



## Improving discrimination in the grading of rat mammary tumors using two-dimensional mapping of histopathological observations

Ana C. Lopes<sup>a,c</sup>, Tânia F.G.G. Cova<sup>b</sup>, Alberto A.C.C. Pais<sup>b</sup>, Jorge L.G.F.S.C. Pereira<sup>b</sup>, Bruno Colaço<sup>d</sup>, António M.S. Cabrita<sup>a,b,\*</sup>

<sup>a</sup> Department of Experimental Pathology, Faculty of Medicine, University of Coimbra, Coimbra, Portugal

<sup>b</sup> Chemistry Department, University of Coimbra, Coimbra, Portugal

<sup>c</sup> University School of Vasco da Gama, Castelo Viegas, Coimbra, Portugal

<sup>d</sup> CECAV – University of Trás-os-Montes e Alto Douro, Vila Real, Portugal

### ARTICLE INFO

#### Article history:

Received 5 June 2013

Accepted 26 September 2013

#### Keywords:

Mammary tumors

DMBA

Histological grading

Principal component analysis

### ABSTRACT

This work aims at characterizing rat mammary tumors induced by 7,12-dimethylbenz(a)anthracene (DMBA) and the respective malignancy potential, commonly graded with histopathology features grouped by intensity levels. Tumors were described over fourteen multiple ranged microscopic parameters and a comprehensive characterization of the histological patterns and their relation with tumor grade was carried out by principal component analysis (PCA). The number of histological patterns present on a tumor tends to correlate with malignant features. High grade tumors are characterized by the presence of several structural patterns, with cribriform prevalence and necrosis. The cribriform pattern correlates with grading, i.e., tumors having a higher predominance of the cribriform pattern are likely to be more malignant. The findings may represent a benchmark for similar characterization studies in other models.

© 2013 Elsevier GmbH. All rights reserved.

### 1. Introduction

Rat models have been accepted in oncobiology research for a long time due to characteristics comparable with human breast cancer, such as the degree of morphologic aggressiveness and tumor patterns (Currier et al., 2005; Russo and Russo, 1996, 2000; Withrow and Vail, 2007; Cardiff, 2001; Tavassoli and Devilee, 2003; Costa et al., 2002; Medina, 1976; Thompson and Singh, 2000; Cardiff and Munn, 1995; Cardiff et al., 2000; Russo et al., 1996; Rakha et al., 2010; Martins et al., 2008; Dias et al., 2001). The transposition between rat and human breast cancer is well documented. In these animals, the mammary gland shows age susceptibility for developing hormone dependent adenocarcinomas, which are very similar to those most frequently found in women (Costa et al., 2002; Russo and Russo, 2000). In humans and animals, over 90% of breast cancers are attributed to non-hereditary aetiology such as environmental carcinogen exposure and diet (Currier et al., 2005; Russo and Russo, 1996). In this context, rat models chemically induced by 7,12-dimethylbenz(a)anthracene (DMBA) are particularly

useful for the study of breast cancer (Currier et al., 2005; Russo and Russo, 1996; Thompson and Singh, 2000; Russo et al., 1996). The exposure of terminal end-buds (TEBs) to DMBA causes unrepaired DNA damage, resulting on mutation, translocation, and inactivation of regulatory genes with long-lasting alterations in cell growth and in anti-apoptotic pathways leading to tumorigenesis (Currier et al., 2005; Russo and Russo, 1996; Thompson and Singh, 2000; Russo et al., 1996). Mammary tumors chemically induced by DMBA are heterogeneous and need to be described according to histological pattern or type, degree of differentiation reflected on histological grade and other morphological characteristics. The histological type is classified according to the presence of the most prevalent pattern of growth either on single or mixed pattern tumors (Costa et al., 2002). Histological grade is considered a valuable prognostic factor in breast cancer. The basis of this indicator relies on the degree of differentiation of the tumor tissue useful to predict the tumor malignancy (Cardiff et al., 2000; Rakha et al., 2010). The World Health Organization (WHO) recommends the Nottingham Grading System (NGS) as the grading system for human breast cancer (Tavassoli and Devilee, 2003). NGS is based on the evaluation of three morphological features: (1) degree of tubular or gland formation, (2) nuclear pleomorphism and (3) mitotic index. Nevertheless, there is an increasing need for more accurate prognostic and predictive markers (Webster et al., 2011; Meyer et al., 2005). In order to improve DMBA rat model for breast cancer investigation, some

\* Corresponding author at: Department of Experimental Pathology, Faculty of Medicine, University of Coimbra, Coimbra, Portugal. Tel.: +351 239 857 777; fax: +351 239 857 745.

E-mail address: [amscabrita@gmail.com](mailto:amscabrita@gmail.com) (A.M.S. Cabrita).

groups adapted grading parameters and correlated morphologic features also used in human breast pathology. The most relevant features are the structural pattern, percentage of cribriform pattern, gland secretion, inflammatory cell infiltration, hemorrhage, necrosis, stromal reaction and microcalcification, all possessing ranges of grading dependent on tumoral prevalence (Russo and Russo, 1996).

In human medicine, Breast Cancer Panels typically include ER/PR, a proliferation marker (Ki67, PCNA, or QDNA), Her-2, an angiogenesis marker (CD31, vegf, or factor VIII) and p53. The Annapolis Pathology Panel found that few model systems have been studied with immunohistochemistry (IHC), and even fewer have been characterized using the expected standard biomarker panel. As a result, direct comparison between human and mouse mammary tumors is still not possible (Cardiff, 2001). In this work, the prognostic features based on hematoxylin and eosin (HE) histopathology, a readily accessible method of tumoral evaluation, are maximized. The need to improve traditional histopathology has been complemented using multivariate analysis techniques (Cova et al., 2013; Madsen et al., 2010; Cheng et al., 2005; Méndez et al., 2009; Selaru et al., 2004). Clinical tools often produce large amounts of data, being multivariate in nature. Statistical, multivariate data analysis cannot be dissociated from most of the problems found in cancer studies (Cova et al., 2013). They form a commonly used set of methods for obtaining reliable results. Dimensionality reduction techniques based on variance, as principal component analysis (PCA), have been used for variable selection and data overview, which will reveal outliers, groups and trends.

In this work, a step forward on data analysis of histopathological and morphological parameters is presented. The analysis is carried out by employing a procedure for dimension reduction over a data set composed of routine mammary tumors chemically induced on Sprague Dawley (SD) virgin female rats. All parameters are defined by scores related to their intensity and are then further detailed and rationalized by PCA, which facilitates two- or three-dimensional representation. This approach enables to assess the relation between the histological features of each tumor and its grade. Additionally, inspection of the most discriminant parameters provides a detailed understanding of the tumors profile. An improvement in malignancy prediction, traditionally established by tumor grading, can be achieved if a significant relation is identified between histological parameters and grade. As a result, a better understanding of tumor characteristics enables a more accurate comparison between human breast tumors and those chemically induced in rat models, frequently used in cancer research.

## 2. Methods

Forty-eight outbred Sprague Dawley (SD) virgin female rats were housed randomly into groups of three elements in each cage at a temperature of  $23 \pm 2^\circ\text{C}$  and humidity of 50–55%, controlled facility on a 12 h light, 12 h dark cycle and fed a standard laboratory chow. Food and water were maintained *ad libitum*. All animals were weighted weekly. At 50 days of age, the animals were orally administered with 65 mg/kg of DMBA diluted in virgin olive oil (a maximum of one millilitre per animal). At 20 weeks after carcinogenic induction, tumors started to be detectable by mammary palpation and a weekly registration of the number, size and shape was carried out. At 27 weeks, all animals were ethically euthanized under anaesthesia. A full necropsy was performed and all tumors were excised, measured on three dimensions for volume calculation and weighed. The mammary tissue macroscopically free of tumors was collected for histopathologic evaluation.

### 2.1. Histopathologic evaluation

The Annapolis consensus (Cardiff et al., 2000) clarifies the classification of mammary tumors in genetically engineered mice, especially in what concerns to histological types. Russo adapted the Annapolis consensus to rat mammary tumors, increasing the field of compared pathology (Russo and Russo, 2000). In practical terms, the tissue fragments were fixed in 10% of formalin.  $5\ \mu\text{m}$  sections obtained from paraffin blocks were stained by HE for histologic examination. The rat tumors were graded using NGS in grade 1 (3–5 points), grade 2 (6–7 points) and grade 3 (8–9 points) (Tavassoli and Devilee, 2003). The tubular and gland formation were scored according to its predominance in the tumor as 1, if present in more than 75%, 2 if present in a moderate degree (10–75%) and 3 for lower predominance or absence (less than 10%). Nuclear pleomorphism was classified as 1, for small and regular cells, 2 for moderate size and shape variation, and 3 for a marked shape variation and mitotic counts at the periphery of the tumor over 10 consecutive high-power fields (HPF), with a field of view of 0.625 mm on a Nikon Eclipse E-600 microscope. In these conditions, the total mitotic counts of 0–11 have score 1, for 12–22 and over 23 have scores 2 and 3, respectively (Thompson and Singh, 2000). Similarly to human breast cancer histological evaluation, additional parameters were also assessed (Russo and Russo, 1996; Thompson and Singh, 2000). The structural pattern was classified and graduated according to the type of lesion: non-neoplastic (0), benign neoplastic (1), *in situ* malignant neoplastic (2) and invasive malignant neoplastic (3). Each histological type was identified and registered according to the most prevalent patterns: papillary, cribriform, glandular, tubular, comedo and squamous. The total number of patterns present in each tumor was registered. The proportion of cribriform pattern was identified as low (0–30%, score 1), medium (40–60%, score 2), or high (70–100%, score 3). Microcalcifications and stromal invasion were classified as 1, if present, and 0 otherwise. Secretion, necrosis and hemorrhage were sorted in absent (score 0), focal (10%, score 1), moderate (20–70%, score 2) and extensive (if >80%, score 3). Stromal reaction was classified as absent (0), mild (1), moderate (2) and high (3) according to collagen tissue within and surrounding the tumor. The same score system was repeated for apoptosis and inflammatory cell infiltration.

### 2.2. Data analysis procedure

Principal component analysis (PCA) is a simple, non-parametric procedure of extracting relevant information from data sets (Jolliffe, 2002). PCA computes a compact and optimal description of the data set, providing a roadmap to lay out a complex data set to a lower dimension and reveal the hidden, simplified structure that often is underlying. The most influential variables in the system are highlighted, and the most relevant factors may be identified. This technique is based on the assumption that most of the information about the structure of the data is contained in the directions along which the variations are the largest (Massart, 1988; Breton, 1990; Jolliffe, 2002). In this work, PCA summarizes the information residing in the data corresponding to the histological parameters into a form which may be more easily inspected and used. The original multi-dimensional space, defined by those parameters, is contracted into a few descriptive dimensions, which represent the main variation in the data. Each principal component (PC) can be displayed graphically and analysed separately, and its meaning may often be interpreted according to simple histological descriptors. Essentially, the procedure is carried out by a linear transformation of the  $m$  histological parameters  $\mathbf{x}_i$  into a new set, the principal components  $\mathbf{u}_i$

$$\mathbf{u}_i = w_{i1}\mathbf{x}_1 + w_{i2}\mathbf{x}_2 + \dots + w_{im}\mathbf{x}_m \quad (1)$$

where  $w_{11} \dots w_{im}$  are the loadings, i.e. the weights of the observed parameters in the linear combination (Wang and Paliwal, 2003).

In PCA, the original matrix  $\mathbf{X}(n \times m)$ , in which  $n$  rows correspond to  $n$  tumors and  $m$  columns correspond to the number of histopathological parameters, is decomposed as a product of two matrices,

$$\mathbf{X} = \mathbf{S}\mathbf{W}^T \quad (2)$$

in which  $\mathbf{S}(n \times m)$  is the matrix of the scores, i.e. the coordinates of the samples in the principal components, and  $\mathbf{W}(m \times m)$  is the loadings matrix. Since the first principal components retain most of the variance, several variables can be summarized by a few components and a plot of the first two or three PCs enables the visualization of most of the information contained in the data.

The definition and computation of principal components are thus straightforward, and this apparently simple technique has proved extremely useful in a wide variety of different applications (Campanella et al., 2003; Kokot et al., 1998; Madsen et al., 2010; Polat and Günes, 2008). Note that preliminary data inspection is especially relevant when developing disease diagnosis tools. In this context, PCA should be the first choice to look for the overall characteristics of the data (Wold et al., 1987; Jackson and Wiley, 1991). In medical research, particularly in oncology diseases, PCA has been used for processing data to extract significant variables and to distinguish cancer patients and healthy subjects (Madsen et al., 2010; Cheng et al., 2005; Méndez et al., 2009; Selaru et al., 2004), as an example among various others [see e.g., (Polat and Günes, 2008; Reich et al., 2008; Polat and Günes, 2007; Latifoğlu et al., 2008; Espeland et al., 2009; Spratlin et al., 2009; Edefonti et al., 2010; Xie et al., 2011; Gualberto et al., 2010)].

### 2.3. Data description

The procedure based on PCA requires a description of objects, points in Euclidean space. In this analysis, the tumors correspond to the objects. Each tumor is described on the basis of 14 components related to the prevalence of histological parameters. Specifically, the data set contains information on 148 tumors characterized by 14 microscopic observations, including the number and type of structural patterns (NPAD), proportion of the cribriform pattern (PCRI), presence of microcalcifications (MIC), stromal invasion (IEST), necrosis (NEC), hemorrhage (HEM), inflammatory cell infiltrate (INF), glandular secretion (SEC), tubular areas (ATUB), nuclear grade (GNUC), number of mitosis (NMIT), tumor grade (GRA), stromal reaction (REST) and apoptosis (APO). Inspection of the data set revealed 30 different combinations established between the different structural patterns.

## 3. Results and discussion

A comprehensive characterization of tumors was performed to explore the trends and relations between a set of multiple ranged parameters and grading. The tumoral invasiveness was marginal, and no metastases were found on necropsy examination performed at sacrifices made 27 weeks after carcinogenic induction. The tumor occurrences and some of the microscopic parameters considered in the work together with a brief description of the scores and labels employed are summarized in Table 1.

The system is composed of 148 mammary tumors developed by 75% of rats. There are 70 (47.3%) cases corresponding to tumors of grade 1, 60 (40.5%) cases of tumors with grade 2 and 18 (12.2%) tumors of grade 3. In the global set, 93% of the tumoral lesions are invasive carcinomas. Most of the invasive tumors are of low and intermediate grade and 12% are high grade tumors. All non-invasive

tumors, benign and *in situ* carcinomas, are of grade 1. In this data set, all variables are discrete and depart from a normal behavior. The number of structural patterns present on a tumor tends to correlate with malignant features. Most of single pattern tumors are of low grade, two and three pattern tumors are almost equally distributed on low and moderate grade. Tumors with four and five patterns are highly graded. The presence of the cribriform pattern is correlated with grading, i.e., tumors having a higher predominance of cribriform pattern are likely to be more malignant. Microcalcifications are present in 12% of the tumoral lesions and do not relate to tumoral grade. In contrast, necrosis and hemorrhage intensity are related with more malignant tumors, being the influence of the latter more moderate. The absence of necrosis is associated with low grade tumors. All high grade tumors have inflammatory infiltrating cells, usually lymphocytes and eosinophils, but the intensity of these cell populations as well as secretion does not correlate with malignancy. Fig. 1 shows a representative picture of some patterns.

### 3.1. Discrimination based on histological factors

PCA fulfils most of exploratory data analysis (EDA) tasks, and provides relevant data visualization. Table 2 contains the main results of the PCA analysis, using the correlation matrix. The amount of variation explained by each PC is measured by the eigenvalues and will be largest for the first PC and increasingly smaller for the subsequent PCs. An eigenvalue larger than 1 indicates that the corresponding PC accounts for more variance than that present, on average, after standardization of the original variables. This is commonly used as a threshold for which PCs are to be retained. It is seen that the first two and three principal components are able to recover ca. 43% and 53% respectively, of the data variability, indicating that a graphical representation of the tumors based on these two or three components is clearly meaningful.

From Kaiser's criterion (Jolliffe, 2002) at least four components are required. However, it takes eight components to explain up to 80% of the variance in data. The question arises if all these additional components have relevant information.

The scree graph of the eigenvalues, presented in Fig. 2, shows that the first component is in fact the most relevant to describe the data correctly, suggesting a one-factor solution.

In order to retrieve some relevant information for the most discriminant variables, Table 3 presents the relevance of each parameter, in the first two components ( $PC_1$  and  $PC_2$ ). The criterion for selecting a significant loading is based on the comparison to the average value in each component, i.e., the loading is simply considered significant if above the average value, and not significant otherwise (Cova et al., 2013).

Fig. 3 depicts the contribution of each variable for discrimination of tumors. The first component retains mainly information over tumor and nuclear grade (GRA and GNUC, respectively), hemorrhage (HEM), number of mitosis (NMIT), tubular areas (ATUB), necrosis (NEC), predominance of the cribriform pattern (PCRI) and stromal invasion (IEST). Correlation coefficients indicate that these variables are related, being the highest correlation (0.74) between the tumor grade (GRA) and the number of mitosis (NMIT), which stems from the fact that NMIT is used to define the grade. The lowest correlation (0.51) is established between nuclear grade (GNUC) and stromal invasion (IEST). Further observations can be drawn from the fact that the values of this first component are all positive, suggesting that it represents a measure of the degree of malignancy. High values of these characteristics suggest the increasing of malignancy. The second component is mostly related with glandular secretion (SEC), microcalcifications (MIC),

**Table 1**  
Overview of the data under study, with indication of the number of benign, malignant *in situ* and malignant invasive tumors and the percentage of each score value attributed to the histological factors.

	Scores	N <sup>b</sup> (%)	Grade 1 (70) (%)	Grade 2 (60) (%)	Grade 3 (18) (%)
Type of tumor <sup>a</sup>	1	4.10	8.60	0.00	0.00
	2	2.70	5.70	0.00	0.00
	3	93.2	85.7	100	100
Number of structural patterns (NPAD)	1	31.1	37.1	26.7	22.2
	2	47.3	45.7	50.0	44.4
	3	19.6	17.1	20.0	27.8
	4	0.86	0.00	1.67	0.00
	5	1.35	0.00	1.67	5.56
Predominance of cribriform pattern (PCRI) <sup>c</sup>	1	44.6	68.6	26.7	11.1
	2	33.1	24.3	43.3	33.3
	3	22.3	7.14	30.0	55.6
Microcalcifications (MIC) <sup>d</sup>	0	87.8	85.7	88.3	94.4
	1	12.2	14.3	11.7	5.56
Necrosis (NEC) <sup>e</sup>	0	24.3	40.0	11.7	5.56
	1	34.5	40.0	28.3	33.3
	2	37.2	17.1	53.3	61.1
	3	4.05	2.86	6.67	0.00
Hemorrhage (HEM) <sup>e</sup>	0	37.2	57.1	23.3	5.56
	1	47.3	38.6	55.0	55.6
	2	14.9	4.29	20.0	38.9
	3	0.68	0.00	1.67	0.00
Inflammatory cell infiltration (INF) <sup>e</sup>	0	8.78	15.7	3.33	0.00
	1	50.7	52.9	50.0	44.4
	2	35.8	27.1	41.7	50.0
	3	4.73	4.29	5.00	5.56
Glandular secretion (SEC) <sup>e</sup>	0	32.4	34.3	30.0	33.3
	1	43.2	37.1	46.7	55.6
	2	18.2	21.4	18.3	5.56
	3	6.08	7.14	5.00	5.56
Stromal reaction (EST) <sup>e</sup>	0	2.70	2.86	3.33	0.00
	1	44.6	44.3	45.0	44.4
	2	43.2	38.6	45.0	55.6
	3	9.46	14.3	6.67	0.00

<sup>a</sup> Benign (1), malignant *in situ* (2), and malignant invasive (3).

<sup>b</sup> Percentage of tumor occurrences.

<sup>c</sup> Cribriform pattern predominance: 0–30% (1), 40–60% (2), and 70–100% (3).

<sup>d</sup> Absent (0) and present (1).

<sup>e</sup> Absent (0), mild (1), moderate (2) and intense (3).

apoptosis (APO) and the number of histological patterns (NPAD). SEC displays a highest correlation coefficient with MIC, still a small value (0.32). In what follows, the first component ( $PC_1$ ) will be denoted as a measure of malignancy, reflecting the inter-grade discrimination. The second component reflects the direction along which the intra-grade discrimination occurs.

**Table 2**  
Eigenvalues and data recovery evolution with respect to the number of principal components for the original data set composed by 148 mammary tumors.

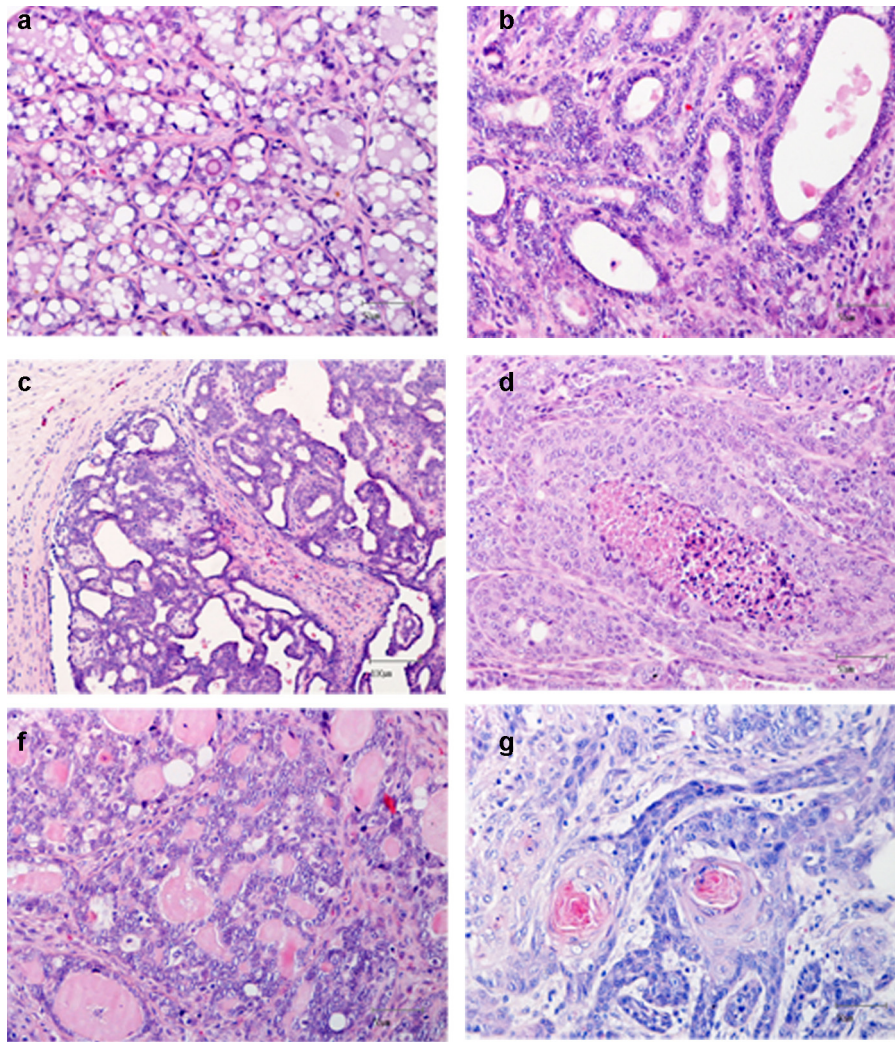
Principal components	Eigenvalues ( $\lambda_i$ )	Explained variance (%)	Cumulative explained variance (%)
$PC_1$	4.21	30.1	30.1
$PC_2$	1.76	12.6	42.7
$PC_3$	1.51	10.8	53.4
$PC_4$	1.12	8.00	61.5
$PC_5$	0.88	6.30	67.7
$PC_6$	0.82	5.88	73.6
$PC_7$	0.79	5.63	79.3
$PC_8$	0.66	4.69	83.9
$PC_9$	0.56	4.03	88.0
$PC_{10}$	0.54	3.87	91.9
$PC_{11}$	0.41	2.91	94.8
$PC_{12}$	0.38	2.73	97.5
$PC_{13}$	0.27	1.90	99.4
$PC_{14}$	0.09	0.62	100

### 3.2. Inter- and intra-grade discrimination

In order to characterize inter- and intra-grade discrimination the analysis proceeds, separately, to the cases related to each tumor pattern. The distributions of the tumors on both directions depend on the score levels of the parameters of highest contribution in the inter- and intra-grade components, GRA and SEC, respectively.

**Table 3**  
Contributions of histological parameters on the first two principal components ( $PC_1$  and  $PC_2$ ). The most relevant contributions are highlighted.

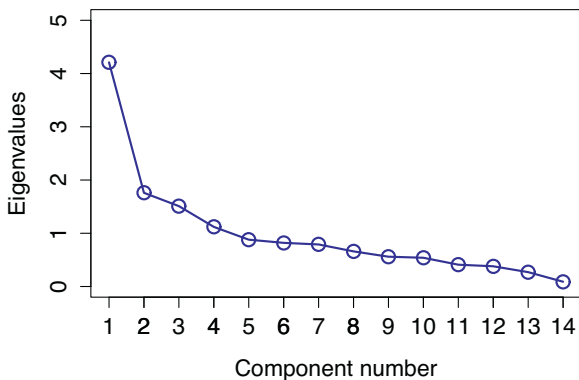
	$PC_1$	$PC_2$
NPAD	0.194	<b>-0.305</b>
PCRI	<b>0.312</b>	0.022
MIC	-0.062	<b>-0.464</b>
IEST	<b>0.282</b>	0.231
NEC	<b>0.320</b>	-0.152
HEM	<b>0.335</b>	-0.107
INF	0.255	-0.161
SEC	-0.026	<b>-0.606</b>
ATUB	<b>0.322</b>	0.059
GNUC	<b>0.352</b>	0.170
NMIT	<b>0.327</b>	-0.141
GRA	<b>0.399</b>	0.008
REST	0.051	0.161
APO	0.067	<b>0.364</b>



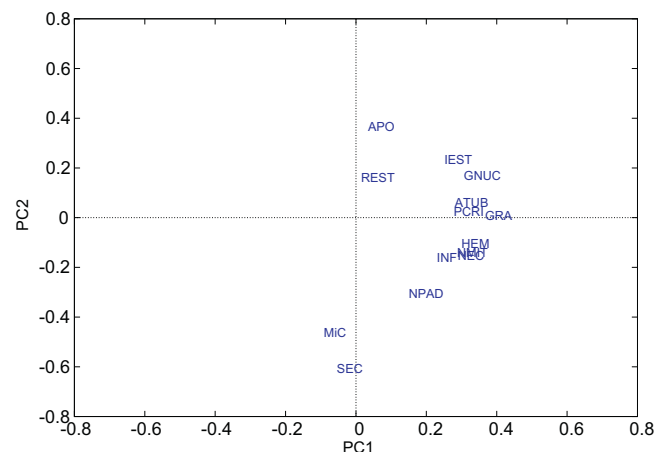
**Fig. 1.** Representative images of some structural patterns present in the rat mammary gland: (a) lactating adenoma with secretion of level 3 (HE, 200×); (b) tubular pattern with mild secretion of level 1 (HE, 200×); (c) papillary pattern (HE, 100×); (d) tumoral necrosis exhibiting the appearance of a comedocarcinoma (HE, 200×); (e) cribriform pattern with high secretion (HE, 200×); and (f) squamous pattern with epithelial gland and squamous components (HE, 200×).

Fig. 4 represents, on the basis of frequency histograms, the projection of the tumors on the inter-grade component. The discrimination is clearly made along this component, in spite of the fact that there is a marked overlap between tumors of grades 1

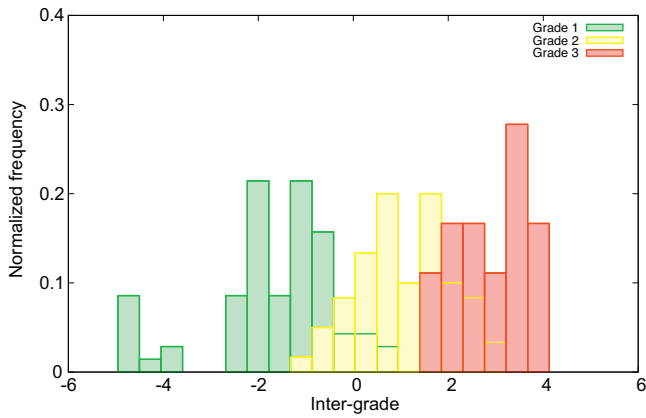
and 2, and also of grades 2 and 3. The overlapping areas correspond to transition areas between tumor grades. This means that the cases included in these areas will have a greater tendency to evolve towards malignancy.



**Fig. 2.** Scree plot of eigenvalues for the 14 histological parameters considered. Principal components are sorted by decreasing fraction of total variance explained.



**Fig. 3.** Loadings of histological observations in the two components of PCA (cumulative variance of 43%).

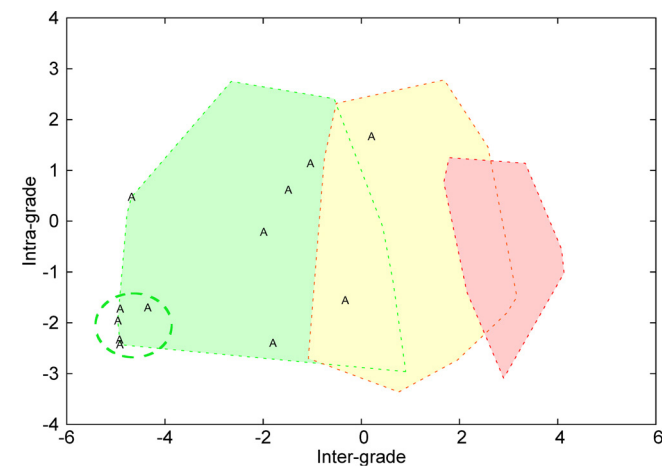


**Fig. 4.** Normalized frequency histograms for the projection of tumors according to its specific grade. Green refers to tumors of grade 1, while yellow and red corresponds to grades 2 and 3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

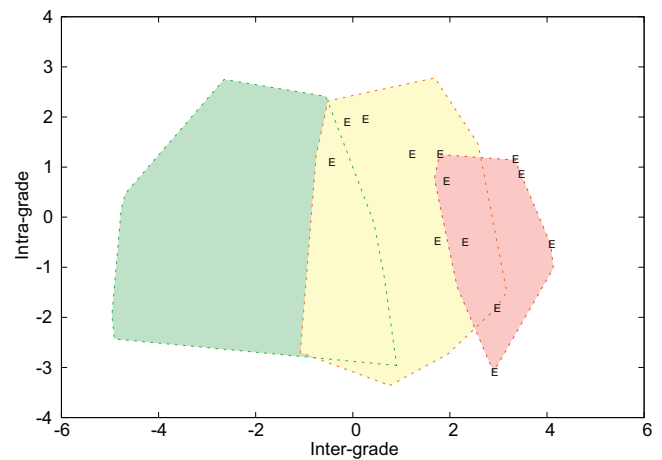
Fig. 5 presents the cases corresponding to adenomas in a plane formed by inter- and intra-grade components. Adenomas are essentially benign and associated to tumors of grade 1. A few are present in the region of grade 2, but these are combined with other patterns. They span a significant range in the intra-grade component, which means that there is some variation in secretion degree. As expected, lactating adenomas (circled in green) exhibit a typical behavior, being usually associated with high levels of secretion. Observations related to necrosis (NEC), hemorrhage (HEM), and stromal invasion (IEST) are absent.

Let us now focus on the most aggressive tumors, that include squamous and comedo types. Comedo cases emerge on mixed patterns while squamous appear mostly as single pattern. Both are almost equally distributed through grades 2 and 3. Fig. 6 provides a simple bi-dimensional representation of the tumors with squamous pattern. On the top, are placed those cases of single squamous pattern, which have negligible or no secretion. The scattering to the bottom is related with variations in apoptosis and microcalcification levels and also with the presence of other patterns.

Regarding the discrimination between tumors of the same single pattern, more observations can be conveyed. The analysis is achieved in two ways, either describing tumors of the same grade, or of different grades. Fig. 7 depicts the cribriform tumor



**Fig. 5.** Relative position of adenomas on the inter- and intra-grade components. Lactating adenomas are circled in green. Colours are related to the increasing of tumor grade from grade 1 (in green) to grade 3 (in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



**Fig. 6.** Representation of squamous tumors on the inter- and intra-grade components with 43% of information recovery. Secretion is not detected for tumors on the top region.

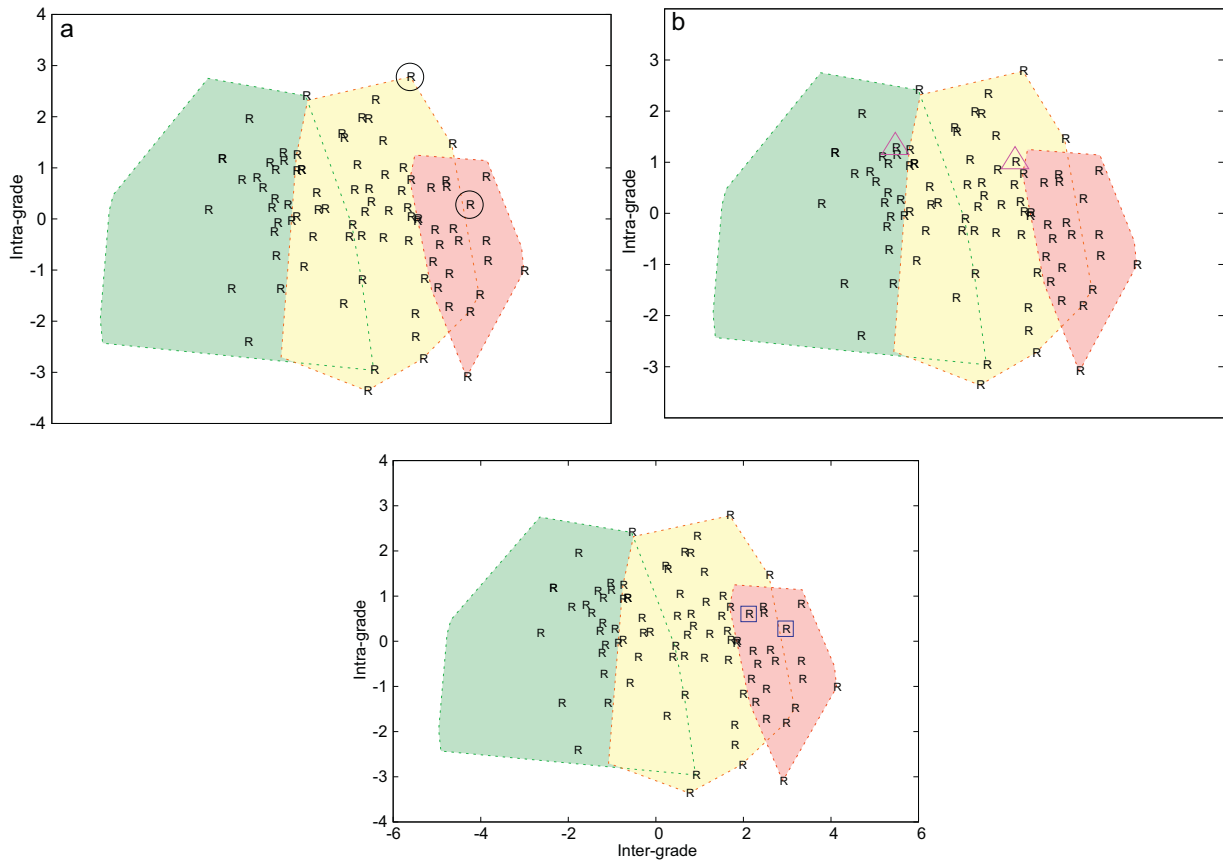
occurrences and the specific cases corresponding to this single pattern are highlighted. In panel (a), both cases have the same levels of PCRI, NEC, GNUC, ATUB and MIC and differ only in NMIT, HEM, APO and SEC. The tumoral position on the top may be explained by the absence of secretion and the presence of high level of apoptosis. This case is placed in the group of intermediate severity (grade 2) due to the lower number of mitosis. Thus, we can state that the inter-grade discrimination of the cribriform pattern lies essentially in the number of mitosis. In panel (b), the tumor included in grade 1 may have necrosis of low level, but the presence of necrosis might be accompanied by a lower level of mitosis. This fact can probably justify the respective positioning in grade 1. The tumor in the group of intermediate severity (grade 2), does not possess necrosis but the absence of this characteristic is compensated by the increased levels of mitosis. In panel (c), the difference between tumors of the same grade is also related to the number of mitosis.

Some interesting remarks can also be reported considering the papillary and tubular patterns. It can be argued that the lipid-rich papillary subtype has the highest level of secretion. The difference between the lipid-rich papillary of grades 1 and 2 lies on the presence of tubular areas (ATUB) and necrosis (NEC) of an intermediate level. If the lipid-rich papillary and the cribriform types are combined, necrosis is absent. Those cases present secretion of low level, and a moderate level of mitosis may probably replace the absence of necrosis. This fact explains the position of the lipid-rich papillary in the group of grade 2. Necrosis and hemorrhage can discriminate tubular tumors within the same group. Among groups of distinct grade, the difference is again probably related to the number of mitosis. The ability of the selected variables to be used for inter- and intra-grade discrimination is summarized in Table 4.

Once identified the factors that contribute most to the discrimination of tumor patterns, we can now assess the overall profile in the global set. Fig. 8 depicts a composed view of the cases pertaining to each structural type (adenoma, papillary, tubular, cribriform,

**Table 4**  
Parameters responsible for discrimination of tumors with similar or distinct severity.

	Similar grade	Different grade
Adenomas	SEC, MIC	–
Papillary	SEC	ATUB, NEC
Tubular	NEC, HEM	NMIT
Cribriform	NMIT	NMIT, NEC
Squamous	SEC, MIC	NEC, HEM, APO
Comedo	NEC, HEM	NMIT



**Fig. 7.** Representation of the cribriform cases in two dimensions, showing three distinct situations: (a) same pattern in groups of distinct severity and different level of secretion, (b) same pattern in groups with distinct severity and different number of mitosis, and (c) same pattern and severity but different number of mitosis.

comedo and squamous), taking into account the single and mixed arrangement. As already known, the tumors of grade 1 are generally considered the least aggressive in behavior. Tumors of grades 1 and 2 are in larger diffuse groups, ranging from one to three combined patterns. Tumors of grade 3 correspond to a small group and their structure varies from one single pattern to five combined patterns.

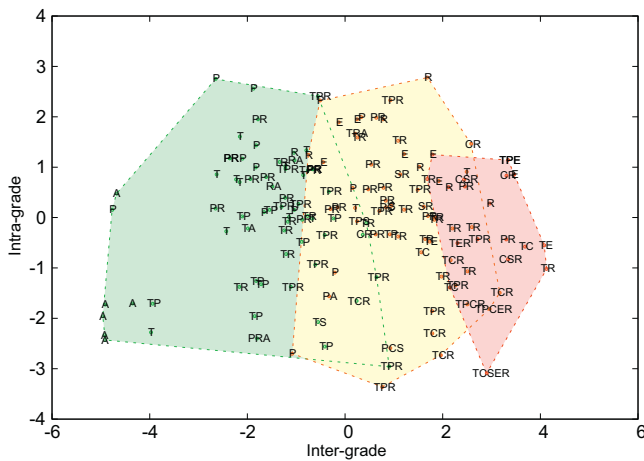
3.3. Overview

The results have shown that PCA provides the means to carry out an in depth analysis of data originating from multiple ranged

microscopic parameters of rat mammary tumors induced by DMBA. The proposed approach allows a direct visualization of the data structure and also provides the relative positioning of the different groups, emphasizing the underlying patterns. The heterogeneity of the tumor based on the number of histological patterns is not a relevant parameter, but the prevalence of the cribriform pattern is important to distinguish malignancy, being a potential factor of discrimination. Another relevant factor is the presence of gland secretion levels, which is a benign feature in all tumors. Such information is already known, but now we can discriminate the relevance of secretion depending on the histological type and grade. Regarding necrosis and hemorrhage, these parameters are pertinent to distinguish the degree of malignancy between tumors of the same grade, especially if they are composed by papillary or tubular types, or even both. Remembering the concerns detailed previously, the proposed numerical analysis improves the NGS. Moreover, the analysis is extremely informative and provides a rationale for the observations.

4. Conclusions

In this work we have proposed a simple and objective approach to improve tumor grading in rats and define informative parameters for malignancy prediction, not only on invasive carcinomas, but on all type of mammary tumors. The analysis reflects tumor heterogeneity building a low-dimensional map allowing the comparison of *in vivo* mammary tumor models. This is particularly important on this model type since it is commonly used for prevention and therapeutic studies of human interest. The discrimination of chemically induced tumors is mapped over relevant parameters for malignancy prediction, being tumor histological type the most important. On each histological type, some parameters correlate



**Fig. 8.** Tumors mapped onto discrimination components, according to the histological pattern (A – adenoma, P – papillary, T – tubular, R – cribriform, C – comedo and E – squamous).

with the grade of the tumor. The prevalence of cribriform pattern is an important factor of discrimination. Necrosis and hemorrhage are pertinent to distinguish potential malignancy between tumors of the same grade, especially if they are composed by papillary or tubular types. The presence of secretion levels is a benign feature responsible essentially for discrimination between tumors of the same type. In sum, the present analysis represents a new practical, reliable and economic approach which improves traditional histopathology.

### Acknowledgement

This work was supported by Fundação para a Ciência e a Tecnologia, Portugal (Grant number: SFRH/BD/74850/2010).

### References

- Brereton R. *Chemometrics: applications of mathematics and statistics to laboratory systems*. Chichester: E. Horwood; 1990].
- Campanella L, De Angelis G, Visco G. Chemometric investigation of the efficiency of different TiO<sub>2</sub>-based catalysts as principal components of TOC photochemical sensors under development. *Analytical and Bioanalytical Chemistry* 2003;376(4):467–75.
- Cardiff R. Validity of mouse mammary tumor models for human breast cancer: comparative pathology. *Microscopy Research and Technique* 2001;52(2):224–30.
- Cardiff R, Anver M, Gusterson B, Hennighausen L, Jensen R, Merino M, et al. The mammary pathology of genetically engineered mice: the consensus report and recommendations from the annapolis meeting. *Oncogene* 2000;19(8):968–88.
- Cardiff R, Munn R. Comparative pathology of mammary tumorigenesis in transgenic mice. *Cancer Letters* 1995;90(1):13–9.
- Cheng L, Burns M, Taylor J, He W, Halpern E, McDougal W, et al. Metabolic characterization of human prostate cancer with tissue magnetic resonance spectroscopy. *Cancer Research* 2005;65(8):3030–4.
- Costa I, Solanas M, Escrich E. Histopathologic characterization of mammary neoplastic lesions induced with 7,12 dimethylbenz(alpha)anthracene in the rat: a comparative analysis with human breast tumors. *Archives of Pathology and Laboratory Medicine* 2002;126(8):915–27.
- Cova T, Pereira J, Pais A. Is standard multivariate analysis sufficient in clinical and epidemiological studies? *Journal of Biomedical Informatics* 2013;46(1):75–86.
- Currier N, Solomon S, Demicco E, Chang D, Farago M, Ying H, et al. Oncogenic signaling pathways activated in DMBA-induced mouse mammary tumors. *Toxicologic Pathology* 2005;33(6):726–37.
- Dias M, Sousa E, Cabrita S, Patrício J, Oliveira C. Chemoprevention of DMBA-induced mammary tumors in rats by a combined regimen of alpha-tocopherol, selenium, and ascorbic acid. *The Breast Journal* 2001;6(1):14–9.
- Edefonti V, Bravi F, Garavello W, La Vecchia C, Parpinel M, Franceschi S, et al. Nutrient-based dietary patterns and laryngeal cancer: evidence from an exploratory factor analysis. *Cancer Epidemiology Biomarkers & Prevention* 2010;19(1):18–27.
- Espeland M, Bray G, Neiberg R, Rejeski W, Knowler W, Lang W, et al. Describing patterns of weight changes using principal components analysis: results from the action for health in diabetes (look ahead) research group. *Annals of Epidemiology* 2009;19(10):701–10.
- Gualberto A, Gustavson M, Dolled-Filhart M, Christiansen J, Wang Y, Hixon M, et al. Molecular analysis of non-small cell lung cancer identifies subsets with different sensitivity to insulin-like growth factor i receptor inhibition. *Clinical Cancer Research* 2010;16(18):4654–65.
- Jackson J, Wiley J. *A user's guide to principal components*. New York: Wiley Online Library; 1991].
- Jolliffe I. *Principal component analysis*, vol. 2. New York: Springer; 2002].
- Kokot S, Grigg M, Panayiotou H, Phuong T. Data interpretation by some common chemometrics methods. *Electroanalysis* 1998;10(16):1081–8.
- Latifoğlu F, Polat K, Kara S, Güneş S. Medical diagnosis of atherosclerosis from carotid artery doppler signals using principal component analysis (PCA), k-NN based weighting pre-processing and artificial immune recognition system (airs). *Journal of Biomedical Informatics* 2008;41(1):15–23.
- Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics – a review in human disease diagnosis. *Analytica Chimica Acta* 2010;659(1–2):23–33.
- Martins F, Teixeira F, Reis I, Galdes N, Cabrita A, Dias M. Chemoprevention and mammary neoplastic aggressiveness. *The Breast Journal* 2008;14(6):608–9.
- Massart D. *Chemometrics: a textbook*, vol. 2. New York: Elsevier Science; 1988].
- Medina D. Preneoplastic lesions in murine mammary cancer. *Cancer Research* 1976;36(7 Part 2):2589–95.
- Méndez E, Houck J, Doody D, Fan W, Lohavanichbut P, Yueh, et al. A genetic expression profile associated with oral cancer identifies a group of patients at high risk of poor survival. *Clinical Cancer Research* 2009;15(4):1353–61.
- Meyer J, Alvarez C, Milikowski C, Olson N, Russo I, Russo J, et al. Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Modern Pathology* 2005;18(8):1067–78.
- Polat K, Güneş S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing* 2007;17(4):702–10.
- Polat K, Güneş S. Computer aided medical diagnosis system based on principal component analysis and artificial immune recognition system classifier algorithm. *Expert Systems with Applications* 2008;34(1):773–9.
- Rakha E, Reis-Filho J, Baehner F, Dabbs D, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research* 2010;12(4):207.
- Reich D, Price A, Patterson N. Principal component analysis of genetic data. *Nature Genetics* 2008;40(5):491–2.
- Russo I, Russo J. Mammary gland neoplasia in long-term rodent studies. *Environmental Health Perspectives* 1996;104(9):938–67.
- Russo J, Russo I. Atlas and histologic classification of tumors of the rat mammary gland. *Journal of Mammary Gland Biology and Neoplasia* 2000;5(2):187–200.
- Russo J, Russo I. Experimentally induced mammary tumors in rats. *Breast Cancer Research and Treatment* 1996;39(1):7–20.
- Selaru F, Yin J, Olaru A, Mori Y, Xu Y, Epstein S, et al. An unsupervised approach to identify molecular phenotypic components influencing breast cancer features. *Cancer Research* 2004;64(5):1584–8.
- Spratlin J, Serkova N, Eckhardt S. Clinical applications of metabolomics in oncology: a review. *Clinical Cancer Research* 2009;15(2):431–40.
- Tavassoli F, Devilee P. *Pathology and genetics of tumors of the breast and female genital organs*, vol. 4. Lyon: World Health Organization; 2003].
- Thompson H, Singh M. Rat models of premalignant breast disease. *Journal of Mammary Gland Biology and Neoplasia* 2000;5(4):409–20.
- Wang X, Paliwal K. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition* 2003;36(10):2429–39.
- Webster J, Dennis M, Dervisis N, Heller J, Bacon N, Bergman P, et al. Recommended guidelines for the conduct and evaluation of prognostic studies in veterinary oncology. *Veterinary Pathology Online* 2011;48(1):7–18.
- Withrow S, Vail D. *Withrow and MacEwen's small animal clinical oncology*. St. Louis: S. Elsevier; 2007].
- Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 1987;2(1–3):37–52.
- Xie Y, Xiao G, Coombes K, Behrens C, Solis L, Raso G, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. *Clinical Cancer Research* 2011;17(17):5705–14.