Second International Conference on Intelligent Computing and Information Systems
March 5-7, 2005. Cairo, Egypt.

FEATURE ANALYSIS AND CLASSIFICATION OF THE HUMAN CHROMOSOME IMAGES

Mohammed Alhanjouri[1], Nadder Hamdy[1], and Fatma E.Z. Abou-Chadi[2]

[1]Dept. of Elec. & Comm. Engineering, Arab Academy for Science & Technology, Alexandria, Egypt
[2]Department of Electronics. & Communications Engineering, Mansoura University, Mansoura, Egypt
Tel. +203-5489586 and
email: Alhanjouri@aast.edu

Abstract

In this work, an automatic system has been developed to perform chromosome feature analysis and classification and to find out the existence of possible chromosomal aberrations. It performs chromosome analysis in three steps: metaphase image pre-processing, object analysis and chromosome classification. A neural network classifier was used to achieve a high performance of 100%.

Keywords: *Image Processing, Pattern Recognition, Neural Network, Majority Filter, Chromosome classification*

I. Introduction

Chromosome imaging is a valuable tool for doctors and cytogenetic technicians. Extra chromosomes, missing chromosomes, broken chromosomes, and translocations (parts of chromosomes breaking off and attaching to other chromosomes) are indicators of radiation damage, cancer, and a wide variety of inherited diseases [1],[2].

Traditional chromosome imaging has been limited to grayscale images, but recently a 5-fluorophore combinatorial labeling technique was developed in which each class chromosomes binds with a different combination of fluorophores, this results in a multi-spectral image, in which each class of chromosomes has distinct spectral components (each chromosome type appears to have a distinct color). This multi-spectral staining technique made analysis of chromosome images easier, not only for visual inspection of the images by humans, but also for computer analysis of the images. The multi-spectral staining technique is called M-FISH (multiplex fluorescence in-situ hybridization.) M-FISH uses five color dyes that attach to various chromosomes differently to produce a multi-spectral image, and a sixth dye that attaches to all chromosomes to produce a grayscale image. Although M-FISH presents significantly more information than was available in traditional grayscale images, little research on multi-spectral chromosome image analysis has been previously reported in the open literature [3-5].

The technique of multicolor fluorescence in situ hybridization (M-FISH) involves generated chromosome specific probes that are labeled differentially with various fluorescent dyes in different proportions. The combination of dyes (DEAC "Diethylamino-Coumarin", FITC "Fluorescein Isothiocyanate", Spectrum Orange, Texas Red and Cy5 "Cyan") enables the detection of all the chromosomes simultaneously with a single hybridization procedure. The separation of different excitation and emission spectra is achieved by 100 mercury lamp and appropriate filter sets (DAPI "4',6-Diamidino-2-phenylindole", DEAC, FITC, Spectrum Orange, Texas Red and Cy5). The resulting un-equivocal color signature for each chromosome enables the analysis of hidden or complex chromosome aberrations. M-FISH requires 3-4 days for hybridization and the analysis of a single metaphase takes about 30 min. Microscopic analysis was performed using an Axioplan II imaging microscope (Zeiss, Germany) [13].

The introduction of in situ hybridization techniques has added significantly to the description of the human karyotype by defining chromosome regions at the molecular level. Recently, in order to simplify chromosome identification, various strategies have been proposed for adding color information to the classical gray-scale description of chromosome banding. Whole chromosome painting can define the gross DNA "Deoxyribonucleic acid" content of an entire chromosome. This method has also been performed in a multi-color format by labeling probes with different hap-tens in a simple Boolean fashion or by including ratios of differentially labeled probes. Recent studies have successfully demonstrated that each member of the entire chromosome set can be simultaneously distinguished by using different combinations of fluorescent labels. An alternative approach has been used to label each chromo-some by sub-regional DNA probes in different colors. This pattern has been named the "chromosome bar code" and has simplified chromosome identification by producing a limited number of bars on each chromosome. Most karyotype analyses, however, still largely rely on the correct interpretation of classical banding patterns, since the new strategies only allow the identification of whole chromosomes without further differentiation of chromosomal sub-regions. In a different approach, FISH with sub-regional probes has been applied to generate a multi-color banding pattern on a single chromosome pair, so that intra-chromosomal rearrangements can be detected.

Under normal circumstances, chromosomes are extremely long and thin and are essentially invisible. However, during the metaphase stage of mitosis (cell division), they contract and become much shorter (around 2-10 μm) and wider

(around 1-2 μm diameter). At this stage, they can be stained to become visible and can be imaged by a microscope (Fig. (1)).

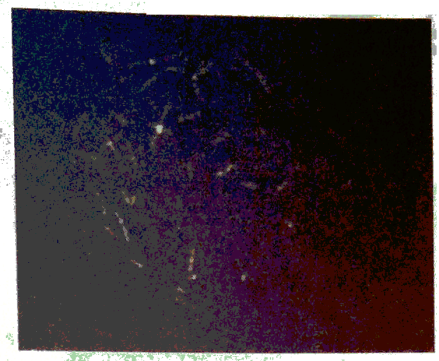


Fig. (1): Typical chromosome image

The cells for producing these images are commonly obtained from blood specimens, bone marrow, and amniotic fluid. In humans, the 46 chromosomes consist of 23 pairs of chromosomes, one of each pair coming from the father and the other from the mother. Of the 46 chromosomes, there are 22 homologous pairs and two sex chromosomes denoted X and Y (Fig. (2)). A normal human female has two X chromosomes, while a normal male has an X and a Y chromosome. By convention, the 22 pairs and the X chromosome and Y chromosome are assigned to 24 distinct classes [1],[6].

In this paper, multi-spectral images (produced by a modern imaging system) are used for training and testing phases, currently, these images aren't available at most the specialized laboratories, therefore, the papers in the references list used the gray-level images as available images.

Our database is taken from "Courtesy advanced imaging research, LLC, League city, Texas 77573", that divided to two sets (400 cases and 300 cases) as a training set and testing set, respectively.

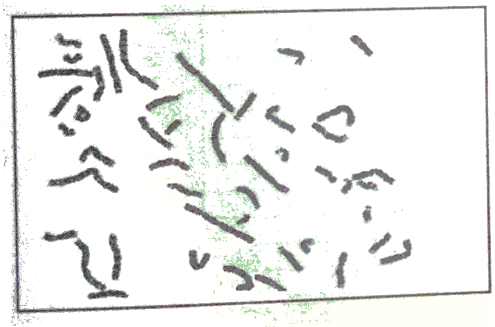
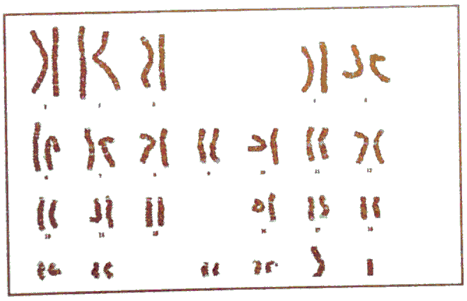


Fig. (2) (a) A chromosome spread image



(b) Karyotype of chromosomes in image (a)

II. Method and Results

A. Pre-processing

The first step is the background cancellation. This is based on adaptive thresholding the gray-level image of biological objects on a microscope slide [6-9] as shown in Fig. (3). Valley searching attempts to find a valley of gray values that represent a separation between two chromosomes, this method often works well for finding accurate boundaries, but it also does not handle overlaps

Fig. (3) Background / Foreground separation

B. Chromosome analysis and classification

Pixel-by-pixel can be classified from the spectra of foreground for gray-level image using green, orange, red, cyan spectra according to Table.1. This table is produced by applying separated adaptive threshold for each spectral image. Selection of the appropriate adaptive threshold is a complex task to obtain best results; therefore, a multiplayer feedforward neural network (NN) was used to classify each pixel. Table.2. compares between the adaptive threshold technique and the Neural network technique.

Table (1) Sizes and spectra of chromosomes (Quoted from [3-5])

Chromosome Class	Relative Size of chromosome	Spectrum Green	Spectrum Orange	Texas Red	Cy5	Cy5.5
1	0.097359	X		X	X	
2	0.084825					X
3	0.073803	X	X		X	X
4	0.065074	X			X	
5	0.063339	X	X	X		X
6	0.068543	X		X	X	X
7	0.047001			X	X	
8	0.046777	X				
9	0.040175	X	X			X
10	0.041182				X	X
11	0.044427	X	X		X	
12	0.036985			X		X
13	0.037936	X	X			
14	0.036594			X		
15	0.02932		X	X	X	
16	0.025347	X		X		
17	0.021822				X	
18	0.025347	X	X	X		
19	0.019248		X		X	
20	0.017066		X			
21	0.017793	X				X
22	0.015555		X	X	X	X
X	0.033964		X			X
Y	0.010519	X			X	X
	=1	Summation of areas				

Table (2) Comparison the threshold technique and Neural network technique

	Threshold technique	Neural network technique
Testing Time	15 minute	12 minute
Recognition rate	95%	100%

C. Neural network topology

Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.
The neural network consists of three layers. The input layer has five neurons corresponding to the five spectra. The hidden layer consists of 25 neurons and the output layer consists of 25 neurons corresponding to 24 chromosomes and the background of the image.
Before training a feedforward network, the weights and biases must be initialized by random values between -1 and 1. Using MATLAB6.5, the network was automatically initialized with the default parameters.

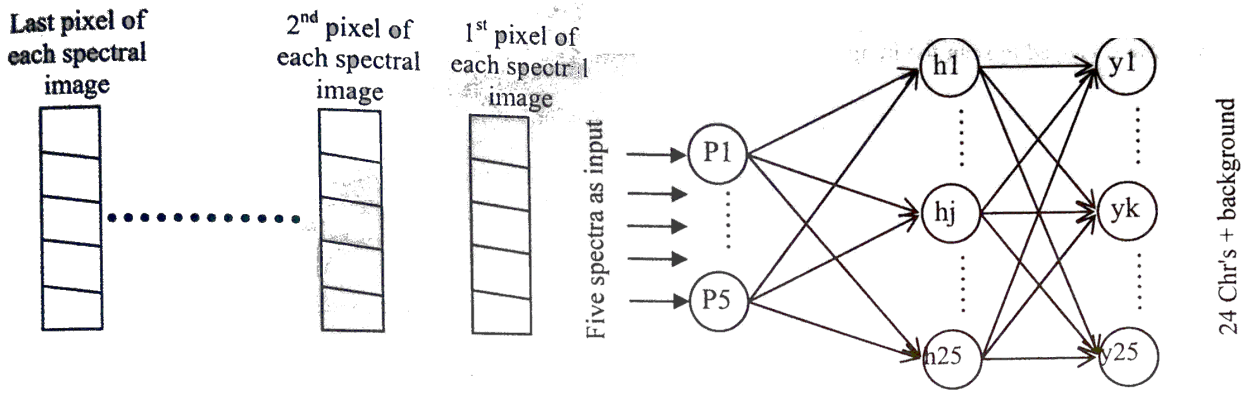


Fig. (4) Multilayer feedforward neural network topology

Once the network weights and biases have been initialized, the network is ready for training. The network can be trained for function approximation (nonlinear regression), pattern association, or pattern classification. The training process requires a set of examples of proper network behavior - network inputs p and target outputs t. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function. The default performance function for feedforward networks is mean square error (MSE) -the average squared error between the network outputs and the target outputs [10-12]. Figure (5) illustrate optimum number of hidden nodes.

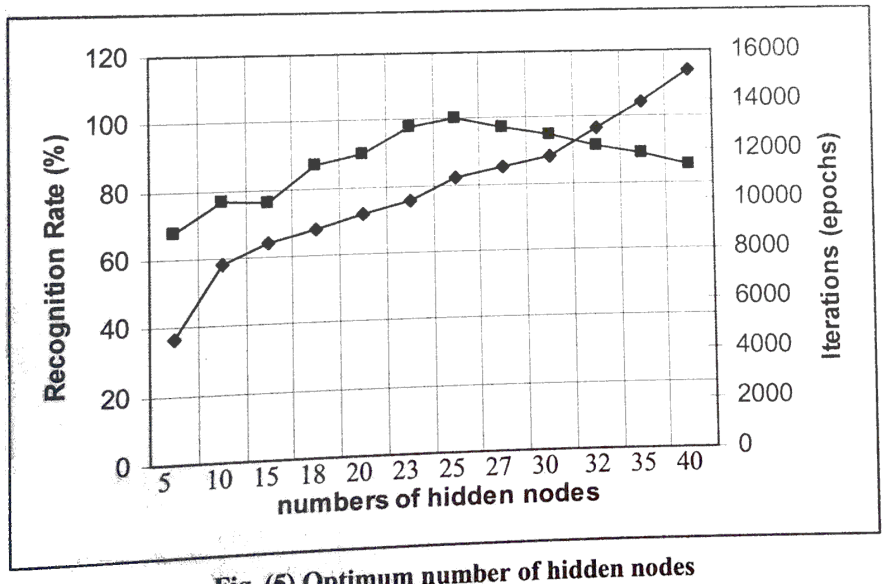


Fig. (5) Optimum number of hidden nodes

D. Denoising

Since pixel classification is an inherently noisy process, some isolated pixels and small segments would be misclassified in this step. To reduce the effect of this noise, a filtering approach must be utilized using a majority filtering approach. The majority filter [3],[4] was chosen because it removes small segments and maintains the shape and position of large scale edges, A majority filter consists of a structuring element H and the image is scanned in raster order, and the class at the center pixel location is replaced by the majority class within spatial extent of the structuring element H, mathematically,

$$y(m)=\underset{k\in H,(m-k)\in\{Oi\}}{maj}\{x(m-k)\} \quad (1)$$

Where x is the input pixel map, y is the output pixel, and *maj* denotes the majority operation, notice that only object pixels are used for calculating the majority, not background pixels $\{Oi\}$, H is the structuring element, for example, a 3×3 square window is defined as

$$H=\{(-1,-1),(-1,0),(-1,1),....(1,0),(1,1)\}$$

A 3×3 window was used to scan the image and to replace a center pixel with the majority values (Fig.(6))

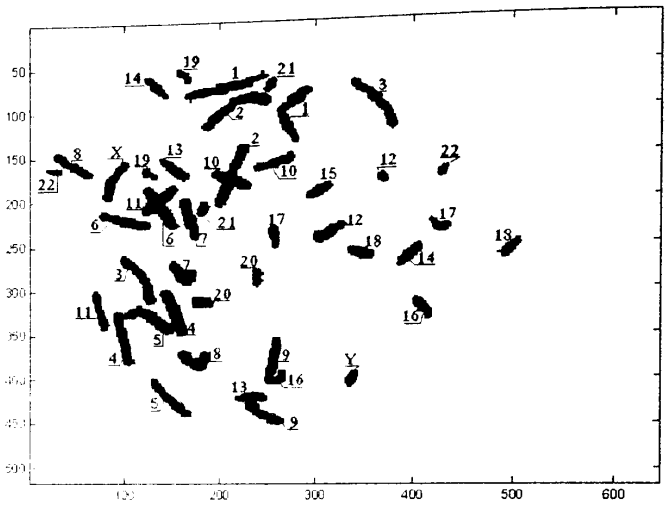


Fig. (6) Classified Image

E. Segmentation

From previous steps, each pixel is now classified as a point in a certain chromosome, therefore each chromosome can be segmented by collecting all the pixels belonging to this chromosome, and again, it was scanned by a 3×3 window majority filter. Fig. (7) Shows the segmented Chromosome No. 1, however, it is difficult to segment Chromosome No. 2 because of the overlapping of more than chromosome as shown in Fig. (8.a) These chromosomes must be separated. Chromosomes that are overlapped will loose some parts of its area.

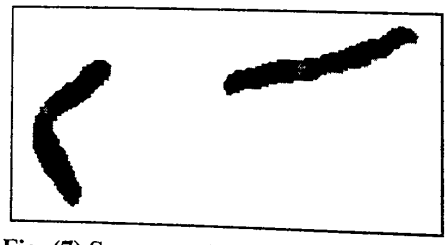


Fig. (7) Segmentation of Chromosome No. 1

To solve the overlapping problem, the medial axis transform can be used to find the medial axis of the cluster by measuring the width along a transverse line perpendicular to the tangent of the edges as shown in Fig. (7b). The upper chromosome can be segmented by extracting the pixels belonging to it. Fig. (8c) shows the lower chromosome alone, but not complete, therefore, completing the lower chromosome must be done by connecting the medial axis and adding parallel lines to medial axis and repeating until filling all the space between the edges of the chromosome as shown in Fig.(8d).

Finally, karyotype of chromosome was done by calculation of edges to extract each chromosome as shown in Fig. (9). From the karyotype image, it can be noted that the image sample was taken from a male subject because of the existence of one chromosome X and one chromosome Y. Moreover, Chromosome 15 is abnormal because one of the chromosome pair is missing. In addition, a part from one of the chromosome 12 is missing. Therefore, this can be considered as an abnormal case comparative with a normal case as shown in figure (2b).

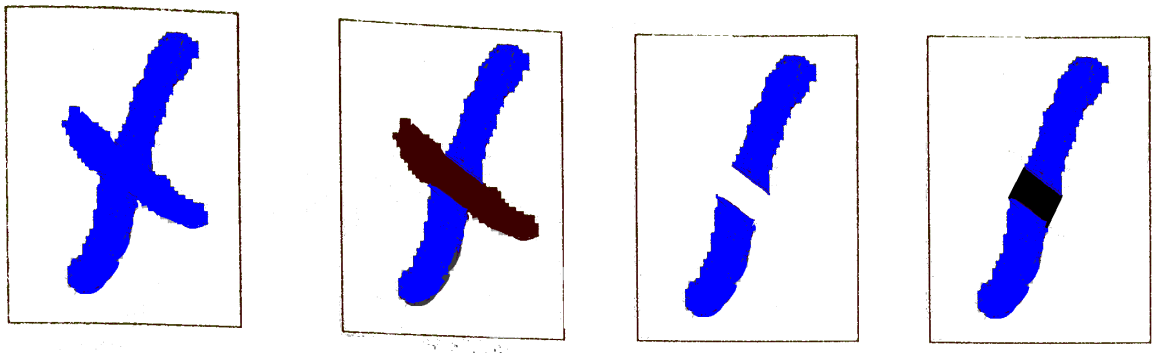


Fig. (8) Separation of overlapped chromosomes

(a) Overlapped chromosomes (b) Medial axis
(c) Separated upper chromosome (d) Connect the medial axis and growth it

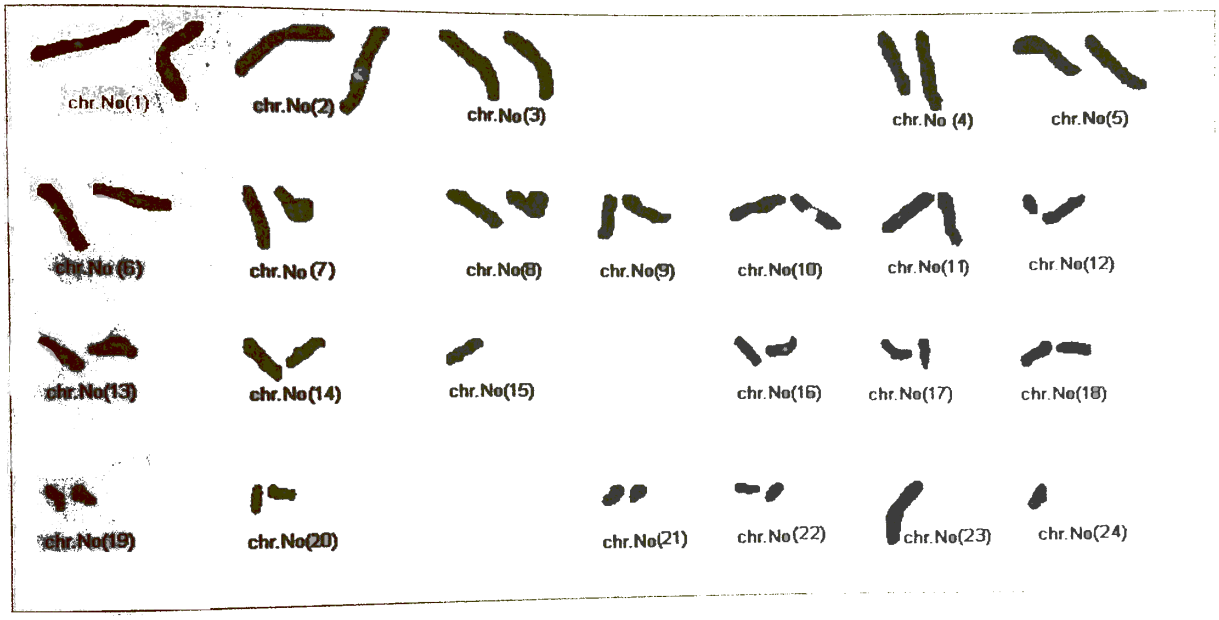


Fig.(9) Karyotype of the chromosomes in Fig.(1)

Lately, a 190 cases as new data [14] is applied to the system to achieved performance of 93.2%. This recognition rate is logical result because these data is produced by another different imaging system.
Each pixel is classified as chromosome (x) takes a certain value. All pixels equal to this value is collected to segment chromosome (x).

III. Conclusion

A system to segment and classify chromosomes based on multi-spectral information is introduced in this work. It utilizes used pixel segmentation, classification and both can be achieved simultaneously whilst in traditional methods the segmentation is accomplished as a separated step before the classification step. The system is able to decompose both overlaps and clusters composed of more than two chromosomes.
The feedforward neural network technique performs better than the threshold technique, because it reduces the computation time and accurate.
The use of the majority filter is appropriate to reduce misclassified pixels without creating interpolation values that may belong to another chromosome. The majority filter applied to segmented chromosome to fill any pixel surrounded with chromosome pixels. The proposed system gives 100% correct chromosome classification and it was tested for about 300 cases (about 19200 individual chromosomes). The system needs about 12 minutes to complete the calculation and present karyotype image using Matlab6.5 software on PC has specification (500MHz, 64 cash and 64 MB as a RAM memory).

References

[1] *J. M. Cho, "Chromosome classification using neural networks", IEEE Trans. on Medicine and Biology, January/February 2000, pp. 28-33*

[2] *P. Mousavi, R. Ward, S. Fels, M. Sameti, P. Lansdrop, "Feature Analysis and centromere segmentation of human chromosome images using an iterative fuzzy algorithm", IEEE Trans. on Biomed. Eng., Vol.49, No.4, April 2002, pp. 363-371*

[3] *W. Schwartzkopf, B.L. Evans, and C. Bovik, "Minimum entropy segmentation applied to multi-spectral chromosome images" Proc. IEEE Int. Conf. on Image Processing. Oct. 7-10, 2001, Vol. II, pp. 865-868, Thessaloniki, Greece*

[4] *W. Schwartzkopf, B.L. Evans, and C. Bovik, "Entropy estimation for segmentation of multi-spectral chromosome images", IEEE southwest symposium on Image Analysis and Interpretation, April 7-9, 2002, pp. 234-238, Santa Fa, NM.*

[5] *M. P. Sampat, K. castleman, and C. Bovik, "Pixel-by-Pixel classification of M-FISH Images", 2nd Joint Conf. of the IEEE Engineering in Medicine and Biology Society and the Biomedical Engineering Society, Oct. 23-26, 2002, Houston.*

[6] *B. Lerner, H. Guterman, and I. Dinstein, "A classification-driven partially occluded object segmentation (CPOOS) method with application to chromosome analysis", IEEE Trans. on signal processing, Vol.46, No. 10, pp. 2841-2847, 1998*

[7] *E. Decheve, "Heuristic knowledge based chromosome classification", Proc. Of the 1st national Conf. INFORMATION'94, Sofia, 1994. pp. 85-89*

[8] *E. Decveva, "Object-Oriented approach to the design of CAIS- an intelligent system for chromosome analysis and classification", Mathematics and education in mathematics, Proc. of the 21st spring conf. of union of Bulgarian mathematicians, Sofia, 1992, pp. 226-230*

[9] *E. Decheva, "CAIS- An intelligent system for chromosome analysis and classification", int. workshop "Artificial intelligent and Humanities", Sozopol, Sept. 1996, pp. 49-54.*

[10] *R. Lippmann, ""An introduction to computing with neural nets", IEEE Acoust., Speech, Signal processing (ASSP) Magn., April, Vol. 61, pp. 4-22,1987.*

[11] *S. Cichoki. Neural networks for optimization and signal processing. John Wiley & Sons Inc., 1993.*

[12] *P. D. Wasserman, "Neural computing; theory and practice", Van Nostrand Reinhold Inc., 1989.*

[13] *M. Speicher, S. Ballard, and D. Ward, "Karyotyping human chromosomes by combinatorial multi-fluor FISH", Nature Genet. Vol. 12, 368-375, 1996.*

[14] *http://info.med.yale.edu/genetics/ward/tavi/n_dyesmfish.html (last visit: 20/12/2004).*