01 Jul 2001

Abnormal Cell Detection using the Choquet Integral

R. Joe Stanley
Missouri University of Science and Technology, stanleyj@mst.edu

James M. Keller

Charles William Caldwell

Paul D. Gader

Recommended Citation

R. J. Stanley et al., "Abnormal Cell Detection using the Choquet Integral," *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference (2001, Vancouver, British Columbia, Canada)*, vol. 2, pp. 1134-1139, Institute of Electrical and Electronics Engineers (IEEE), Jul 2001.
The definitive version is available at https://doi.org/10.1109/NAFIPS.2001.944764

Abnormal Cell Detection Using the Choquet Integral

Ronald Stanley	James Keller	Charles W. Caldwell	Paul Gader
Department of Electrical and Computer Engineering University of Missouri-Rolla	Department of Computer Engineering and Computer Science University of Missouri-Columbia	Department of Pathology and Anatomical Sciences University of Missouri-Columbia	Department of Computer Engineering and Computer Science University of Missouri-Columbia
Rolla, MO U.S.	Columbia, MO U.S.	Columbia, MO U.S.	Columbia, MO U.S.
stanleyr@umr.edu	keller@cecs.missouri.edu	CaldwellC@.missouri.edu	gader@cecs.missouri.edu

Abstract

Automated Giemsa-banded chromosome image research has been largely restricted to classification schemes associated with isolated chromosomes within metaphase spreads. In normal human metaphase spreads, there are 46 chromosomes occurring in homologous pairs for the autosomal classes, 1-22, and X chromosome for females. Many genetic abnormalities are directly linked to structural and/or numerical aberrations of chromosomes within metaphase spreads.

Cells with the Philadelphia chromosome contain an abnormal chromosome for class 9 and for class 22, leaving a single normal chromosome for each class. A data-driven homologue matching technique is applied to recognizing normal chromosomes from classes 9 and 22. Homologue matching integrates neural networks, dynamic programming and the Choquet integral for chromosome recognition. The inability to locate matching homologous pairs for classes 9 and 22 provides an indication that the cell is abnormal, potentially containing the Philadelphia chromosome. Applying this technique to 50 normal and to 48 abnormal cells containing the Philadelphia chromosome yields 100.0% correct abnormal cell detection with a 24.0% false positive rate.

1. Introduction

Automated Giemsa-banded chromosome image research has been largely restricted to classification schemes associated with isolated chromosomes within metaphase spreads. Many existing approaches for performing automated chromosome image analysis presuppose a fixed number of chromosomes per class, two, and 46 chromosomes within a metaphase spread for achieving better classification [1,2,3], which is true for normal cells. Many genetic abnormalities are directly linked to structural and numerical anomalies of chromosomes within the metaphase spread. An example of a numerical anomaly is Down's Syndrome, where individual cells contain three chromosome 21s.

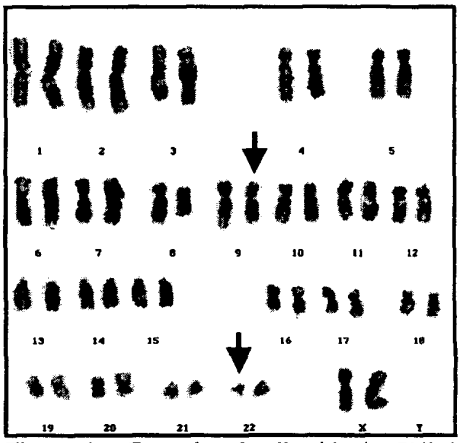

Figure 1: Example of cell with the Philadelphia chromosome. The arrows point to chromosomes 9 and 22 involved in the reciprocal translocation.

This research focuses on a specific structural anomaly referred to as the Philadelphia chromosome. The Philadelphia chromosome is the remaining piece of a chromosome 22 that results from a reciprocal translocation between a chromosome 22 and a chromosome 9, described by t(9;22). The reciprocal translocation is the exchange of a specific portion from the chromosome 9 with a specific portion of the chromosome 22 [4,5]. The Philadelphia chromosome was first reported in 1960 and is characteristic of several types of leukemia [4,5]. Figure 1 provides an image example of the Philadelphia chromosome, where all chromosomes within a cell are paired with their homologues. The arrows in Figure 1 point to the distorted chromosomes 9 and 22 that resulted from the reciprocal translocation. Inspecting Figure 1, the homologues for classes 9 and 22 do not have the same

degree of correspondence as homologues for the other classes.

In this research homologue matching is applied to detect abnormal cells potentially containing t(9;22) associated with the Philadelphia chromosome. For the experiments performed in this study, the Philadelphia chromosome is present in all abnormal cells. In actual practice, the Philadelphia chromosome is not necessarily present in every cell because some residual normal cells may be present. The homologue matching method identifies chromosomes from selected classes based on the similarity between homologues and uses the chromosome recognition results for the corresponding classification of cells. The method capitalizes on the principle of least commitment [6] and avoids the two chromosome per class assumption.

Chromosome identification integrates neural networks, banding pattern and centromeric index criteria checking, homologue matching and information fusion. For the selected class, the best representative or primary chromosome is found within the metaphase spread. Homologue candidates are obtained using simple criteria. The candidates are matched to the primary chromosome for homologue determination using the Choquet integral to fuse multiple, similar primary chromosome to homologue match approaches into one match score. The homologue found is rematched using the same process. With the purpose of aiding a cytogenetic expert, making no decision for chromosome assignment is better than an incorrect assignment. Experimental results are presented applying an extension of this approach to the identification of abnormal cells potentially containing the Philadelphia chromosome.

2. Method and Materials

2.1 Homologue Matching Algorithm

The homologue matching technique employed in this research is an extension of prior applications of identifying isolated and overlapped chromosomes from selected classes [7,8]. The basic homologue matching algorithm is as follows. The feature extraction process utilizes three program inputs: 1) the original metaphase spread image, 2) the segmented image of the metaphase spread image, and 3) the skeletons determined from the segmented image.

Following segmentation and skeletonization, feature extraction is performed for each isolated chromosome found within the metaphase spread image. The features computed and used for analysis include: 1) chromosome size (length and area), 2) centromeric index, 3) polarity, 4) band features including total number of bands, p-arm bands, and q-arm bands, 5) density, shape, binary band, and width profiles along the medial axis based on orthogonal lines to that axis, and 6) weighted density distribution (WDD) function values [2,9] from the density and shape profiles. Definitions and algorithms for implementation are extensively described in [7,10].

The WDD features computed for each isolated chromosome are linearly scaled and input to a feed forward neural network. The weights used in the neural network for assigning confidence values to chromosomes within the metaphase spread are obtained using a standard back propagation neural network, training 45 G-banded bone marrow chromosomes per class from the University of Missouri chromosome image library. In addition, the band features and centromeric index are computed over the same training data, providing max/min ranges for the total number of bands, p-arm bands, q-arm bands, and centromeric index for the selected class.

Confidence values for each chromosome belonging to the class of interest are determined from the feed forward neural network. Chromosomes with greater than zero confidence in the selected class are the initial candidates chosen. Candidates with band features or centromeric indices outside the max/min range found for the selected class (classes 9 and 22 in this research) are eliminated from consideration. From the remaining candidates, the chromosome with the greatest margin of victory in neural network confidence is chosen as the primary chromosome. If no candidates remain, no chromosome is assigned to that class for the cell under consideration.
Determining a primary chromosome and a set of candidate chromosomes, the remaining candidates are automatically inspected to determine the matching homologue using dynamic programming. Density, shape, and binary band profiles are matched between the primary chromosome and each of the candidates using the computed chromosome polarity or orientation to coordinate the matching process. Based on the dynamic programming match scores for all candidates, a confidence distribution is formed for each profile. A confidence distribution is also formed from a distance measure relating to the chromosome scaled length, scaled area, and scaled centromeric index.

Final confidence values for each candidate for each profile-based classifier are computed as the product of the dynamic programming confidence value and the distance measure confidence value. The final

confidence values from the three profile-based classifiers are combined using the Choquet integral [16,17]. The candidate with the highest final confidence value from the Choquet integral is deemed homologue to the primary chromosome. Upon homologue determination, the homologue is rematched to find its closest fit. The homologue rematching process is the same as the primary chromosome matching process.

2.2 Choquet Integral

In the context of this research, the Choquet integral was used to combine the final confidence value for each candidate from the density, shape, and binary band profile-based classifiers. The Choquet integral provided information fusion accounting for the credibility of the information source and the worth of the information from the source. Fuzzy integrals, including the Choquet integral, are non-linear functions defined with respect to a fuzzy measure [11].

The concept of a fuzzy measure, denoted as g, is applied to this research as follows: 1) $X = \{x_1, x_2, x_3\}$ which represented the density profile based classifier, shape profile based classifier, and binary band profile based classifier, respectively, 2) $g(\{x_1\})$ was chosen as 0.4, $g(x_2)$ and $g(x_3)$ are computed based on homologue matching results using the density, shape, and binary profile-based classifiers as independent classifiers for assigning chromosome 17 within normal metaphase spreads [7], and 3) λ is computed using the expressions above and, then, the measure g is generated. The mass values in step 2 are calculated by subtracting the number of cells with wrong chromosome assignments from the total number of cells with correct chromosome assignments and dividing the difference by the total number of cells used in assigning chromosome 17. Fifty normal metaphase spreads were used for the chromosome 17 analysis. The ratio of 0.4 to the mass value for the density profile is used to scale the binary band and shape profile confidence values. The mass values for the various combinations of the classifiers and λ are determined using the union expression above.

Based on the defined λ-fuzzy measure, the Choquet integral [11] is used to fuse the final confidence scores from the density, shape, and binary band profile-based classifiers. The Choquet integral, as applied to this research, can be expressed as:

$$e = \sum_{i=1}^{3} [h(x_{(i)}) - h(x_{(i-1)})] g(A_i)$$

with $x_{(1)}$ the profile-based classifier with lowest final confidence value, $x_{(2)}$ the profile-based classifier with middle final confidence value, $x_{(3)}$ the profile-based classifier with highest confidence value, $h(x_{(i)})$ $i \in \{1,2,3\}$ the confidence value corresponding to profile-based classifier, $h(x_{(1)}) \leq h(x_{(2)}) \leq h(x_{(3)})$, $h(x_{(0)}) = 0$, and $A_i = \{x_i, \ldots, x_3\}$.

2.3 Experiments Performed

The first set of experiments examines the homologue matcher's ability to label cells as normal when the cells are known to be normal. The experimental procedure is: 1) apply the homologue matcher to the normal 50 metaphase spreads for assigning chromosomes to classes 9 and 22 and 2) combine the results for classes 9 and 22 using the "AND" approach for determining cell normalcy.

The second set of experiments determines cell normalcy for metaphase spreads known to contain the Philadelphia chromosome. The experimental procedure is: 1) apply the homologue matcher to the 48 abnormal metaphase spreads for assigning chromosomes to classes 9 and 22 and 2) combine the homologue matching results for classes 9 and 22 using an "AND" approach for cell normalcy evaluation. Note that some chromosome clusters are manually separated in 20 of the 48 metaphase spreads containing the Philadelphia chromosome. The automatic segmentation algorithm is applied to the resulting images. The homologue matching approach does not improve on segmentation, and is not intended to do this. Instead, it is meant to automatically detect abnormal cells based on the failure to match specific homologues.

2.4 Cell Normalcy Determination Procedure

The homologue matching algorithm is applied to the detection of abnormal cells potentially containing the Philadelphia chromosome. Specifically, the homologue matching algorithm is applied to assigning chromosomes from class 9 and from class 22 within metaphase spread images. In normal metaphase spread images, matching homologues exist for classes 9 and 22. In scoring the metaphase spreads tested, a 'y' is obtained when the primary chromosome is properly determined, the correct homologue is found, and the homologue matched to the primary chromosome. An 'i' is assigned when only the correct primary chromosome is found, and the homologue matched to a chromosome other than the primary. An 'm' is obtained when at least one chromosome is incorrectly assigned. In abnormal metaphase spread images containing the Philadelphia chromosome, matching homologues do not exist for classes 9 and 22. In scoring the abnormal metaphase spreads tested, an 'i' is assigned to reflect that a single normal chromosome is present for class 9 and for class 22.

An 'm' is assigned to show that matching homologues are found in the cell when matching homologues do not exist in the cell.

An "AND" approach is used to fuse the class 9 and 22 results for determining if cells are normal. Cells are flagged as abnormal when classes 9 and 22 have indeterminate assignments, i.e. classes 9 and 22 have an 'i'. Otherwise, the cells are labeled normal. The evaluation rules are adjusted for the situation that no primary chromosome is found for either class 9 or class 22 (or both) within a metaphase spread. There are two cases where no primary chromosome 9 is found. First, if the homologue matcher finds only one chromosome 22 and the top two neural network winners do not agree with the homologue matcher, the cell is labeled abnormal. Second, if the homologue matcher found two chromosome 22's and they are the top neural network winners, the cell is labeled normal. The same rules are applied if no primary chromosome 22 is found. If no primary chromosome is found for class 9 or class 22, the cell is labeled abnormal. The additional constraints provide cross validation for evaluating cell normalcy. Experiments are performed using 50 normal metaphase spread images and 48 abnormal metaphase spread images containing the Philadelphia chromosome from the University of Missouri-Columbia chromosome image library and from the Cytogenetic Laboratory at The University of Texas M. D. Anderson Cancer Center.

3. Results

The experimental results for applying the homologue matcher to classes 9 and 22 and for determining cell normalcy are as follows. The normal metaphase spread results are shown in Table 1. Table 1 contains the experimental results for 10 normal metaphases with the composite results over the same 50 normal metaphase spread images from applying the homologue matcher to classes 9 and 22 and the combinatorial results using the "AND" approach for cell normalcy evaluation. The results shown for the 5 cells in Table 1 are representative of the results for the 50 cells tested.

The abnormal metaphase spread results are shown in Table 2. Table 2 presents the experimental results for the homologue matching algorithm and the corresponding cell normalcy description for 5 cells with the composite results over the 48 metaphase spreads tested. The evaluation rules are adjusted for the situation that no primary chromosome is found for either class 9 or class 22 (or both) within a metaphase spread.

Individual Cell Results			
Cell	**Class 9**	**Class 22**	**Cell Id**
1	y	y	Correct
2	y	y	Correct
3	i	y	Correct
4	y	m	Correct
5	y	y	Correct

Composite Results			
	Class 9	**Class 22**	**Cell Id**
Correct	36	41	49
Indeterminate	12	8	
Incorrect	2	1	1

Table 1: Homologue matching and cell normalcy evaluation for 50 known normal metaphase spread images. Results are shown for 5 cells. Composite results are presented for the 50 metaphase spreads.

Table 1 Key:
Column 2 contains the chromosome 9 homologue matching results. Column 3 has the chromosome 22 homologue matching results. Column 4 contains the cell identification results.
Key for columns 2 and 3:
y: correct homologues found
i: no matching homologues found
m: incorrect homologues found

The rule adjustments utilize the top two neural network winners and the "AND" combination results from the homologue matcher. If the homologue matcher finds only one chromosome 22 and the top two neural network winners do not agree with the homologue matcher, the cell is labeled abnormal. Also, if the homologue matcher found two chromosome 22's and they are the top neural network winners, the cell is labeled normal. The same rules are applied if no primary chromosome 22 is found. If no primary chromosome is found for class 9 or class 22, the cell is labeled abnormal. The additional constraints provide cross validation for evaluating cell normalcy.

There are two cases where no primary chromosome 9 is found. The relatively poor quality of the metaphase spread images and small number of chromosomes used for neural network training may be contributing factors to the inability to identify primary chromosomes in those cells.

Individual Cell Results			
Cell	**Class 9**	**Class 22**	**Cell Id**
1	i	i	Correct
2*	i	i	Correct
3*	m	i	Incorrect
4	m	m	Incorrect
5	i	i	Correct

Composite Results			
	Class 9	**Class 22**	**Cell Id**
Indeterminate	30	37	34
Incorrect	8	10	14

Table 2: Homologue matching and cell normalcy identification results for 48 known abnormal metaphase spreads containing the Philadelphia chromosome.

Table 2 Key:

Results are shown for 5 cells. Composite results are presented for the 50 metaphase spreads. Cell numbers with * are manually segmented.

Key for columns 2 and 3:

i: no matching homologues found (one chromosome assigned)

m: matching homologues found (two chromosomes assigned but cell only has one normal chromosome)

4. Discussion

The similarity between homologues is important for evaluating chromosomes within the context of a given cell. If the homologue found and the primary chromosome match each other, there is high similarity between the homologues. Both chromosomes are assigned to the selected class. If the homologue found does not match the primary chromosome, there is low similarity between the homologues. Only the primary chromosome is assigned to the selected class. In normal cell analysis, chromosomes occur in homologous pairs. Only assigning the primary chromosome to the selected class, an 'i' in Table 1, is a recoverable error. The homologue matcher makes no decision in assigning the second chromosome to the selected class in these situations. For the abnormal cells examined in this research, one of homologues for classes 9 and 22 are distorted. Identifying a single chromosome for classes 9 and 22 provides a lack of correspondence between homologues, which is characteristic of normal cells. Thus, the homologue matching approach identifies abnormal cells based on the lack of correspondence between homologues, not by identifying the actual abnormality. A false positive chromosome assignment to class 9 and/or class 22 results in an incorrect diagnosis of the patient. Making no chromosome assignment leads to a false positive diagnosis, not a false positive chromosome classification. Further analysis can be used to overcome false positive diagnoses for determining that a patient does not have a specific condition. It is assumed with the homologue matching approach that support expertise is available to overcome false positive diagnostic cases but is unavailable to overturn normal interpretations. Making an incorrect chromosome assignment is an unrecoverable error, an 'm' in Table 1. The "AND" approach yields a better overall cell identification rate (98.0%) than either class 9 or 22 individually.

The goal for abnormality analysis is to maximize labeling cells as anomalous when they are anomalous (true negative rate) and to concurrently minimize labeling cells as normal when they are abnormal (false negative rate). True positive, false positive, false negative, and true negative cell assessments represent the following: 1) true positive: correctly identifying normal cells as normal, 2) false positive: incorrectly labeling normal cells as abnormal, 3) false negative: incorrectly identifying abnormal cells as normal, 4) true negative: correctly labeling abnormal cells as abnormal. Table 3 breaks down the cell normalcy evaluation results using the "AND" approach to present the true positive, false positive, false negative, and true negative rates for the normal and abnormal cells analyzed.

Cell Type	% Identified as Normal	% Identified as Abnormal
Normal	98.0 (TP)	2.0 (FP)
Abnormal	25.0 (FN)	75.0 (TN)

Table 3: Breakdown of recognition of 50 normal metaphase spreads and 48 abnormal metaphase spreads for the "AND" approach for evaluating cell normalcy.

Table key:

TP:true positive rate FP:false positive rate

FN:false negative rate TN:true negative rate

The homologue (as well as the homologue rematch) is the candidate with the highest Choquet integral confidence value. This constraint for homologue selection is tightened in order to improve the false negative rate. Specifically, the number of homologues and rematched homologues with confidence values exceeding the winning margin diminishes as the winning margin increases. Thus, the number of indeterminate cases increases as the winning margin requirement increases. The "AND" method requires chromosome classes 9 and 22 to have indeterminate matches for the cell to be labeled abnormal. Figure 2

presents the receiver operator characteristic (ROC) curve showing the relationship between the true negative rate (vertical axis) and the false positive rate (horizontal axis) for cell normalcy. The plotted points shown in Figure 2 represent 0.01 incremental increases in the winning margin starting from zero. Note that the plotted points represent one or more 0.01 increments. As seen from Figure 2, The "AND" approach flags 100.0% of the cells with t(9;22) by raising the winning margin to 0.20. Correctly identifying 100.0% of the abnormalities results in an increase in the false positive rate to 24.0%.

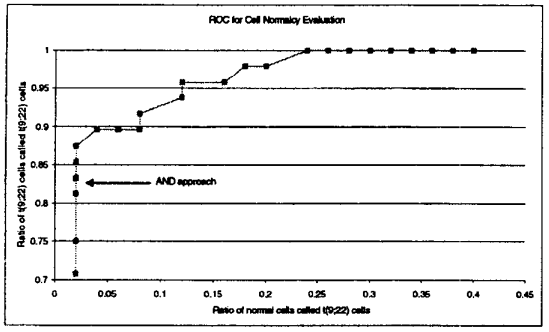

Figure 2: ROC curve showing the relationship between true negative rate to false positive rate for cell normalcy. From left to right, each plotted point represents an incremental increase of 0.01 in the winning margin beginning with 0.

5. Conclusion

In this research a novel homologue matching approach is introduced for intelligent flagging of abnormal cells containing the Philadelphia chromosome using metaphase spread image analysis. Specifically, the algorithm is successfully applied to flagging anomalous cells containing a distorted chromosome 9 and a distorted chromosome 22, where the anomalous chromosomes yield the Philadelphia chromosome.

Future research will focus on detecting cells with different types of chromosome aberrations and will attempt to address the following questions: 1) Is it possible to detect which regions between two homologues differ? 2) Is the size of the translocation and/or deletion that can be detected dependent on the size of the chromosome itself? Answering these questions will provide insight to the overall effectiveness of the homologue matching technique for abnormality analysis.

6. References

[1] C. Lundsteen, A.M. Lind, E. Granum, "Visual classification of banded human chromosomes i. Karyotyping compared with classification of isolated chromosomes," *Annals of Human Genetics*, Vol. 40, 1976, pp. 87-97.

[2] J. Piper, E. Granum, "On fully automatic feature measurement for banded chromosome classification," *Cytometry*, Vol. 10, 1989, pp. 242-255.

[3] J. Graham, J. Piper, "Automated karyotype analysis," In Gosden JR (ed): *Methods in Molecular Biology: Chromosome Analysis Protocols*. Totowa, NJ, Humana Press, Inc, 1994, Vol. 29.

[4] J.D. Rowley, "A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining," *Nature*, Vol. 243, 1973, pp. 290-293.

[5] S. Heim, F. Mitelman, *Cancer Cytogenetics*, Second Edition, Wiley-Liss, New York, 1995.

[6] J. Keller, Gader P, Caldwell CW, "The principle of least commitment in the analysis of chromosome images," *Proceedings of the SPIE Symposium on OE/Aerospace Sensing and Dual Use Photonics*, Orlando, FL, 1995, pp. 178-186.

[7] R.J. Stanley, J. Keller, P. Gader, C.W. Caldwell, "Data-driven homologue matching for chromosome identification," *IEEE Trans. Medical Imaging*, Vol. 17, No. 3, 1998, pp. 451-462.

[8] R.J. Stanley, J. Keller, P. Gader, C.W. Caldwell, "Homologue matching applications: recognition of overlapped chromosomes," *Pattern Analysis and Applications*, Vol. 1, 1998, pp. 206-217.

[9] E. Granum, T. Gerdes, C. Lundsteen, "Simple weighted density distributions, WDDs for discrimination between G-banded chromosomes," *Proceedings of the Fourth European Chromosome Analysis Workshop*, Edinburgh, 1981.

[10] R.J. Stanley, J. Keller, C.W. Caldwell, P. Gader, "Centromere attribute integration based chromosome polarity assignment," *JAMIA*, Vol. 3 (supplement), 1996, pp. 284-288.

[11] M. Grabisch, H.T. Nguyen, E.A. Walker, *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Kluwer Academic Publishers, Boston, 1995.

Acknowledgment: Special thank you to Dr. Armand Glassman at the Cytogenetic Laboratory at The University of Texas M. D. Anderson Cancer Center for providing Philadelphia chromosome data.