ORIGINAL ARTICLE

Comparision of Inter-Observer Variability between Computer Aided Image Analysis and Manual Counting of AgNORs in Grading of Oral Leukoplakia: A Double Blind Study

Vineet Raj[1], Kavita Nitish Garg[1], Shaleen Chandra[1], Abhishek Khare[1] and Amrita Bose[2]

ABSTRACT

Background: Histopathological diagnosis of oral leukoplakia has always been considered subjective with large inter-observer variability. Argyrophilic Nuclear Organizing Region (AgNOR) count has been suggested as an objective method in differentiating dysplastic lesions from non-dysplastic lesions. Computer aided image analysis systems are a recent introduction in diagnostic pathology which might help in increasing the objectivity of histological techniques.

Material and Method: This study was designed to assess the inter-observer variability in counting of AgNOR dots and in diagnosis of dysplastic oral leukoplakia using manual as well as image analysis technique and hence to assess the efficiency of these methods as objective diagnostic tools. Paraffin embedded tissue sections from 30 oral lesions, diagnosed clinically as leukoplakia, were stained for AgNOR and analyzed by five different observers. Hundred nuclei per section were counted manually in the parabasal layer and the results were expressed as mean AgNOR count/nucleus. Counting of AgNOR dots was also done using image analysis software on photomicrographs of the same sections. Based on the counts obtained each observer categorized the lesion as dysplastic or non-dysplastic leukoplakia, using an established cut off, which was correlated with standard histopathological diagnosis.

Results: Results of the study showed reduced inter-observer variability and higher overall efficiency of image analysis assisted AgNOR technique in grading leukoplakias as compared to manual counting.

Conclusion: Computer assisted image analysis system was found to be an effective tool in achieving high reproducibility and minimum inter-observer variability, and hence improving the diagnostic efficacy of AgNOR technique in oral leukoplakia.

Keywords: AgNOR, Computer Aided Image Analysis, Inter-Observer Variability, Leukoplakia.

[1]Department of Oral & Maxillofacial Pathology, Saraswati Dental College & Hospital, Lucknow, India, [2]Centre of Clinical & Diagnostic Oral Sciences, Institute of Dentistry, Queen Mary University, London, UK.

Address for Correspondence:
Dr. Vineet Raj,
Flat no. 402, Rohtas Vintage Apartment, 29-Jopling Road, Lucknow-226001 (UP), India.
Contact No: +91-9918293535, E-mail: vineetraj@yahoo.com
Date of Submission : 17-03-2011
Reviews Completed : 06-04-2011
Date of Acceptance : 18-04-2011

INTRODUCTION

Oral cancer is one of ten most common cancers in the world and shows marked geographic differences in occurrence. Oral cancer is common where habits of betel quid chewing, bidi smoking, alcohol, and tobacco consumption are high. Thus it is a common cancer in south-east Asia, where more than 100,000 new cases are reported every year. Among all malignancies reported in India, oral cancer ranks number one among males and number three among females. The National Cancer Registry project revealed oral cancer to constitute 12% of all cancers in males and 8% of all cancers in females.[1] Cancer being a genetic disorder involves multiple alteration of the genome progressively accumulating during a protracted period, the overall effects of which surpass the inherent reparative ability of the cell. In 1978, WHO defined Oral Premalignant lesion as "a morphologically altered tissue in which cancer is more likely to occur than in its apparently normal counterpart."[2] Oral Leukoplakia is defined as a nonscrapable white patch or plaque that cannot be characterized clinically or pathologically as any other disease and which is not associated with any physical or chemical agent except the use of tobacco.[3]

The diagnosis of precancers is primarily based on clinical appearance of lesion and architectural changes on histological examination (degree of dysplasia). Oral pathologists use the term epithelial dysplasia to indicate microscopic features in a biopsy specimen that are associated with risk of malignant

changes and then assign a grade of severity. Dysplasia is defined as architectural disturbance accompanied by cytological atypia. Conventionally, dysplasia is divided into grades of mild, moderate and severe.[4] Despite the fact that this estimation is subjective and therefore carries a low prognostic value of an impending malignancy, it is still widely practiced to estimate the risk of malignant potential of such lesions. Till date, no clear markers for distinguishing various grades of epithelial dysplasia have evolved, and histological criteria for diagnosing a dysplastic lesion are still subjective. In absence of any objective criteria, there is often an inter-observer bias between pathologists in diagnosing such lesions.[5]

Nucleolar Organizing Regions (NORs) are specific portions of DNA, called rDNA, that by using the enzyme RNA polymerase 1, code for the transcription of ribosomal RNA. This rRNA inside the ribosome is responsible for protein synthesis of the cell which in turn is a necessary step in the process of cell proliferation.[6] Silver-stained Nucleolar Organizing Region (AgNOR) counts have been found to be a useful histological index of protein synthesis, and hence proliferative potential, in malignancy of various organs. The mean number of AgNORs per nucleus is reported to be higher in malignant lesions as compared to benign tissue. It increases with increasing grade of malignancy and is reported to be higher in tumors with poor prognosis compared to those with good prognosis. AgNOR counts have been suggested as an objective method in differentiating dysplastic lesions from non-dysplastic lesions. A cut-off point of 2.37 AgNORs/nuclei has been reported as a threshold for diagnosing dysplastic oral lesions.[7]

For any diagnostic test to be considered objective, it should have minimum inter-observer variability. Computerized image analysis may be a useful tool in enhancing the objectivity and hence the usefulness of AgNOR technique in early diagnosis of oral premalignancy. The present study is designed to assess the inter-observer variability in diagnosis of dysplastic oral leukoplakia, based on previously reported AgNOR cut-off point, using manual as well as image analysis technique and hence to assess the reliability of image analysis assisted AgNOR test as an objective diagnostic tool.

MATERIALS AND METHOD

Five oral pathologists participated in this double-blind study. The study material constituted of thirty biopsy samples of oral leukoplakia obtained from the archives of Department of Oral and Maxillofacial Pathology, Saraswati Dental College and Hospital, Lucknow, India. The histopathological findings were reviewed and agreed upon by two oral pathologists having at least five years of experience in histopathology of oral premalignant and malignant lesions. These diagnoses were considered as gold standard and used for comparing with AgNOR based diagnosis. Based on the above findings the cases were divided into two groups:

Non-dysplastic Leukoplakia (NDLK) included lesions clinically diagnosed as leukoplakia and on histological examination showed either benign hyperkeratosis/ hyperplasia, or mildly dysplastic epithelium (hence considered to be of low risk).

Dysplastic Leukoplakia (DLK) included lesions clinically diagnosed as leukoplakia and on histological examination showed either moderate or severe epithelial dysplasia (hence considered to be at higher risk of malignant transformation).[7]

Three micron (3 µm) thick AgNOR stained sections, coded with unique identification numbers, in order to conceal the identity of the case, were then made available to all the observers.

AgNOR Staining and Counting

For each case, 3µm thick formalin-fixed, paraffin-embedded sections were dewaxed in xylene and rehydrated through alcohol to deionised water. The sections were subjected to AgNOR staining technique as suggested by Ploton *et al*.[8] Following incubation in silver solution the sections were washed in running deionized water, dehydrated in alcohol, cleared in xylene and mounted with DPX. The AgNORs were visualized as intranuclear brown to black dots of different sizes under light microscopy (Fig 1). For each specimen, the number of AgNOR dots was counted in 100 nuclei at 100X magnification manually in the parabasal layer by five different observers between 2nd to 5th day of staining. NORs were directly counted under a light microscope according to the parameters established by Crocker *et al.*,[9] i.e., well-defined black dots in the nucleus were counted, with aggregations (overlapping or fused black dots) being considered a single

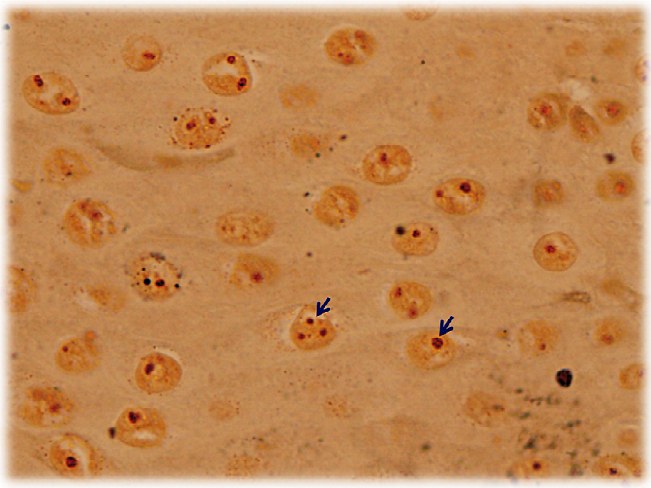

Figure 1: Histological section stained with AgNOR technique. AgNOR dots (arrows) are seen within the nuclei as well delineated brown to black dots.

structure and the mean AgNOR count per nucleus was calculated for each specimen. Photomicrographs of the same fields were taken on the 2nd day of staining using Olympus BX 51 trinocular microscope with Olympus E 330 camera and counting was performed digitally using Image Pro Express 6.0 from Media Cybernetics. Counting of AgNOR dots was performed using the point selection tool of the software with the same criteria used for manual counting.

A mean count of 2.37 AgNORs per nucleus, as suggested by Ray *et al.*,[7] was used as cut-off point for AgNOR-based diagnosis of epithelial dysplasia. Thus, specimens scoring a mean of more than 2.37 AgNORs per nucleus were considered positive, and those with less than 2.37 were considered negative. Each observer was blind to the other's diagnosis till the end of the study. Inter-observer variability in counting AgNOR dots using manual and image analysis methods was analyzed statistically using *paired-t-test* (p values < 0.05 were considered significant). The AgNOR based diagnosis was then compared with the histopathological diagnosis and inter-observer agreement between the two methods was computed by *kappa* (κ) analysis.

RESULTS

The study group consisted of thirty cases diagnosed clinically as leukoplakia. Observers were blind to the histopathological diagnosis during the time of study. Fig. 2 shows the mean AgNOR count/nucleus obtained by the five observers using manual counting as well as image analysis. With manual method the mean count obtained by observers A, B, C, D and E were 1.97±0.56, 1.94±0.51, 1.91±0.62, 2.07±0.56 and 2.13±0.54 respectively. The mean counts ranged from 1.91 (observer C) to 2.13 (observer E) with a difference of 0.22 between the maximum and minimum counts. With image analysis method the mean counts for observers A, B, C, D and E were 2.14±0.51, 2.19±0.50, 2.16±0.49, 2.12±0.49 and 2.13±0.51 respectively. The mean counts ranged from 2.12 (observer B) to 2.19 (observer D) with a difference of 0.07 between the maximum and minimum counts. There was no statistically significant interobserver variability in either techniques, as assessed by *paired-t-test* (p>0.05), but the image analysis showed a lower degree of variability in

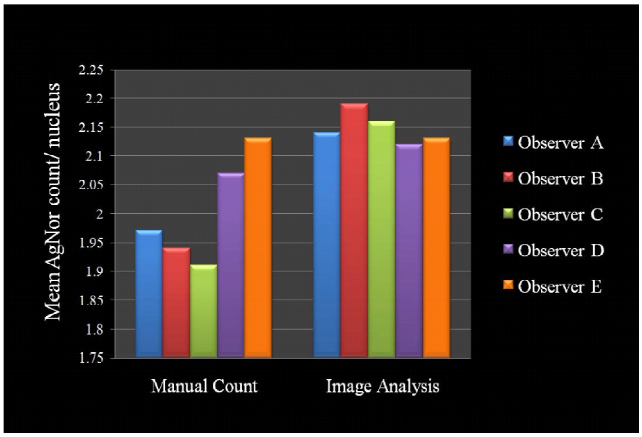

Figure 2: Graph showing mean AgNOR count/ nucleus for different observers with manual and image analysis techniques

assessing AgNOR counts between the five observers when compared to manual technique.

Based on histopathological diagnosis 17(56.67%) cases were classified as dysplastic leukoplakia (DLK) and 13(43.33%) cases as non dysplastic leukoplakia (NDLK), and these were taken as gold standard for comparing AgNOR based diagnosis. Table 1 shows classification of lesions into NDLK and DLK by all five observers based on AgNOR counts using manual observation method. Data from this table reveals that a higher percentage of lesions were diagnosed as NDLK by all the observers with percentage varying from 63.3%-73.3% which was higher than those achieved by histopathological findings (43.3%). Consequently, the diagnosis of DLK ranged from 26.67%-36.67% which is considerably less than what was achieved by histopathology. Table 2 shows classification of lesions into NDLK and DLK by all the observers based on AgNOR counts using image analysis method. Data from this table reveals that 43.3%-50% of all cases were diagnosed as NDLK and consequently, the diagnosis of DLK was made ranging from 50%-56.67% of cases which was considerably similar to that achieved by histopathological finding.

The interobserver variability among the observers in distinguishing NDLK from DLK, for both the techniques, was assessed by using *kappa* (κ) analysis, where (κ) values range from -1 to +1. Negative values indicate disagreement while

Table 1: Grading of leukoplakia done by five observers by using manual AgNOR count and compared to histopathological diagnosis (NDLK – Non dysplastic leukoplakia; DLK – dysplastic leukoplakia)

	Observer					Histopathological Diagnosis
	A	B	C	D	E	
NDLK (%)	21 (70%)	22 (73.33%)	19 (63.33%)	19 (63.33%)	19 (63.33%)	13 (43.33%)
DLK (%)	9 (30%)	8 (26.67%)	11 (36.67%)	11 (36.67%)	11 (36.67%)	17 (56.67)

Table 2: Grading of leukoplakia done by five observers by using image analysis based AgNOR count and compared to histopathological diagnosis (NDLK – Non dysplastic leukoplakia; DLK – dysplastic leukoplakia)

	Observer					Histopathological Diagnosis
	A	B	C	D	E	
NDLK (%)	14 (46.7%)	14 (46.7%)	15 (50%)	13 (43.3%)	14 (46.7%)	13 (43.33%)
DLK (%)	16 (53.3%)	16 (53.3%)	15 (50%)	17 (56.67%)	16 (53.3%)	17 (56.67)

the positive values indicate agreement. Table 3 shows the agreement rates among observers in differentiating DLK from NDLK using AgNOR test by manual method as expressed by κ value. The values were in the range of -0.066 to 0.713 with the lowest κ-value was between the observer A and B (κ = - 0.066) while highest value was between observer D and E (κ = 0.713). Overall a mild to moderate agreement was seen. Table 4 shows the agreement rates among observers using image analysis. κ values were in the range of 0.867 to 1.0. Overall a good agreement was seen between all observers in distinguishing DLK from NDLK.

Table 3: Inter-obesrver Agreement (kappa values) between different observers in differentiating DLK and NDLK by Manual Observations. (κ<0.4 - Poor agreement, κ >0.4 - Moderate agreement, κ>0.6 - Substantial agreement, κ >0.8 - Good agreement, κ =1.0 Complete agreement)

	Obs A	Obs B	Obs C	Obs D	Obs E
Obs A		-0.066	0.254	0.254	0.403
Obs B	-0.066		0.619	0.467	0.467
Obs C	0.254	0.619		0.426	0.426
Obs D	0.254	0.467	0.426		0.713
Obs E	0.403	0.467	0.426	0.713	

The diagnostic efficacies of both the methods were calculated by estimating the sensitivity, specificity, positive and negative predictive values, and overall correct classification. For manual method sensitivity was found to be 55.29%, specificity 93.85%, positive predictive value of 94%, and negative predictive value of 62%. Overall correct classification was 72.7%. Image analysis method was found to be 92.94% sensitive and 98.46% specific with a positive predictive value of 98.75%, negative predictive value of 91.43% and overall correct classification rate was 95.3% (Fig 3).

Table 4: Inter-obesrver Agreement (kappa values) between different observers in differentiating DLK and NDLK by Image analysis. (κ<0.4 - Poor agreement, κ>0.4 - Moderate agreement, κ>0.6 - Substantial agreement, κ>0.8 - Good agreement, κ=1.0 Complete agreement)

	Obs A	Obs B	Obs C	Obs D	Obs E
Obs A		1	0.933	0.933	1
Obs B	1		0.933	0.933	1
Obs C	0.933	0.933		0.867	0.933
Obs D	0.933	0.933	0.933		0.867
Obs E	1	1	0.933	0.933	

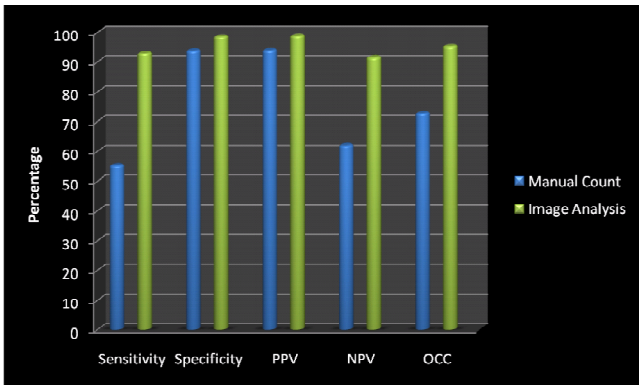

Figure 3: Diagnostic Efficacy of Manual count vs Image analysis (PPV – Positive predictive value; NPV – negative predictive value; OCC – overall correct classification)

DISCUSSION

Leukoplakia is a clinical diagnosis with histological presentation ranging from mild hyperkeratosis to severe dysplasia. It is important to distinguish between these lesions because of differing malignant potential, clinical management and prognosis. The diagnosis of oral epithelial dysplasia is irrelevent without the assessment of its severity, and hence there is a need to distinguish between various grades because treatment for moderate and severe dysplasia needs to be more aggressive as compare to hyperplastic or mildly dysplastic lesions. The current grading scheme of dysplasia is primarily based on the subjective evaluation of nuclear characteristics such as pleomorphism, enlargement, and hyperchromasia and spatial features such as orientation and stratification. In absence of any objective criteria, under current standards there occurs a wide variation between oral pathologists in diagnosing such lesion.[10] Hypertrophy of the nucleolus is one of the most distinctive cytological feature of cancer cells. Dysplastic cells more frequently display a larger nucleolus than benign cells and nucleolar size might represent a morphological parameter of cell proliferation rate in cancer

tissue.[11] AgNOR staining has been applied to a variety of lesions in order to provide more information on their nucleolar activity.[8] The silver staining tends to highlight NORs which are actively transcribing rRNA and increased NOR count has been reported in lesions with active proliferation including those of oral tissues.[12-15] Since H&E is a basic and, in experienced hands, an extremely useful staining method with wide utility, there is little use in considering AgNOR as a replacement of H & E per se, but as H&E based grading is subjective, it is conceivable that it may misclassify some of these lesions. AgNORs are thought to be sound replicatory markers and therefore, a useful adjunct in grading the severity of epithelial proliferation.

For any test to become a reliable adjunct to diagnosis in clinical setting high degree of reproducibility with minimum interobserver variability is a prerequisite. Computerized image analysis system has emerged in recent years as a powerful tool for bias free objective and reproducible quantification of histological features. It has shown considerable potential for diagnostic application in diverse histological situations especially in quantification of subvisual features for diagnostic applications.[16] The method involves taking photographs of slides at a fixed focal distance and projecting them onto a screen which produces an enlarged image ready for scoring and then the images are transferred to software where all the parameters are calculated. It has been suggested that automated image analysis may help to reduce inter and intra-examiner variability and will allow both enumeration and quantification of AgNORs.[17] Our study revealed that though there was no statistically significant interobserver variability in either of the two methods of counting AgNORs there was lower interobserver difference in mean AgNOR counts with image analysis as compared to manual counting (Fig 2). A possible reason for this difference can be explained by the fact that in manual observation method a bias in selection of field can cause variability in the counts of different observers. On the other hand using image analysis all the observers assessed similar fields which had been selected by an experienced oral pathologist. This also reduces the variability arising due to difference in the level of experience and training of the observer. Hence, an appropriately trained technical hand may be able to carry out the test under guidance of an experienced oral pathologist thus saving time and workforce of more specialized personnel.

The results also revealed that the mean AgNOR counts for all the observers except observer E were higher when using image analysis as compared to those obtained by manual observation technique. The reason for this variability could be the fact that when using manual observation technique it is often difficult to differentiate between small AgNOR stained dots and nonspecifically stained particles and hence these were usually excluded from the count. The drawback of nonspecific staining in AgNOR technique has been discussed previously.[18] The option of image enhancing and magnification in image analysis helps in differentiating between specific and nonspecific staining and hence smaller specific dots could also be counted. This advantage of high magnification with enhanced resolution also helped us in differentiating closely placed dots which on manual observation might have been counted as single dots. Another limitation of AgNOR technique is the lack of permanence of the record as the stain tends to fade away with time, sometimes within few days[18], making reviewing of the slides unreliable. This limitation can be overcome by using image analysis in which images can be taken at a time when the staining is crisp and can then be stored permanently for future reference.

When we assessed the ability to differentiate dysplastic lesions from nondysplastic lesions, using established AgNOR cut off value, it was seen that all observers classified considerably lesser number of lesions to be dysplastic by manual counting as compared to image analysis. When these findings where compared with standard H&E based diagnoses, image analysis appeared to correlate more closely to histopathology than manual counts (Table 1 and 2). When the level of agreement between the various observers was assessed using κ-value a much higher interobserver agreement in differentiating DLK from NDLK was observed while using image analysis (Table 3 and 4). Reduced interobserver variability in grading of dysplastic lesions using computer assisted image analysis has been advocated previously.[19-21] When the efficacy of the two methods was compared image analysis scored higher in all parameters i.e. sensitivity, specificity, positive predictive value, negative predictive value, and overall correct classification (Fig 3). High diagnostic efficiency of computer aided image analysis systems have been reported previously in various diseases,[16,22,23] and hence its use in oral lesions needs to be explored further.

CONCLUSION

To conclude, AgNORs have a high potential to be used as an adjunct diagnostic tool for assessment of proliferative potential in oral epithelial lesions. Computer assisted image analysis system can be an effective tool in achieving high reproducibility and minimum inter-observer variability, thus improving the diagnostic efficacy of AgNOR technique leading to timely diagnosis and proper management of potentially malignant oral lesions.

REFERENCES

1. Sankaranarayanan R. Oral cancer in India: an epidemiologic and clinical review. Oral Surg Oral Med Oral Pathol 1990; 69: 325-30.
2. Rajendran R. Benign and malignant tumors of the oral cavity. In. Shafer's Textbook of Oral Pathology. Rajendran R, Shivapathasundharam B, (eds). 5th ed, Reed Elsevier India Private Limited, New Delhi. 2006, p. 113-308.

3. Karabulut A, Reibel J, Therkildsen MH, Praetorius F, Nielsen HW, Dabelsteen E. Observer variability in the histologic assessment of oral premalignant lesions. J Oral Pathol Med 1995; 24: 198-200.
4. Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. J Oral Pathol Med 2008; 37: 127-33.
5. Chattopadhyay A, Ray JG, Caplan DJ. AgNOR count as objective marker for dysplastic features in oral leukoplakia. J Oral Pathol Med 2002; 31: 512-7.
6. Derenzini M, Sirri V, Trerè D, Ochs RL. The quantity of nucleolar proteins nucleolin and protein B23 is related to cell doubling time in human cancer cells. Lab Invest 1995; 73: 497-502.
7. Ray JG, Chattopadhyay A, Caplan DJ. Usefulness of AgNOR counts in diagnosing epithelial dysplasia. J Oral Pathol Med 2003; 32: 71-6.
8. Ploton D, Menager M, Jeannesson P, Himber G, Pigeon F, Adnet JJ. Improvement in the staining and in the visualization of the argyrophilic proteins of the nucleolar organizer region of the optical level. Histochem J 1986; 18: 5-14.
9. Crocker J, Bold DAR, Egan MJ. How should we count AgNORs? Proposals for a standardized approach. J Pathol 1989; 158: 185–8.
10. Chattopadhyay A, Ray JG. AgNOR cut-point to distinguish mild and moderate epithelial dysplasia. J Oral Pathol Med 2008; 37: 78-82.
11. Derenzini M, Trerè D, Pession A, Govoni M, Sirri V, Chieco P. Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. J Pathol 2000; 191: 181-6.
12. Crocker J, Egan MJ. Correlation between NOR sizes and numbers in non-Hodgkin's lymphomas. J Pathol 1988; 156: 233-9.
13. Giri DD, Nottingham JF, Lawry J, Dundas SAC, Underwood JCE. Silver-binding nucleolar organizer regions (AgNORs) in benign and malignant breast lesions: correlations with ploidy and growth phase by DNA flow cytometry. J Pathol 1989; 157: 307-13.
14. Teixeira G, Antonangelo L, Kowalski L, Saldiva P, Ferraz A, Silva Filho G. Argyrophilic nucleolar organizer regions staining is useful in predicting recurrence-free interval in oral tongue and floor of mouth squamous cell carcinoma. Am J Surg 1996; 172: 684-8.
15. Xie X, Clausen OP, Sudbö J, Boysen M. Diagnostic and prognostic value of nucleolar organizer regions in normal epithelium, dysplasia, and squamous cell carcinoma of the oral cavity. Cancer 1997; 79: 2200-8.
16. Gupta N, Sarkar C, Singh R, Karak AK. Evaluation of diagnostic efficiency of computerized image analysis based quantitative nuclear parameters in papillary and follicular thyroid tumors using paraffin-embedded tissue sections. Pathol Oncol Res 2001; 7: 46-55.
17. Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. J Oral Pathol Med 2008; 37:127-33.
18. Linder LE. Improvements in the silver-staining technique for nucleolar organizer regions (AgNOR). J Histochem Cytochem 1993; 41: 439-45.
19. van Sandick JW, Baak JP, van Lanschot JJ, Polkowski W, ten Kate FJ, Obertop H, *et al*. Computerized quantitative pathology for the grading of dysplasia in surveillance biopsies of Barrett's oesophagus. J Pathol 2000; 190: 177-83.
20. Polkowski W, van Lanschot JJ, van Sandick JW, Baak JP. The role of quantitative pathology in clinical decision making for Barrett's oesophagus. Anal Cell Pathol 2003; 25: 123-7.
21. Sabo E, Beck AH, Montgomery EA, Bhattacharya B, Meitner P, Wang JY, Resnick MB. Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in Barrett's esophagus. Lab Invest 2006; 86: 1261-71.
22. Marchevsky AM, Hauptman E, Gil J, Watson C. Computerized interactive morphometry as an aid in the diagnosis of pleural effusions. Acta Cytol 1987; 31: 131-6.
23. Walts AE, Marchevsky AM. Computerized interactive morphometry. An expert system for the diagnosis of lymphoid-rich effusions. Am J Clin Pathol 1989; 92: 765-72.