminative power and may lead to highly variable results. These issues are accurately

handled in our algorithm by the antigen expression probabilities across diseases, which rather quantitative task that may not be easily managed by a human. The same applies to the disease incidence rates. It is not a simple task to account accurately for disease incidence rates during diagnosis. Our method provides a solution to these problems, while being flexible enough to allow a pathologist to have full control of the decision process through pathologist-in-the-loop interface.

Our method is very flexible in the sense that it just eliminates diseases that do not support revealed facts. The overall goal of the method is not to suggest a single diagnosis, but to streamline pathology decision-making. For example, in a traditional rule-based expert system, the inference process produces "useful" results only when information about all "key" antecedent terms is entered into the system. If the antecedent part of a rule is only partially satisfied, the rule may not fire and it is hard to extract any benefit from the system in this case. Our method, on the other hand, finds the shortest diagnostic paths using only available facts. It "optimizes" the process by finding the "most efficient" factors at each step (the ones that will lead to ruling out larger number of possible diagnoses). The more such facts are available, the more diseases can be ruled out and the leaf nodes of a corresponding EAS-Tree will have fewer possible diagnostic solutions. A pathologist is free to remove from or add diseases/antigens into consideration and the algorithm will dynamically rebuild the EAS-Tree. This gives the pathologist control over the decision process. Moreover, being involved in an interactive decision-making provides the pathologist with better understanding of why one or another disease should be ruled out or ruled in and with what confidence level. ATOS is not one-answer fit system with predefined rules and answers. Based on the differential diagnosis provided by a pathologist, it builds diagnostic trees that reflect the individual practice patterns of that pathologist. Furthermore, the Bayesian probabilistic model used in our method naturally incorporates disease incidence rates as prior probabilities. ATOS can be easily adjusted to account for specific rates of disease occurrences across different demographical and geographical categories.

We share the view that personalized medicine involves the integration of diagnostics and therapeutics into what became known as theranostics.[27] As such, a more granular subclassification of diseases based on an increasingly complex proteomic knowledge would be expected to contribute significantly to this process. Granted, our current state of practice of personalized medicine in pathology largely seems to be dependent upon the identification of the expression or activation of proteins that may be unique to an individual's particular cellular pathways rather than representative of a population. Therefore, our algorithm could not be expected to provide much assistance. However, as diagnostics evolves to include pathway-based molecular biological evidence,[28] new patterns of protein expression may need to be evaluated in which computational support in the manner of this algorithm may be useful.

CONCLUSION

We have developed ATOS - a novel informatics tool to help pathologists streamline antibody selection process. Using the ATOS-GUI alone, pathologists can significantly speed up antibody selection process. Moreover, our entropy maximization probabilistic algorithm brings up to 40% of an additional decrease in the number of antibody tests required to reach a diagnostic conclusion. Furthermore, it significantly reduces inter- and intrapathologist variability, which makes the process more consistent and predictable. A comparative analysis of our method with the World Health Organization classification guidelines and conventional approaches showed that the proposed tool brings approximately threefold reduction in number of antibody tests required to reach a diagnostic conclusion. Therefore, ATOS has great potential to streamline antibody test selection and significantly decrease associated costs.

ACKNOWLEDGEMENT

Authors thank Dr. Ann Havey, MD and Dr. Magda Esebua, MD experienced pathologists, for their help in evaluation of our method and participating in the experiments. Authors also thank Dr. John Fresen for his constructive comments.

REFERENCES

1. Foucar E. Classification in anatomic pathology. Am J Clin Pathol 2001;116Suppl:S5-20.
2. Brown RE. Morphogenomics and morphoproteomics: A role for anatomic pathology in personalized medicine. Arch Pathol Lab Med 2009;133:568-79.
3. Swerdlow C. International Agency for Research on, and O. World Health, WHO classification of tumours of haematopoietic and lymphoid tissues: France: International Agency for Research on Cancer; 2008.
4. Taylor CR. IHC and the WHO classification of lymphomas: cost effective immunohistochemistry using a deductive reasoning "decision tree" approach. Appl Immunohistochem Mol Morphol 2009;17:366-74.
5. Taylor CR. The WHO classification of lymphomas: cost-effective immunohistochemistry using a deductive reasoning "decision tree" approach: Part II: The decision tree approach: Diffuse patterns of proliferation in lymph nodes. Appl Immunohistochem Mol Morphol 2009;17:470-82.
6. Higgins RA, Blankenship JE, Kinney MC. Application of immunohistochemistry in the diagnosis of non-Hodgkin and Hodgkin lymphoma. Arch Pathol Lab Med 2008;132:441-61.
7. Human Cell Differentiation Molecules. Available from: http://www.hcdm.org/MoleculeInformation/tabid/54/Default.aspx. [Last accessed on 2011 Sep 2].
8. Kishimoto T, Goyert S, Kikutani H, Mason D, Miyasaka M, Moretta L, CD antigens 1996. Blood 1997;89:3502.
9. Zola H, Swart B. The human leucocyte differentiation antigens (HLDA) workshops: The evolving role of antibodies in research, diagnosis and therapy. Cell Res 2005;15:691-4.

10. Zola H, Swart B, Banham A, Barry S, Beare A, Bensussan A, *et al*. CD molecules 2006--human cell differentiation molecules. J Immunol Methods 2007;319:1-5.
11. Elsheikh TM, Asa SL, Chan JK, DeLellis RA, Heffess CS, LiVolsi VA, *et al*. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. Am J Clin Pathol 2008;130:736-44.
12. Friedman BA. A survey of the myriad forces changing anatomic pathology and their consequences. Arch Pathol Lab Med 2008;132:735-8.
13. Jara-Lazaro AR, Thamboo TP, Teh M, Tan PH. Digital pathology: Exploring its applications in diagnostic surgical pathology practice. Pathology 2010;42:512-8.
14. Pantanowitz L. Digital images and the future of digital pathology. J Pathol Inform 2010;1.pii: 15.
15. Feldman MD. Beyond morphology: Whole slide imaging, computer-aided detection, and other techniques. Arch Pathol Lab Med 2008;132:758-63.
16. Heckerman DE, Horvitz EJ, Nathwani BN. Toward normative expert systems: Part I. The Pathfinder project. Methods Inf Med 1992;31:90-105.
17. Vollmer RT. Differential diagnosis in immunohistochemistry with Bayes theorem. Am J Clin Pathol 2009;131:723-30.
18. Vollmer RT. Primary lung cancer vs metastatic breast cancer: A probabilistic approach. Am J Clin Pathol 2009;132:391-5.
19. SEER Cancer Statistics Review. Available from: http://www.seer.cancer.gov/csr/1975_2006/. [Last accessed on 2006]
20. Shannon C. A Mathematical Theory of Communication. CSLI Publications; 1948.
21. Dunham MH. Data mining introductory and advanced topics. Upper Saddle River, N.J: Prentice Hall/Pearson Education; 2003.
22. Consortium TU. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 2010;38:D142-8.
23. Huss JW 3rd, Orozco C, Goodale J, Wu C, Batalov S, Vickers TJ, *et al*. A gene wiki for community annotation of gene function. PLoS Biol 2008;6:e175.
24. Foucar E. Diagnostic decision-making in anatomic pathology. Am J Clin Pathol 2001;116 Suppl:S21-33.
25. Pena GP, Andrade-Filho Jde S. How does a pathologist make a diagnosis? Arch Pathol Lab Med 2009;133:124-32.
26. Kapke G, Stoddard JJ. Biomarkers - boon or bane for researchers? Good Clin Pract J 2008. p. 27-31.
27. Warner S. Diagnostics + therapy = theranostics: Strategy requires teamwork, partnering, and tricky regulatory maneuvering. New York: The Scientist; 2004.
28. Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, A pathway based classification of human breast cancer. Proc Natl Acad Sci U S A 2010;107:6994-9.