

## ANALYSIS OF MALIGNANCY-ASSOCIATED DNA CHANGES IN INTERPHASE NUCLEI OF BUCCAL EPITHELIUM IN PERSONS WITH BREAST DISEASES

Natalya Boroday<sup>1,\*</sup>, Dmitriy Klyushin<sup>2</sup>, Yuriy Petunin<sup>2</sup>, Roman Andrushkiw<sup>3</sup>

<sup>1</sup>R.E.Kavetsky Institute of Experimental Pathology, Oncology and Radiobiology, NAS of Ukraine, Kyiv 03022, Ukraine

<sup>2</sup>Kyiv National Shevchenko University, Kyiv 01022, Ukraine

<sup>3</sup>Department of Mathematical Sciences and Center for Applied Mathematics and Statistics, New Jersey Institute of Technology, Newark, NJ 07102

**Aim:** To study from the point of view of statistical and geometrical theory of pattern recognition, the peculiarities of the distribution of optical density of DNA in the interphase nuclei of mammary buccal epithelium upon pathology. **Methods:** Cytogenetic investigation of buccal smears and computer-based image analysis were used. **Results:** It is shown that in malignant neoplasms of the mammary glands compared with its values in benign, the optical density of DNA in the nuclei of buccal epithelium increase in a range from 0.15 to 0.30 conventional units. The sensitivity of that criterium varied from 80.5% to 96.8%, and specificity was 92.3%. **Conclusion:** The method proposed may be recommended as an additional one for improvement of the diagnosis of mammary pathologies. **Key Words:** breast cancer, malignancy-associated changes, buccal epithelium, confidence interval, pattern recognition.

The work is aimed on the study of the peculiarities for the distribution of optical density of DNA in the interphase nuclei of buccal epithelium upon the pathology of mammary glands, from the point of view of statistical and geometrical theory of pattern recognition. Two new indices characterizing this distribution — ratio of modal class volumes and relief index — are proposed.

The smears of buccal epithelium were obtained from patients with breast diseases cured in the Department of Breast Cancer of the Institute of Oncology (Kyiv, Ukraine). Age of the patients was in the range of 25 to 50 years. The scanograms of the interphase nuclei of buccal epithelium in patients suffering from fibroadenoma (FA,  $n = 12$ ), fibroadenomatosis (FAM,  $n = 14$ ), and cancer of the mammary gland ( $n = 41$ ): infiltrative ductal cancer (IDC,  $n = 11$ ), infiltrative lobular cancer (ILC,  $n = 18$ ), infiltrative ductal-lobular cancer (IDLC,  $n = 10$ ) and scirrhous ( $n = 2$ ) are investigated.

For the purposes of this study smears from various depth of the spinous layer were obtained (conventionally they were denoted as median and deep), after gargling and removing the superficial cell layer of the buccal mucous. The smears were dried out under room temperature and fixed for 30 min in Nikiforov's mixture. Then, a Feulgen reaction was made with cold hydrolysis in 5 N HCl for 15 min at 21–22 °C. Optical density of the nuclei was registered by scanning cytospectrophotometry at wave length of 575 nm and probe 0.05 μm.

In each preparation, 10 to 20 nuclei were investigated. The DNA-fuchsine content in the nuclei of the epitheliocytes was defined as a product of the density and area (in terms of conventional units). The scanograms obtained as a results of the investigations of the nuclei of the cells were analyzed using statistical and geometrical methods of pattern recognition [1–4].

The scanogram of the DNA distribution is a rectangular matrix

$$R = \left\| r_{ij} \right\|_{i=1, \overline{m}}^{j=1, \overline{n}}$$

where  $r_{ij}$  are values of pointwise optical density of chromatin in interphase nuclei of the cell, expressed in terms of conventional unit of measure, and  $n, m$  are the numbers of points of the scanogram along vertical and horizontal lines, respectively. Usually the scanogram contains 8 or 9 rows and columns, hence it consists of 64 or 81 numbers.

The first index, termed the ratio of modal class volumes (RMCV), is obtained by considering the set of all scanograms as an unarranged set of random values from some general population, and by distributing this set into 3 modal classes consisting of the random values from the predefined ranges

$M_1 = \{s_{ij} : 0 \leq s_{ij} < 0.15\}$ ,  $M_2 = \{s_{ij} : 0.15 \leq s_{ij} \leq 0.30\}$ ,  $M_3 = \{s_{ij} : s_{ij} > 0.30\}$ , and, finally, by calculating the ratio of volumes of the modal classes  $M_1$  and  $M_2$  in  $k$ th scanogram:

$$V_k = \frac{\text{card } M_1^{(k)}}{\text{card } M_2^{(k)}}$$

where  $\text{card } M_j^{(k)}$ ,  $j = 1, 2$  is the number of the elements from the modal class  $M_j^{(k)}$  (for example,  $\text{card } M_2^{(k)}$  is the number of the points in  $k$ th scanograms where the DNA

Received: December 8, 2003.

\*Correspondence: E-mail: inst@iepor.kiev.ua, vm214@dcp.kiev.ua

**Abbreviations used:** FA — fibroadenoma; FAM — fibroadenomatosis; IDC — infiltrative ductal cancer; ILC — infiltrative lobular cancer; IDLC — infiltrative ductal-lobular cancer.

optical density varies from 0.15 to 0.30). RMCV is calculated by the average of all scanograms for each patient:

$$V = \frac{1}{N} \sum_{k=1}^N V_k.$$

This statistical index contains the information about the distribution of the DNA optical density in the inter-phase nuclei of epitheliocytes in buccal epithelium.

The second index, called relief index (RI), is based on the geometrical interpretation of the features of the patient's scanogram. To calculate this index, we consider the patient's scanogram as a surface of the function of two arguments  $s_{ij} = s(i, j)$ , where  $(i, j)$  are the coordinates of the points in scanogram (Fig. 1). To characterize the relief of the surface that corresponds to  $k^{\text{th}}$  scanogram, we calculate the average slope of its slices with respect to the coordinates  $i$  and  $j$ :

$$R_k = \frac{1}{n^2} \left( \sum_{i=1}^n \sum_{j=1}^{n-1} |s_{ij+1} - s_{ij}| + \sum_{i=1}^{n-1} \sum_{j=1}^n |s_{i+1j} - s_{ij}| \right).$$

RI is determined as the average of all scanograms for each patient by a formula:

$$R = \frac{1}{N} \sum_{k=1}^N R_k,$$

where  $N$  is the number of scanograms. By calculation of mentioned indices for all scanograms from the training samples, we determine the corresponding confidence regions.

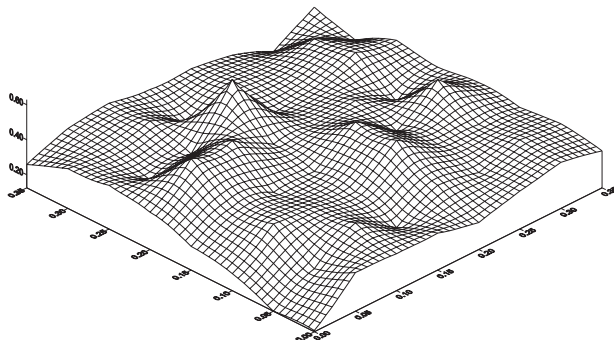


Fig. 1. The surface of DNA optical density in the scanogram of a patient suffering from infiltrative ductal cancer of the mammary gland

The confidence regions constructed are half planes bounded by order statistics of the samples consisting of the indexes  $V$  and  $R$ , respectively. The confidence limits for these indexes are equal to the minimal and maximal order statistics. According to [5], the probability that  $(n+1)^{\text{th}}$  sample value of index from the same population will exceed the maximal order statistics, is equal to

$$P(x_{n+1} > x_{(n)}) = \frac{1}{n+1},$$

where  $x_{(n)}$  is the maximal order statistics (the  $n^{\text{th}}$  order statistics) and  $x_{n+1}$  is the  $(n+1)^{\text{th}}$  sample value. Thus, the significance level of the confidence rectangle is equal to

$$\alpha = \frac{1}{n+1}.$$

The evaluation of the above indices was based on patients' scanograms, and checking whether the point  $(V, R)$  belongs to the corresponding confidence region

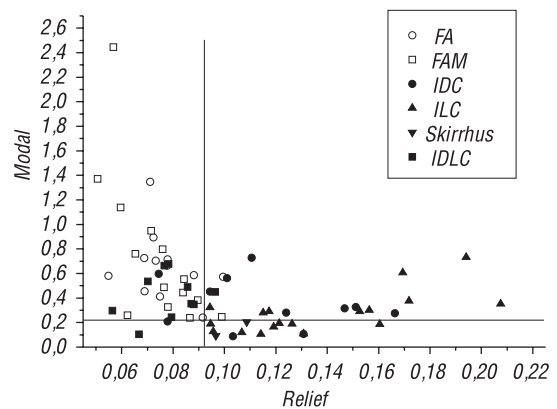


Fig. 2. Confidence regions for the relief indices and ratios of modal class volumes for patients suffering from diseases of the mammary gland (including IDLC)

(Fig. 2) constructed for the bulk of the distribution corresponding to the given significance level with the help of ordered statistics or the  $3\sigma$ -rule. We use the confidence intervals constructed with the help of ordered statistics.

Let us calculate the probabilities of the errors of the first and second kinds of proposed method. Suppose the main hypothesis  $H$  is that the patient is suffering from breast cancer, and the alternative competing hypothesis  $H'$  is that the patients is suffering from fibroadenomatosis. As it is well known [6], the probability of error of the first kind is defined as the probability of rejection of the hypothesis  $H$  when it holds true. In our case it is equal to  $p(H' | H)$ . To estimate the probability  $p(H | H)$  we use the formula shown below.

The significance levels for the confidence regions corresponding to breast cancer and benign diseases was computed for the case when IDLC was ignored. For the case when IDLC was included the confidence region was the same. So, we have

- 1) for breast cancer  $\alpha_1 = p(R < 0.0925) = 3/32 = 0.09375$ ;
- 2) for benign diseases  $\alpha_2 = p(V > 0.2090) = 3/27 = 0.11111$ ,

(the order statistics  $x_{(n-1)}$  and  $x_{(n)}$  from the sample corresponding to breast cancer and the order statistics  $y_{(1)}$  and  $y_{(2)}$  from the sample corresponding to benign diseases were ignored).

We can estimate the sensitivity and specificity of the method of diagnosis of breast cancer:

- 1) in the case when IDLC is ignored
  - $p(H | H) = p(R > 0.0925 \ \& \ V < 0.2090) = 30/31 \approx 0.9677$ ,
  - $p(H' | H') = p(R < 0.0925 \ \& \ V > 0.2090) = 24/26 \approx 0.9231$ .
- 2) in the case when IDLC is taken into account
  - $p(H | H) = p(R > 0.0925 \ \& \ V < 0.2090) = 33/41 \approx 0.8049$ ,
  - $p(H' | H') = p(R < 0.0925 \ \& \ V > 0.2090) = 24/26 \approx 0.9231$ .

Thus, we see that the IDLC may be recognized among others hystological forms of breast cancer on the basis of the analysis of the relief index of scanograms.

In conclusion, it is shown that the optical density of DNA in the nuclei of buccal epithelium from breast can-

cer patients compared with that in patients with benign diseases of mammary gland is characterized by an increase for 0.15–0.30 conventional units. The sensitivity of the proposed criterium in the studied cases (excluding IDLC) was 96.8 % and its specificity was 92.3%, whilst including IDLC — 80.5% and 92.3%, respectively.

#### REFERENCES

1. Petunin YuI, Klyushin DA, Andrushkiw RI, Gani-na KP, Boroday NV. Computer-aided differential diagnosis of breast cancer and fibroadenomatosis based on malignancy associated changes in buccal epithelium. *Automedica* 2001; **19**: 135–64.
2. Petunin YuI, Klyushin DA, Andrushkiw RI. Non-linear algorithm of pattern recognition for computer-aided diagnosis of breast cancer. *Nonlinear Analysis* 1997; **30**: 5431–6.
3. Avtondilov GG, Kholodova JL, Lysenko ON, Strijova NV. Morphometric characteristic (evaluation of ploidy) for differentiation stage of uteral adenocarcinoma. *Arkhiv Patologii* 2002; **64**: 27–33 (In Russian).
4. Gundorova LV, Avtondilov GG, Saniev KB, Zay-rantyants OV. Morphometric diagnosis of precancer and cancer of prostate by data for study of ploidy. *Arkhiv Patologii* 2003; **65**: 46–50 (In Russian).
5. Madreimov I, Petunin YuI. Characterization of the uniform distribution with the help of the order statistics. *Theory Probab Math Statist* 1982; **27**: 96–102.
6. Bairamov I, Petunin YuI. Statistical tests based on training samples. *Cybernetics* 1991; **3**: 74–7.

## АНАЛИЗ ОПУХОЛЬАССОЦИИРОВАННЫХ ИЗМЕНЕНИЙ ДНК ИНТЕРФАЗНЫХ ЯДЕР БУККАЛЬНОГО ЭПИТЕЛИЯ У ЛИЦ С ЗАБОЛЕВАНИЯМИ МОЛОЧНОЙ ЖЕЛЕЗЫ

**Цель:** изучить статистические и геометрические особенности распределения оптической плотности ДНК в интерфазных ядрах буккального эпителия при патологии молочной железы. **Методы:** проводилось клиническое обследование пациентов, цитогенетическое исследование мазков буккального эпителия и компьютерный анализ изображений. **Результаты:** показано, что злокачественные новообразования молочной железы приводят к изменениям в буккальном эпителии, которые характеризуются повышением оптической плотности ДНК в диапазоне от 0,15 до 0,30 условных единиц по сравнению с таковым в доброкачественных новообразованиях. Чувствительность предложенного критерия диагностики рака молочной железы колеблется в пределах от 80,5% до 96,8%, специфичность равна 92,3%.

**Ключевые слова:** рак молочной железы, опухольассоциированные изменения, буккальный эпителий, доверительный интервал, распознавание образов.