Computers in Biology and Medicine 40 (2010) 318–330

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients

Daniele Soria [a,*], Jonathan M. Garibaldi [a], Federico Ambrogi [c], Andrew R. Green [b], Des Powe [b], Emad Rakha [b], R. Douglas Macmillan [f], Roger W. Blamey [f], Graham Ball [d], Paulo J.G. Lisboa [e], Terence A. Etchells [e], Patrizia Boracchi [c], Elia Biganzoli [c], Ian O. Ellis [b]

[a] School of Computer Science, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, UK
[b] School of Molecular Medical Sciences, Nottingham University Hospitals and University of Nottingham, Queens Medical Centre, Derby Road, Nottingham NG7 2UH, UK
[c] Institute of Medical Statistics and Biometry, University of Milan, Via Venezian 1, 20133 Milan, Italy
[d] School of Science and Technology, Nottingham Trent University, Clifton Campus, Clifton Lane, Nottingham NG11 8NS, UK
[e] School of Computing and Mathematical Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
[f] The Breast Institute, Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB, UK

ARTICLE INFO

Article history:
Received 18 June 2009
Accepted 4 January 2010

Keywords:
Breast cancer
Molecular classification
Clustering methods
Consensus clustering
Validity indices

ABSTRACT

Single clustering methods have often been used to elucidate clusters in high dimensional medical data, even though reliance on a single algorithm is known to be problematic. In this paper, we present a methodology to determine a set of 'core classes' by using a range of techniques to reach consensus across several different clustering algorithms, and to ascertain the key characteristics of these classes. We apply the methodology to immunohistochemical data from breast cancer patients. In doing so, we identify six core classes, of which several may be novel sub-groups not previously emphasised in literature.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer, the most common cancer in women [1,2], is a complex disease characterised by multiple molecular alterations. Current routine clinical management relies on availability of robust clinical and pathologic prognostic and predictive factors to support decision making. Recent advances in high-throughput molecular technologies supported the evidence of a biologic heterogeneity of breast cancer.

Following the seminal paper of Eisen and colleagues [3], in which hierarchical clustering and visual inspection of the dendrogram were performed to discover unknown pattern of gene associations, the use of clustering has become more and more popular, especially for discovering profiles in cancer with respect to high-throughput genomic data. Perou et al. [4] identified four molecular distinct breast cancer groups based on gene expression profiles using a hierarchical clustering algorithm: luminal epithelial/estrogen (ER) positive, HER2 positive, basal-like and normal breast-like. A subsequent study extended this by dividing the luminal/ER positive group into three subtypes: luminal-A, B, and C [5], but the luminal-C group was later eliminated [6]. Sotiriou et al. [7] showed six similar groups, with two basal-like subgroups and no normal breast-like group. Whilst numerous studies have reported these and other novel molecular subtypes, and assigned a prognostic significance to the proposed classes [8–10], they remain varied in their detailed classification [11]. An alternative approach to gene expression profiling is to use established robust laboratory technology, such as immunocytochemistry on formalin fixed paraffin embedded patient tumour samples. We and others have applied protein biomarker panels with known relevance to breast cancer, to large numbers of cases using tissue microarrays, exploring the existence and clinical significance of distinct breast cancer classes [12–19]. In particular, in [12] five breast cancer classes were identified and characterised. Note that a sixth group of only four cases was also identified but considered too small for further detailed assessment. However, these studies have not addressed the stability of the proposed classifications across different case sets, assay methods and data analysis procedures. Such an issue appears of critical relevance considering the increase in the number of features involved in bionformatics analyses.

In order to deal with the stability of classifications and in particular of clustering techniques, several studies have focused on the comparison and concordance among different clustering methods defining what is now known as the 'consensus clustering'. Monti and colleagues presented a new methodology of class

* Corresponding author. Tel.: +44 115 95 14229.
E-mail address: dqs@cs.nott.ac.uk (D. Soria).

0010-4825/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2010.01.003

discovery and clustering validation tailored to the task of analysing gene expression data [20]. The new methodology, termed 'consensus clustering', provides a method, in conjunction with resampling techniques, to represent the consensus across multiple runs of a clustering algorithm and to assess the stability of the discovered clusters. The basic assumption of this method was the following: if the data represent a sample of items drawn from distinct sub-populations, and if a different sample drawn from the same sub-populations were to be observed, the induced cluster composition and number should not be radically different. Therefore, the more the attained clusters are robust to sampling variability, the more one can be confident that these clusters represent real structure.

Swift and colleagues used consensus clustering to improve confidence in gene-expression analysis, on the assumption that microarray analysis using clustering algorithms can suffer from lack of inter-method consistency in assigning related gene-expression profiles to clusters [21]. To assess gene-expression cluster consistency, the use of the weighted-kappa metric was analysed. This metric rates the agreement between the classification decisions made by two or more observers. In this case the two observers are the clustering methods.

Filkov and Skiena proposed a methodology for consensus clustering as an approach to integrating diverse sources of similarity clustered microarray data [22]. They proposed to exploit the popularity of cluster analysis of biological data by integrating clusterings from existing data sets into a single representative clustering based on pairwise similarities of the clusterings. Under reasonable conditions, the consensus cluster should provide additional information to that of the union of individual data analyses. The goals of consensus clustering are to integrate multiple data sets for ease of inspection, and to eliminate the likely noise and incongruencies from the original classifications. In terms of similarity the consensus partition should be close to all given ones, or in terms of distance, it must not be too far from any of them. One way to do this is to find a partition that minimises the distance to all the other partitions. So, given k different partitions, the target one was identified as the consensus partition.

In another approach [23], robust clusters were identified by the implementation of a new algorithm termed 'Clusterfusion'. 'Clusterfusion' takes the results of different clustering algorithms and generates a set of robust clusters based upon the consensus of the different results of each algorithm. Firstly, an agreement matrix was generated with each cell containing the number of agreements amongst methods for clustering together the two variables represented by the indexing row and column indices. This matrix was then used to cluster variables based upon their cluster agreement. In essence, a clustering technique was applied to the clustering results.

The idea of combining and comparing the results of different clustering algorithms is particularly important in order to evaluate the stability of a proposed classification. In this paper, a methodology is presented to evaluate the stability of six breast cancer classes by comparing the clustering solutions provided by different algorithms. In order to address the standard problem of consensus clustering in which the label of classes is arbitrary, a label was assigned using the six clusters characterised in the work of Abd El-Rehim [12], as a reference for the description of our resulting groups.

2. Material and methods

The four-step methodology for elucidating core, stable classes (groups) of data from a complex, multi-dimensional dataset was as follows:

1. A variety of clustering algorithms were run on the data set (see Section 2.1).
2. Where appropriate, the most appropriate number of clusters was investigated by means of cluster validity indices (see Section 2.2).
3. Concordance between clusters, assessed both visually and statistically, was used to guide the formation of stable 'core' classes of data.
4. A variety of methods were utilised to characterise the elucidated core classes.

The methodology was applied to a well-known set of data concerning breast cancer patients [12] (see Section 2.5) in order to obtain core classes. Once these core classes were obtained, the clinical relevance of the corresponding patient groups were investigated by means of associations with related patient data. All statistical analysis was done using *R*, a free software environment for statistical computing and graphics [24].

2.1. Clustering algorithms

Five different algorithms were used for cluster analysis:

(i) Hierarchical (as per our previous study [12]).
(ii) K-means (KM).
(iii) Partitioning around medoids (PAM).
(iv) Adaptive resonance theory (ART).
(v) Fuzzy c-means (FCM).

2.1.1. Hierarchical clustering

The hierarchical clustering algorithm (HCA) begins with all data considered to be in a separate cluster. It then finds the pair of data with the minimum value of some specified distance metric; this pair is then assigned to one cluster. The process continues iteratively until all data are in the same (one) cluster. A conventional hierarchical clustering algorithm (HCA) was utilised, utilising Euclidean distance on the raw (unnormalised) data with all attributes equally weighted.

2.1.2. K-means clustering

The K-means (KM) technique aims to partition the data into K clusters such that the sum of squares from points to the assigned cluster centres is minimised. The algorithm repeatedly moves all cluster centres to the mean of their Voronoi sets (the set of data points which are nearest to the cluster centre). The objective function minimised is

$$J(V) = \sum_{j=1}^{k} \sum_{i=1}^{c_j} \|x_i - v_j\|^2$$

where x_i is the i-th datum, v_j is the j-th cluster centre, k is the number of clusters, c_j is the number of data points in the cluster j and $\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j.

The j-th centre v_j can be calculated as

$$v_j = \frac{1}{c_j} \sum_{i=1}^{c_j} x_i, \quad j = 1, \ldots, k$$

K-means clustering is dependent on the initial cluster centres setting (which, in turn, determines the initial cluster assignment). Various techniques have been proposed for the initialisation of clusters [25], but for this study we used a fixed initialisation of the cluster centres obtained with hierarchical clustering. The number of clusters is an explicit input parameter to the K-means algorithm.

2.1.3. Partitioning around medoids

The partitioning around medoids (PAM) algorithm (also known as the k-medoids algorithm) is a technique which attempts to

minimise the distance between points labelled to be in a cluster and a point designated as the centre of that cluster. In contrast to the K-means algorithm, PAM chooses data points as centres (the so-called medoids) and then assigns each point to its nearest medoid. A medoid is defined as the object within a cluster for which the average dissimilarity to all other objects in the cluster is minimal, i.e. it is the most centrally located datum in the given cluster. Dissimilarities are nonnegative numbers that are small (close to zero) when two data points are 'near' to each other and become large when the points are very different [26]. Usually, a Euclidean metric is used for calculating dissimilarities between observations.

The algorithm consists of two phases: the *build* phase in which an initial set of k representative medoids is selected and the *swap* phase in which a search is carried out to improve the choice of medoids (and hence the cluster allocations). The algorithm is described in detail in [26, pp. 102–104]. The number of clusters is an explicit input parameter to the PAM algorithm.

2.1.4. Adaptive resonance theory

The adaptive resonance theory (ART) algorithm has three main steps [27]. First, the data are normalised to a unit hypersphere, thus representing only the ratios between the various dimensions of the data. Second, data allocated to each cluster are required to be within a fixed maximum solid angle of the group mean, controlled by a so-called 'vigilance parameter' ρ, namely $X_k \cdot P^i \leq \rho$. However, even when the observation profile and a prototype are closer than the maximum aperture for the group, a further test is applied to ensure that the profile and prototype have the same dominant covariates. This is done in a third step by specifying the extent to which the nearest permissible prototype allocation for the given observation must be on the same side of the data space from the diagonal comprising a vector of ones, $\hat{1}$, using a pre-set parameter, λ:

$$X_k \cdot P^i \leq \lambda X_k \cdot \hat{1}$$

The ART algorithm is initialised with no prototypes and creates them during each successive pass over the data set. It has some, limited, sensitivity to the order in which the data are presented and converges in a few iterations. In the ART algorithm the clusters are determined automatically: the number of clusters is not an explicit parameter, although there are parameters that can adjust the number obtained.

2.1.5. Fuzzy c-means

The fuzzy c-means (FCM) algorithm is a generalisation of the K-means algorithm which is based on the idea of permitting each object to be a member of *every* cluster to a certain degree, rather than an object having to belong to only one cluster at any one time. It is based upon the concept of fuzzy logic promulgated by Zadeh [28] and aims to minimise the objective function

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{i,j})^m \|x_i - v_j\|^2$$

where n is the number of data points, x_i and v_j are the data points and cluster centres and $\mu_{i,j}$ is the membership degree of data x_i to the cluster centre v_j ($\mu_{i,j} \in [0,1]$). m is called the 'fuzziness index' and the value of $m = 2.0$ is usually chosen. An exhaustive description of this method can be found in [29]. As for K-means, the number of clusters is an explicit input parameter to FCM.

2.2. Cluster validity

Clustering validity is a concept that is used to evaluate the quality of clustering results. If the number of clusters is not known prior to commencing an algorithm, a cluster validity index may be used to determine the best number of clusters for the given data set. Although there are many variations of validity indices, they are all either based on considering the data dispersion in a cluster and between clusters, or considering the scatter matrix of the data points and the one of the clusters centres. In this study, the following indices were applied to those algorithms for which the number of clusters is an explicit parameter, over a range of number of clusters:

1. Calinski and Harabasz [30].
2. Hartigan [31].
3. Scott and Symons [32].
4. Marriot [33].
5. Trace W [34,35].
6. Trace $W^{-1}B$ [35].

For each index, the number of clusters to be considered was chosen according to the rule reported in Table 1 where i_n is the validity index value obtained for n clusters [36].

2.3. Derivation of classes

Concordance among solutions was evaluated using the Cohen's kappa coefficient κ [37]. This coefficient is a statistical measure of inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation since κ takes into account the agreement occurring by chance.

To enable cluster visualisation, the original data space (consisting of a large number of dimensions) was transformed by principal component analysis (PCA) [38], and then the points plotted at their projected position on axes of the first and second principal components. As PCA transforms data such that the first principal component (PC) carries the maximum amount of variance in the data and the second PC carries the next largest variance (etc.), such a plot provides a picture in which the clusters have been 'spread out' as much as possible.

The previously obtained clustering results from Abd El-Rehim and colleagues [12], the cluster validity indices (where appropriate), visualisation of the new clustering results themselves, and the concordance among clustering solutions were then all used heuristically to guide the formulation of a set of rules to define core class membership from the various cluster assignments.

2.4. Characterisation of classes

2.4.1. Class characterisation by visualisation

For inspection of the patient characteristics in each class, the distribution of each variable in the class was compared with its distribution in the total sample, using boxplots. A boxplot shows the median expression level (solid horizontal bar), the upper

Table 1
Different validity indices and their associated decision rules.

Index	Decision rule
Calinski and Harabasz	$\min_n((i_{n+1}-i_n)-(i_n-i_{n-1}))$
Hartigan	$\min_n((i_{n+1}-i_n)-(i_n-i_{n-1}))$
Scott and Symons	$\max_n(i_n-i_{n-1})$
Marriot	$\max_n((i_{n+1}-i_n)-(i_n-i_{n-1}))$
Trace W	$\max_n((i_{n+1}-i_n)-(i_n-i_{n-1}))$
Trace $W^{-1}B$	$\max_n(i_n-i_{n-1})$

quartile and lower quartile range (shaded grey bar), the highest non-outlier and lowest non-outlier (smaller ticks joined by dashed lines), and any outliers (open circles). For a full description of boxplots, including the statistical definition of outliers see, for example [39].

2.4.2. Class characterisation by OSRE (orthogonal search rule extraction)

Orthogonal search rule extraction (OSRE) [40] is a computationally efficient algorithm to search for hypercubes in data space, since they map directly onto Boolean rules. A general description of this method is given below, while a more detailed one can be found in [40]. This methodology initially returns a rule for each data point, which triggers a pruning process to keep only those rules which represent large proportions of the data in the clusters, and do so with minimal mixing between clusters. The result is a set of low-order rules containing the covariates that characterise the sub-group of the cluster. The proposed interpretation is that these rules identify the drivers for cluster allocation, which may vary across the cluster but are, in general, well-defined.

Note that this method contrasts with widely used rule induction methods in two ways: firstly, there are no univariate cut-offs for groups of data, as in OSRE a sequential univariate search is carried out at the level of each individual data point which returns a multivariate hyperbox around that point, without the need to partition the data along a sequence of univariate covariates; and secondly, that the rules are overlapping, rather than constrained to mutual exclusivity as is usually the case in rule tree induction.

2.4.3. Class characterisation by ANN (artificial neural networks)

A conventional multi-layer perceptron artificial neural network (MLP-ANN) model was utilised such that individual H-scores derived from the tissue microarray analysis of the clinical samples were set as inputs and the class was set as the output using Boolean notation (i.e. 1 represented membership of a given class, 0 represented non-membership). This allowed the identification of markers that drive membership of a given class and that discriminate the class from the others. A three-layer MLP-ANN (featuring eight nodes in the hidden layer) with a back-propagation algorithm and a sigmoid activation function was used. The approach used in this work is similar to the ones used in [41,42].

2.5. Patients and clinical methods

A series of 1076 patients from the Nottingham Tenovus Primary Breast Carcinoma Series presenting with primary operable (stages I, II and III) invasive breast cancer between 1986 and 1998 was used to evaluate the methodology. Immunohistochemical reactivity for 25 proteins, with known relevance in breast cancer including those used in routine clinical practice, were previously determined using standard immunocytochemical techniques on tumour samples prepared as tissue microarrays [12]. Levels of immunohistochemical reactivity were determined by microscopical analysis using the modified H-score (values between 0 and 300), giving a semiquantitative assessment of both the intensity of staining and the percentage of positive cells. For the intensity, a score of 0–3, corresponding to negative, weak, moderate and strong positivity, was recorded. In addition, the percentage of positive cells at each intensity category was estimated. The H-score is calculated as follows [43]

$$\begin{aligned} \text{H}-\text{score} = {} & (1 \times \% \text{ of cells stained at intensisty category 1}) \\ & + (2 \times \% \text{ of cells stained at intensisty category 2}) \\ & + (3 \times \% \text{ of cells stained at intensisty category 3}) \end{aligned} \quad (1)$$

The range of possible scores is thus 0–300, where 300 equals 100% of tumour cells stained strongly [44]. The complete list of variables used in this study is given in Table 2, while data extracted from three patients are reported as an example in Table 3.

This is a well-characterised series [12] of patients who were treated according to standard clinical protocols. Patient management was based on tumour characteristics using Nottingham Prognostic Index (NPI) and hormone receptor status. Patients with an NPI score ≤ 3.4 received no adjuvant therapy, those with a NPI score > 3.4 received hormone therapy if oestrogen receptor (ER) positive or classical cyclophosphamide, methotrexate and 5-fluorouracil (CMF) if ER negative and fit enough to tolerate chemotherapy. Hormonal therapy was given to 420 patients (39%) and chemotherapy to 264 (24.5%). This study was approved by the *Nottingham Research Ethics Committee 2* under the title 'Development of a molecular genetic classification of breast cancer'.

Table 2
Complete list of antibodies used and their dilutions.

Antibody, clone	Short name	Dilution
Luminal phenotype		
CK 7/8 [clone CAM 5.2]	CK7/8	1:2
CK 18 [clone DC10]	CK18	1:50
CK 19 [clone BCK 108]	CK19	1:100
Basal phenotype		
CK 5/6 [cloneD5/16134]	CK5/6	1:100
CK 14 [clone LL002]	CK14	1:100
SMA [clone 1A4]	Actin	1:2000
p63 ab-1 [clone 4A4]	p63	1:200
Hormone receptors		
ER [clone 1D5]	ER	1:80
PgR [clone PgR 636]	PgR	1:100
AR [clone F39.4.1]	AR	1:30
EGFR family members		
EGFR [clone EGFR.113]	EGFR	1:10
HER2/c-erbB-2	HER2	1:250
HER3/c-erbB-3 [clone RTJ1]	HER3	1:20
HER4/c-erbB-4 [clone HFR1]	HER4	6:4
Tumour suppressor genes		
p53 [clone DO7]	p53	1:50
nBRCA1 Ab-1 [clone MS110]	nBRCA1	1:150
Anti-FHIT [clone ZR44]	FHIT	1:600
Cell adhesion molecules		
Anti E-cad [clone HECD-1]	E-cad	1:10/20
Anti P-cad [clone 56]	P-cad	1:200
Mucins		
NCL-Muc-1 [clone Ma695]	MUC1	1:300
NCL-Muc-1 core [clone Ma552]	MUC1co	1:250
NCL muc2 [clone Ccp58]	MUC2	1:250
Apocrine differentiation		
Anti-GCDFP-15	GCDFP	1:30
Neuroendocrine differentiation		
Chromogranin A [clone DAK-A3]	Chromo	1:100
Synaptophysin [clone SY38]	Synapto	1:30

Table 3
H-score of each variable for three different patients.

Variable name	Patient 1	Patient 300	Patient 1061
CK7/8	200	280	100
CK18	300	190	0
CK19	200	145	0
CK5/6	0	0	125
CK14	0	0	210
Actin	0	0	0
p63	0	0	0
ER	60	110	0
PgR	0	0	0
AR	0	45	0
EGFR	0	0	0
HER2	300	0	0
HER3	170	72	250
HER4	100	130	150
p53	150	0	270
nBRCA1	200	45	0
FHIT	120	0	0
E-cad	100	115	0
P-cad	0	15	0
MUC1	180	255	160
MUC1co	210	210	80
MUC2	0	0	0
GCDFP	90	0	0
Chromo	0	0	0
Synapto	0	0	0

3. Results

3.1. Clustering results

3.1.1. HCA, K-means, PAM and ART

The HCA results from Abd El-Rehim et al. [12] were utilised, unaltered. Both the K-means and PAM algorithms were run with the number of clusters varying from 2 to 20, as the number of clusters is an explicit input parameter of the algorithms. Given that both algorithms can be sensitive to cluster initialisation and in order to obtain reproducible results, both techniques were initialised with the cluster assignments obtained by hierarchical clustering. For the ART algorithm, the parameters were set in order to obtain six clusters in order to match the number of clusters previously obtained by HCA. The best validity index obtained for repeated runs of the algorithm with 20 random initialisations was used to select the final clustering assignment.

3.1.2. Fuzzy c-means

The fuzzy c-means algorithm did not perform as hoped. When the number of clusters was set as two and three, it appeared that reasonable results were obtained. However, from examination of the membership function of each point assigned to these clusters, it could be seen that it was very close to either $\frac{1}{2}$ or $\frac{1}{3}$, respectively. In other words, every data point was assigned to all the clusters with the same membership. Moreover, when the number of clusters was above three, non-zero memberships were evident for only three clusters and these memberships were similar to the three cluster solution—i.e. for $n > 3$, the $n = 3$ cluster solution was obtained, but with $n-3$ empty clusters. These results indicated that the fuzzy c-means was not able to obtain clear cluster partitions.

The fuzziness index m was altered in an attempt to improve the results obtained, but it was found that little difference in the results was observed until m was close to one. Given that when $m = 1$ fuzzy c-means is equivalent to K-means, this result was not useful. As there are many applications for which the fuzzy c-means technique has been successful (see, for example [45]), these results are not easy to explain, but they may have been caused by the fact that our data contain a lot of values close to the extremes of each variable. Although the fuzzy c-means algorithm is widely used in literature, we decided to drop it from further analysis due to its poor performance on our data.

3.2. Cluster validity

The values of the decision rule obtained for various values of the validity indices for both K-means and PAM, for 2–20 clusters are shown in Fig. 1; (a) shows the validity decision rule values obtained for K-means and (b) shows those obtained for PAM. The best number of clusters according to each validity index, for each clustering algorithm, corresponds to the either the maximum or minimum decision rule value (depending on the index), as indicated by the solid circle in Fig. 1.

It can be seen that, while there was not absolute agreement among the indices as to which was the best number of clusters for the K-means method, there is good agreement that the best number of clusters for the PAM method is four. Although the best number of clusters varies according to validity index for K-means, on further inspection, it can be seen from Fig. 1 that there is more agreement than might be immediately apparent. For example, the Scott and Symons index (which indicated that the best number of clusters was three) indicated that the second best number of clusters was six. Consequently, the indices were used to rank order the number of clusters and the minimum sum of ranks was examined. It was found that the minimum sum of ranks (a form of consensus among the indices) indicated that the overall best number of clusters was six for K-means and four for PAM. Furthermore, careful examination of Fig. 1(b) confirms that the six cluster solution for PAM is of relatively poor quality.

3.3. Derivation of classes

The correspondence of patients assigned in the six cluster solution for each of the methods was then examined. Cohen's kappa and weighted-kappa indices were computed to measure the degree of agreement among algorithms. For the weighted-kappa index, weights were set in decreasing order from one (perfect agreement) to zero (complete disagreement) with a 0.2 step between levels. Results are reported in Table 4. From this table, a better agreement between K-means and hierarchical algorithms is evident compared to that between ART and hierarchical. It is also evident that the PAM six cluster solution has lower concordance with the original HCA results than either K-means or ART, and that the concordance of PAM with K-means and ART is also correspondingly lower.

The cluster numbers were aligned with those obtained previously by Abd El-Rehim et al. in [12] in order to minimise differences and to aid visualisation. Biplots of the aligned clusters are shown in Fig. 2 for the six cluster solution from each algorithm. From these plots, it can be seen that the most similar results were obtained from the Hierarchical, K-means and ART. In fact, all these three methods obtain two clusters (1 and 2) split over the left-hand side of the biplots. A third cluster (cluster 6) is evident towards the bottom of the biplot. Then various splits of remaining data into three clusters (3, 4 and 5) can be seen. The PAM algorithm, instead, obtains three clusters (1, 2 and 4) split over the left-hand side, one group is visible towards the bottom (cluster 6) and one is spread in the centre of the biplot (cluster 3). PAM places all patients on the right-hand side into a single cluster (cluster 5).

The biplots further confirm that the six cluster solution obtained from the PAM algorithm was the most dissimilar among

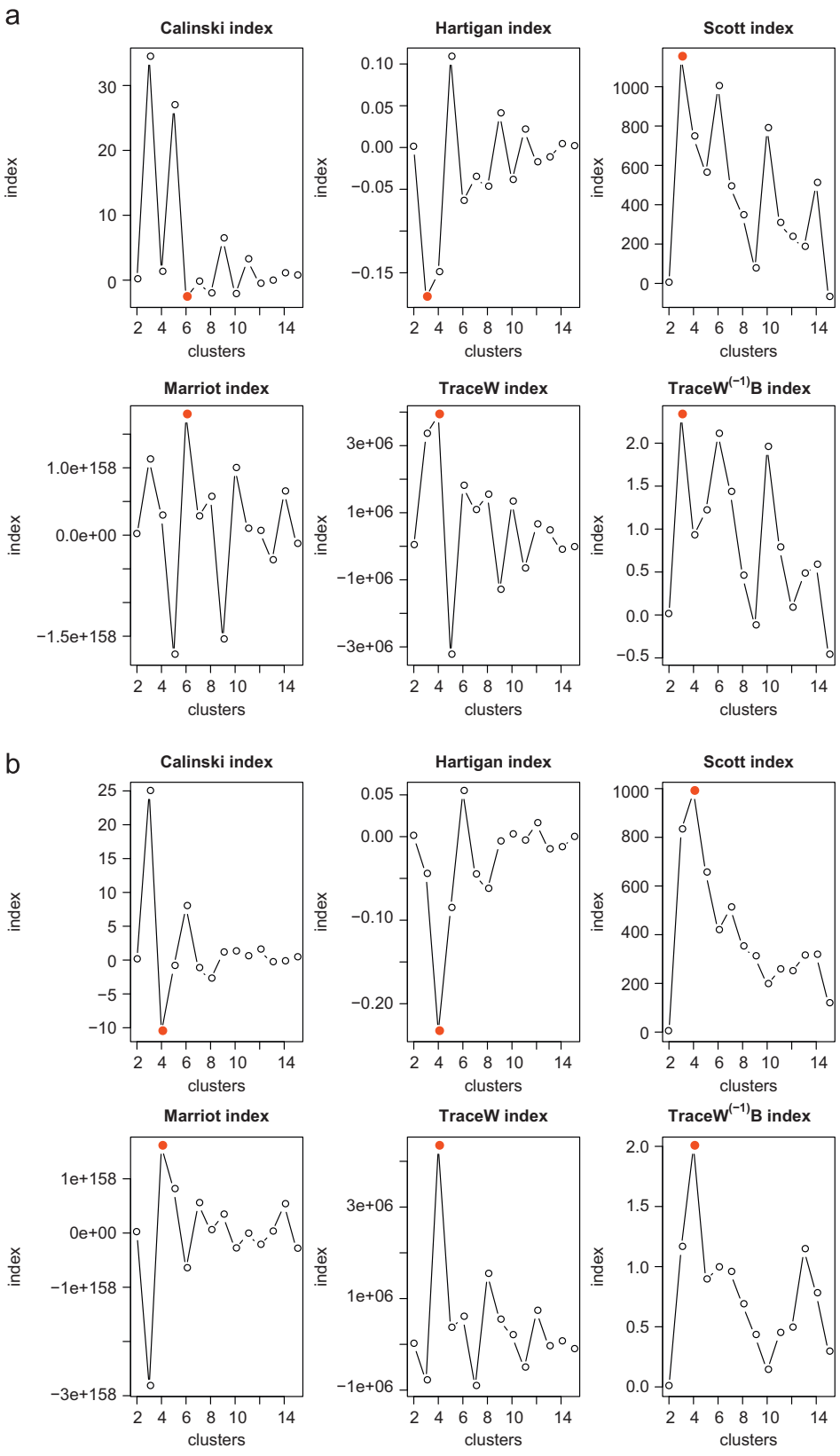


Fig. 1. Cluster validity indices obtained for K-means and PAM clustering, for varying cluster numbers from 2 to 20: (a) K-means indices behaviorsand and (b) PAM indices behaviors.

the considered techniques. Taking into account the results of validity indices analysis, the concordance analysis and the visual analysis, we decided to remove the six clusters determined by PAM from further analysis.

The cluster distributions (number of patients in each cluster) obtained for the original hierarchical clustering and those obtained for the K-means and ART methods are shown in Table 5.

Table 4
Kappa and *weighted kappa* index among different classifications.

	K-means	ART	PAM
HCA	0.497	0.296	0.325
	0.548	*0.401*	*0.332*
K-means	–	0.494	0.420
	–	*0.599*	*0.525*
ART	–	–	0.224
	–	–	*0.376*

Table 5
Number of cases in each cluster.

Cluster	HCA	K-means	ART
1	336	301	238
2	180	282	408
3	139	138	111
4	4	97	96
5	183	124	35
6	234	134	188

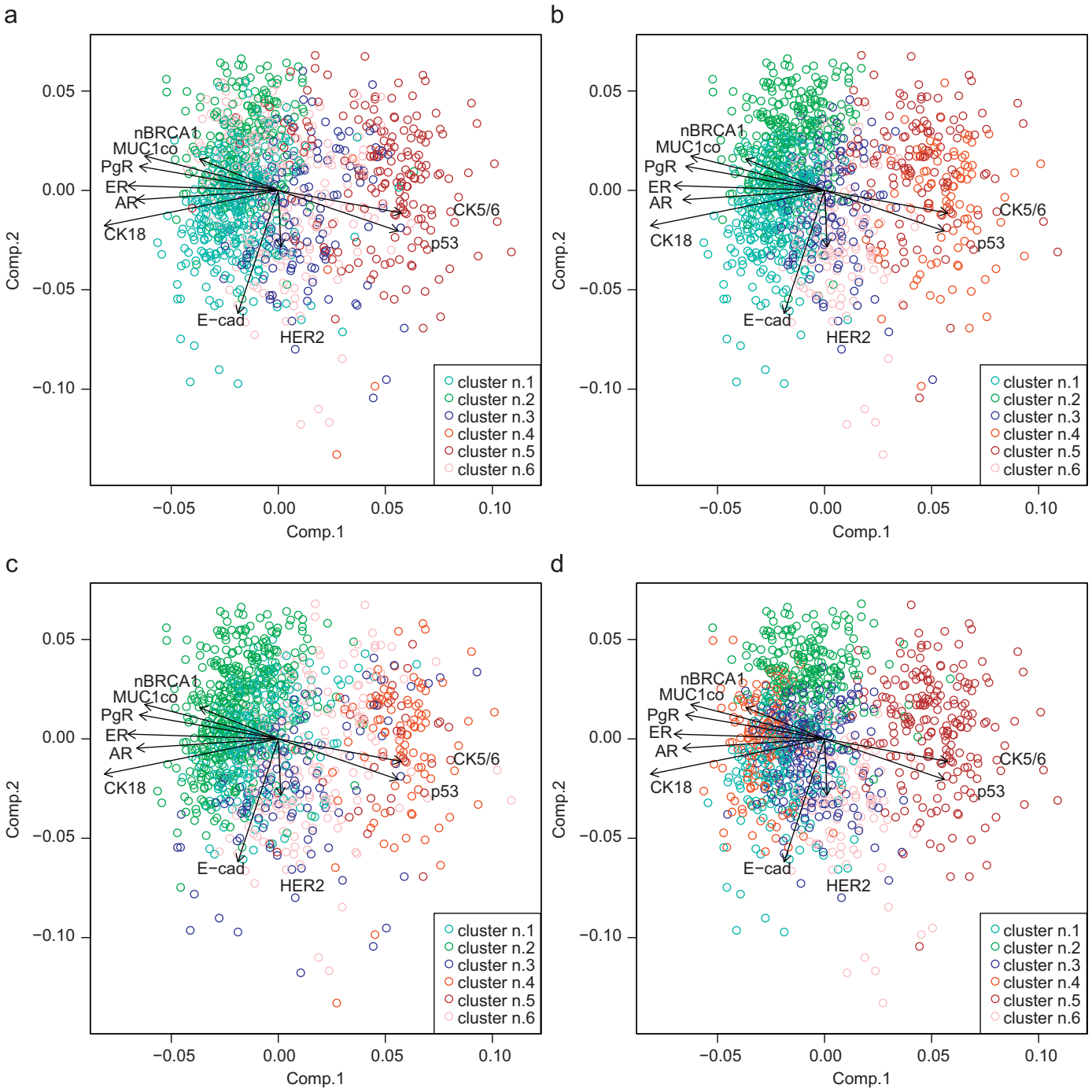


Fig. 2. Biplots of clusters projected on the first and second principal component axes: (a) Hierarchical clustering; (b) K-means clustering; (c) ART clustering and (d) PAM clustering.

Focusing on these cluster correspondences, we wanted to define core classes containing the biggest possible number of patients. In a first attempt, considering agreement among the three clustering techniques (HCA, KM and ART) and looking at those patients assigned to the same group by different methods, a total of 382 patients were classified if hierarchical group 4 was considered and 463 if not. After that, for each labelled group, concordances between all pairs of methods were analysed. It was found that the sum of the number of patients assigned to the same group ranged between 459 (pairing HCA and ART) and 645 (pairing KM and ART). These results are again reflected in Table 4. Two principles were used to guide the definition of consensus classes: (i) to consider all the clustering techniques analysed and (ii) to get the highest number of patients assigned to any class. These principles conflict, in that strict application of the first principle leads to a decrease in the number of patients assigned to classes. Hence, a heuristic trade-off between the two was employed. As a result, hierarchical group 4 was omitted (being replaced by group 5), and the ART assignments were not considered in a strictly conjunctive manner. Consequently, a set of six core breast tumour classes was derived following the specific rules reported in Table 6, in which the resultant number of patients in each class is shown.

It was found that almost the 62% of data was assigned to these core classes; the remaining patients were placed into a 'not classified' (NC) group. It must be stressed that the derivation of class assignments was made on the basis of the clustering results alone (which are, obviously, based on the 25 markers only)—class assignments, although somewhat subjective, were made *blind* to all clinical and outcome data. It should be also noted that around a third (actually 38%) of all patients were not assigned to any of the core classes.

Table 6
Rules for determining consensus classes.

If cluster ...	Class	No. of cases
HCA1 & KM1 & (ART1 | ART2)	1	202
HCA2 & KM2 & (ART1 | ART2)	2	153
HCA3 & KM3	3	80
HCA5 & KM4 & ART4	4	82
HCA5 & KM5	5	69
HCA6 & KM6 & ART6	6	77
Total number of cases assigned to classes 1–6		663
Total number of cases not classified		413

3.4. Characterisation of classes

Biplots of the six consensus classes were produced and are shown in Fig. 3, in order to provide a visualisation of the separation of the classes.

Fig. 3(a) shows the biplot obtained for all patients, in which the cases not assigned to any class (NC) have been coloured grey. It can be seen that these fall mainly into the centre region of the biplot. Fig. 3(b) shows the biplot obtained for only patients assigned to classes 1–6. It can be seen that the classes appear more spread out. The first axis was mainly determined, on the left, by luminal markers including luminal cytokeratins (CK18, CK7/8, CK19), hormone receptors (ER, AR, PgR), and MUC1 over-expression and, on the right, by basal cytokeratins (CK14 and CK5/6) and partly by p53 over-expression. The second axis is determined, on the top, partly by nuclear BRCA1 (nBRCA1) over-expression and, on the bottom, by HER2 and E-cad over-expression (also HER3 and HER4, although these are not shown as they overlap HER2).

Fig. 4 shows boxplots of all 25 markers, (a) for all cases, (b) for those cases assigned to classes 1–6, and (c–h) for each class separately. By inspection of both the biplots and the boxplots, we derived a description of each class. For example, classes 1 and 2 are characterised by strong expression of the luminal CK markers, as well as moderate to strong MUC1 expression (as per the population). However, there is a distinct difference regarding HER3 and HER4 expression. It can also be seen that classes 4 and 5 both exhibit higher expressions of the basal CKs (CK5/6 and CK14). Triple negative patients with high p53 levels are grouped in class 4, whereas class 5 consist of triple negative patients with low p53 levels. A summary of the class characteristics obtained by visual inspection of the boxplots is given in Table 7.

The results obtained from the automated characterisation methods (MLP-ANN and OSRE) are reported in Table 8.

A proposed summary of the essential characterisations of the classes obtained is given in Fig. 5, according to the available bio-pathological knowledge. It is worth noting that class 2, labelled as Luminal-N, and the split of the basal group into two different subgroups depending on p53 levels, appear to be novel findings not previously emphasised in literature.

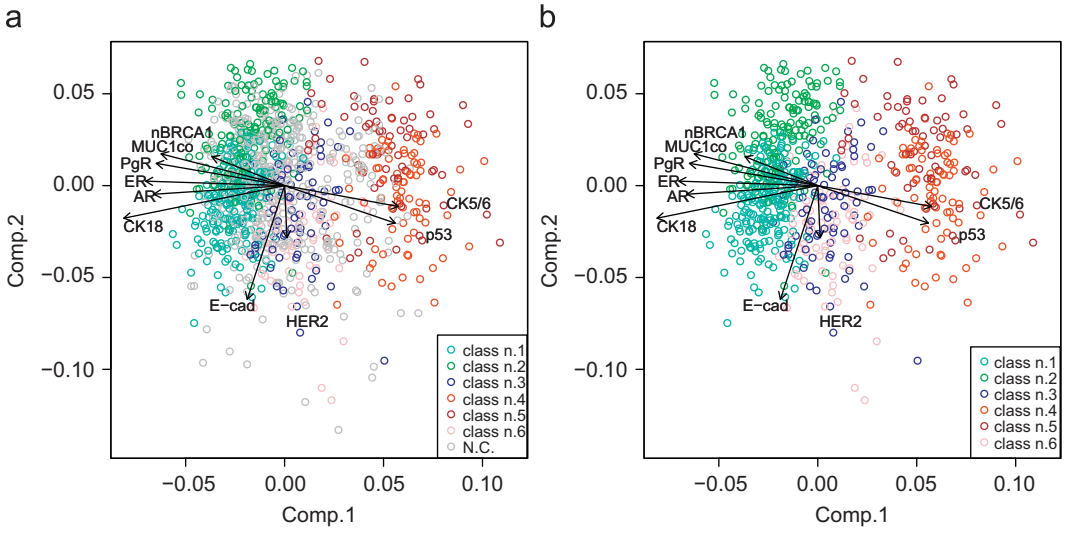


Fig. 3. Biplots of classes projected on the first and second principal component axes: (a) for all patientsand (b) for only patients in classes 1–6.

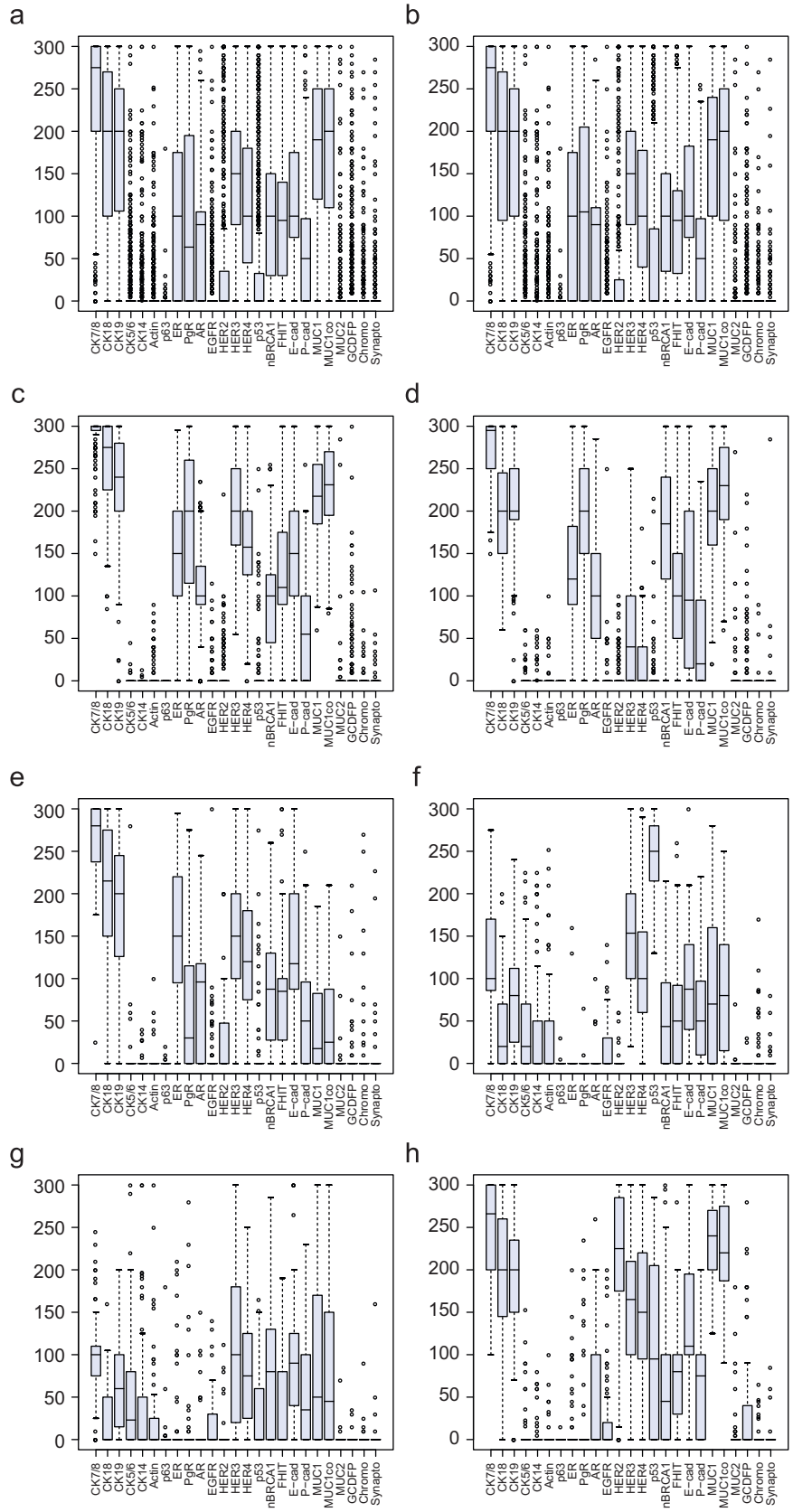


Fig. 4. Boxplot for all markers, whole data and grouped by class: (a) For all patients; (b) for patients in classes 1–6; (c) class 1; (d) class 2; (e) class 3; (f) class 4; (g) class 5 and (h) class 6.

Table 7
Description of classes as determined by statistical characterisation.

Class	Over-expressed	Under-expressed	Other
1	ER, AR, PgR, HER3, HER4		
2	ER, AR, PgR, nBRCA1	HER3, HER4	
3	ER, AR	MUC1, MUC1co	PgR normal
4	p53	ER, PgR, HER2, MUC1, MUC1co, CK18, CK7/8, CK19	
5		ER, PgR, HER2, MUC1, MUC1co, CK18, CK7/8, CK19	p53 absent
6	HER2, HER3, HER4		ER, AR, PgR absent; p53 widely spread

Table 8
A summary of rules obtained from the automated methods for defining class memberships.

Class	Over-expressed	Under-expressed
1 (ANN)	PgR, HER3, HER4, MUC1co	
1 (OSRE)	PgR, HER3, HER4, CK18, CK19, MUC1co	HER2
2 (ANN)	PgR, nBRCA1	HER3, HER4
2 (OSRE)	PgR, nBRCA1, MUC1co	HER3, HER4
3 (ANN)	ER	MUC1
3 (OSRE)	CK7/8, CK18	
4 (ANN)	p53	
4 (OSRE)	HER3, p53	ER, HER2
5 (ANN)	CK5/6	CK7/8
5 (OSRE)		p53; CK7/8, CK19 or HER2, HER4
6 (ANN)	HER2	
6 (OSRE)	HER2, p53, MUC1co	ER

4. Clinical evaluation

4.1. Patient clinical outcome

Patient age ranged from 18 to 72 years (median 54 years). Of the available cases, 708 (66%) cases were aged 50 years or more. At the time of diagnosis, 160 (14.9%) tumours were histological grade 1, 343 (31.9%) grade 2 and 572 (53.2%) grade 3. A total of 654 (60.8%) patients had lymph node-negative disease and 419 (38.9%) had positive lymph nodes (332 cases with between one and three positive nodes, 87 cases with four or more positive). Frequencies for histological tumour types were: 649 invasive ductal carcinomas of no special type (NST), 171 tubular mixed carcinomas, 30 medullary carcinomas, 112 lobular carcinomas, 27 tubular carcinomas, 11 mucinous carcinomas, five cribriform carcinomas, three papillary carcinomas, 37 mixed NST and lobular carcinomas, 24 mixed NST and special type carcinomas and four miscellaneous tumours. A total of 736 (68.4%) had tumour size more than 1.5 cm and distant metastases was observed in 111 cases.

4.2. Clinical characterisation of patients by class

Significant associations, as expected, were found between the classes with respect to patient age, tumour grade, size, lymph node stage and histological tumour type (see Table 9).

A boxplot of the Nottingham Prognostic Index (NPI) split by class is shown in Fig. 6. It can be seen that the NPI for classes 1 and 2 is lower than that of classes 3–6 (overall Kruskal-Wallis $p \ll 0.001$). It can also be seen that classes 1 and 2 have similar NPI, and classes 3–6 have similar NPI (to each other). This is an interesting observation for two reasons. Firstly, it confirms that the NPI is providing discriminant information between classes 1 and 2, and classes 3–6. Secondly, it suggests that the class divisions are providing *additional* information to the NPI.

5. Discussion

This study has extended our previous work [12], with the application of different clustering techniques to address the issue of the non-existence of the 'perfect' clustering algorithm. In particular, in this work four different clustering methods (in addition to the hierarchical method used in [12]) were applied to a multidimensional dataset of protein biomarker data, in order to evaluate the stability of results coming from different techniques. Different clustering algorithms result in different clusters, particularly when large multidimensional data sets are considered.

To explore the extent of the differences among different algorithms, an informal consensus clustering was used, grouping together patients that were assigned to 'similar' clusters by different clustering algorithms. The consensus approach was similar to the one used by Kellam et al. [23], but instead of building an agreement matrix, the previously published hierarchical clustering solution (and associated labelling) was used as a fixed reference. In this way, a set of six core classes of breast cancer was elucidated. This consensus methodology was used to combine results obtained by different clustering algorithms rather than as a comparison with previously published approaches. Another important issue that emerges when cluster analysis is performed, is the best number of clusters to consider. Several validity indices have been proposed in recent years (see, for example [36]) to evaluate the compactness of clusters and the separation among them. For the algorithms which take an explicit number of clusters as an input parameter (i.e. K-means and PAM), cluster validity indices were used to guide the choice of the 'best' number of clusters. Note that cluster validity indices would have been applied to the fuzzy c-means algorithm had it not been dropped from analysis for the reasons outlined in Section 3.1.2.

Furthermore, this study confirmed, as already highlighted in [13], that cluster analysis should be treated with caution, as different clustering algorithms will lead to different groupings of tumours. In particular, in our case, the PAM algorithm, when run with six clusters as an input, provided groups that were different from those obtained using the other techniques. In addition, the hierarchical algorithm, commonly used in standard bioinformatics applications of cluster analysis, such as [4,9], seems to provide a dissimilar and skewed

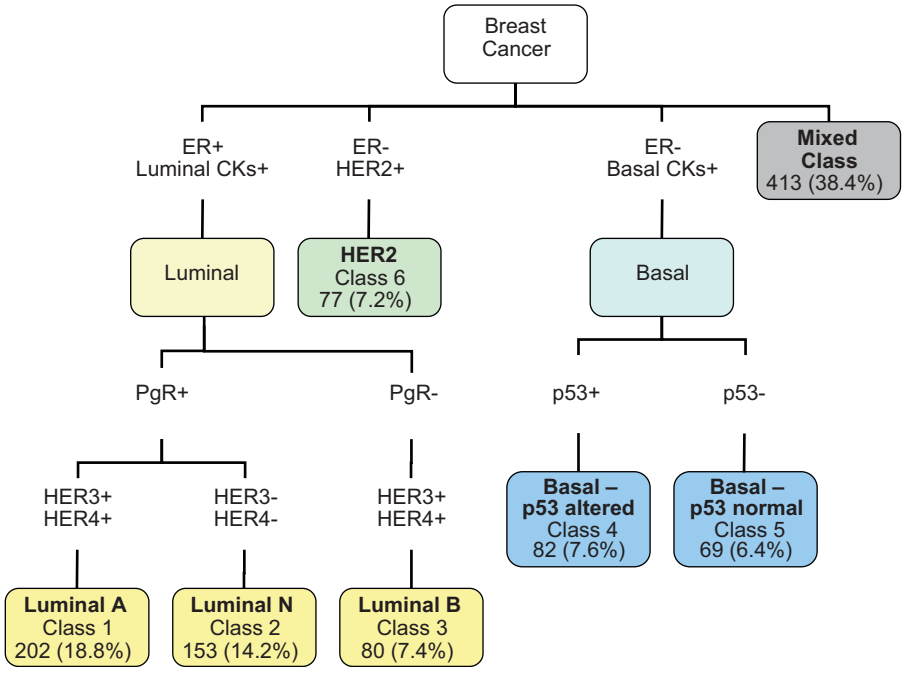


Fig. 5. A summary of the classes of breast cancer obtained, with indicative class interpretations.

Table 9
Breast Cancer Class distribution in relation to clinicopathological parameters (NST: No Special Type).

	Breast cancer class						
	1	2	3	4	5	6	ϕ
Age							
$\leq$ 50	76 (37.6)	63 (41.2)	24 (30.0)	55 (67.1)	33 (47.8)	37 (48.1)	0.209
$>$ 50	126 (62.4)	90 (58.8)	56 (70.0)	27 (32.9)	36 (52.2)	40 (51.9)	
Total	202	153	80	82	69	77	
Grade							
1	58 (28.9)	43 (28.1)	2 (2.5)	0 (0)	2 (2.9)	1 (1.3)	
2	81 (40.2)	89 (58.2)	18 (22.5)	1 (1.2)	7 (10.1)	12 (15.6)	0.660
3	62 (30.8)	21 (13.7)	60 (75.0)	81 (98.8)	60 (87.0)	64 (83.1)	
Total	201	153	80	82	69	77	
Size							
$\leq$ 1.5 cm	79 (39.1)	65 (42.5)	20 (25.0)	12 (14.6)	15 (21.7)	16 (20.8)	0.225
$>$ 1.5 cm	123 (60.9)	88 (57.5)	60 (75.0)	70 (85.4)	54 (78.3)	61 (79.2)	
Total	202	153	80	82	69	77	
Lymph node stage							
1	132 (65.3)	108 (70.6)	39 (48.7)	50 (61.0)	52 (75.4)	36 (46.8)	
2	58 (28.7)	37 (24.2)	35 (43.8)	23 (28.0)	10 (14.5)	30 (39.0)	0.217
3	12 (5.9)	7 (4.6)	6 (7.5)	9 (11.0)	7 (10.1)	10 (13.0)	
Total	202	152	80	82	69	76	
Tumour type							
Invasive ductal/NST	97 (48.0)	45 (29.4)	64 (80.0)	70 (85.4)	53 (76.8)	68 (88.3)	
Tubular mixed	52 (25.7)	50 (32.6)	8 (10.0)	0 (0)	1 (1.5)	5 (6.5)	
Medullary	0 (0)	0 (0)	0 (0)	10 (12.2)	5 (7.2)	2 (2.6)	
Lobular	18 (8.9)	34 (22.2)	6 (7.5)	0 (0)	4 (5.8)	1 (1.3)	0.622
Special types	19 (9.4)	11 (7.2)	0 (0)	1 (1.2)	0 (0)	0 (0)	
Mixed NST and lobular	6 (3.0)	7 (4.6)	2 (2.5)	1 (1.2)	3 (4.3)	0 (0)	
Mixed NST and special type	9 (4.5)	5 (3.3)	0 (0)	0 (0)	1 (1.5)	0 (0)	
Miscellaneous	0 (0)	1 (0.7)	0 (0)	0 (0)	2 (2.9)	0 (0)	
Total	201	153	80	82	69	76	

classification with respect to the others, thus reducing the degree of overall concordance and the number of subjects assigned to the core classes.

In conclusion, we have clearly demonstrated that different clustering algorithms can produce quite different solutions on such multi-dimensional data. It should be noted that no feature

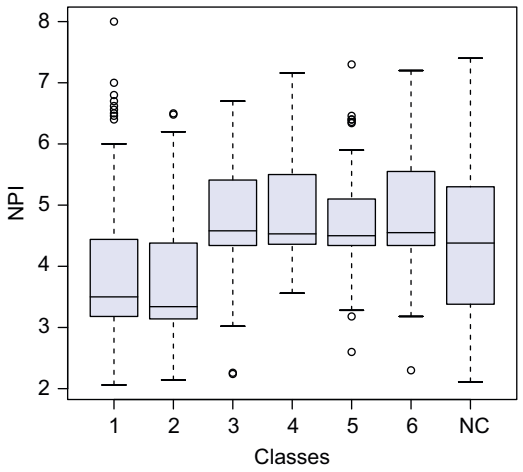


Fig. 6. Boxplots of Nottingham Prognostic Index (NPI) by class.

extraction was performed in this study, so avoiding a possible cause of diverse techniques not converging into similar results. We have proposed a methodology for reaching consensus from the various results that may be obtained from clustering algorithms, and have illustrated this consensus methodology on a well-known set of breast cancer data. In doing so, we have identified possible new sub-classes of breast cancer which warrant further investigation. We emphasise that this consensus methodology, by its heuristic nature, should be considered as an exploratory technique, and must not be considered as providing any form of definitive answer. Further work exploring, for example, the statistical properties of the considered algorithms may provide relevant information on the structure on this complex biological problem.

6. Summary

In a previous study [12], we have suggested that immunohistochemical analysis may be used to identify distinct biological classes of breast cancer. The objectives of this work were to verify the stability of groups obtained from four different unsupervised clustering algorithms, applied to the same data, and to compare and combine the different solutions with the ones available from the previous study. The clustering techniques were used to divide our dataset into six groups, which were labelled according to our previous classification [12]. Moreover, where appropriate, validity indices were used to explore the best data subdivision and to validate the obtained classification. Despite the fact that fuzzy c-means is one of the most widely used clustering techniques, results obtained from the algorithm were quite poor and were dropped from further investigation. The PAM algorithm produce somewhat different results to the other techniques, so that correspondences between classifications were also difficult. Then, only considering hierarchical, K-means and ART methods, a set of six core classes was elucidated by a form of consensus clustering in which labels assigned to the groups were aligned to find correspondences among patients grouped in similar clusters by different techniques. It was found that around the 62% (663 patients) of the available data was assigned to classes, while the remaining 413 women (38%) presented mixed class characteristics. The use of different clustering methods has, once again, demonstrated that diverse algorithms will in general produce different clusters, particularly when large multi-dimensional data sets are considered.

Conflict of interest statement

None declared.

Acknowledgements

This study was, in part, funded by the Breast Cancer Campaign, and was supported by the BIOPATTERN FP6 Network of Excellence (FP6-IST-508803) and the BIOPTRAIN FP6 Marie-Curie EST Fellowship (FP6-007597).

References

[1] D. Parkin, F. Bray, J. Ferlay, P. Pisani, Estimating the world cancer burden: Globocan 2000, Int. J. Cancer 94 (2001) 153–156.
[2] F. Kamangar, G. Dores, W. Anderson, Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world, J. Clin. Oncol. 24 (2006) 2137–2150.
[3] M. Eisen, P. Spellman, P. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. USA 95 (1998) 14863–14868.
[4] C. Perou, T. Sorlie, M. Eisen, M. Van De Rijn, S. Jeffrey, C. Rees, J. Pollack, D. Ross, H. Johnsen, L. Akslen, Ø. Fluge, A. Pergamenschikov, C. Williams, S. Zhu, P. Lonning, A. Børresen-Dale, P. Brown, D. Botstein, Molecular portraits of human breast tumours, Nature 406 (2000) 747–752.
[5] T. Sorlie, C. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. Eisen, M. Van De Rijn, S. Jeffrey, T. Thorsen, H. Quist, J. Matese, P. Brown, D. Botstein, P. Eystein Lonning, A. Børresen-Dale, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, Proc. Natl. Acad. Sci. USA 98 (2001) 10869–10874.
[6] T. Sorlie, R. Tibshirani, J. Parker, T. Hastie, J. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. Perou, P. Lonning, P. Brown, A. Børresen-Dale, D. Botstein, Repeated observation of breast tumor subtypes in independent gene expression data sets, Proc. Natl. Acad. Sci. USA 100 (2003) 8418–8423.
[7] C. Sotiriou, S.-Y. Neo, L. McShane, E. Korn, P. Long, A. Jazaeri, P. Martiat, S. Fox, A. Harris, E. Liu, Breast cancer classification and prognosis based on gene expression profiles from a population-based study, Proc. Natl. Acad. Sci. USA 100 (2003) 10393–10398.
[8] S. Calza, P. Hall, G. Auer, J. Björhle, S. Klaar, U. Kronenwett, E. Liu, L. Miller, A. Ploner, J. Smeds, J. Bergh, Y. Pawitan, Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients, Breast Cancer Res. 8 R34, 2006.
[9] L. van't Veer, H. Dai, M. van de Vijver, Y. He, A. Hart, M. Mao, H. Peterse, K. van der Kooy, M. Marton, A. Witteveen, G. Schreiber, R. Kerkhoven, C. Roberts, P. Linsley, R. Bernards, S. Friend, Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (2002) 530–536.
[10] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. Olson Jr., J. Marks, J. Nevins, Predicting the clinical status of human breast cancer by using gene expression profiles, Proc. Natl. Acad. Sci. USA 98 (2001) 11462–11467.
[11] Z. Hu, C. Fan, D. Oh, J. Marron, X. He, B. Qaqish, C. Livasy, L. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M. Ewend, L. Sawyer, J. Wu, Y. Liu, R. Nanda, M. Tretiakova, A. Ruiz Orrico, D. Dreher, J. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J. Quackenbush, M. Ellis, O. Olopade, P. Bernard, C. Perou, The molecular portraits of breast tumors are conserved across microarray platforms, BMC Genomics 7 (2006) 96.
[12] D. Abd El-Rehim, G. Ball, S. Pinder, E. Rakha, C. Paish, J. Robertson, D. Macmillan, R. Blamey, I. Ellis, High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses, Int. J. Cancer 116 (2005) 340–350.
[13] F. Ambrogi, E. Biganzoli, P. Querzoli, S. Ferretti, P. Boracchi, S. Alberti, E. Marubini, I. Nenci, Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods, Clin. Cancer Res. 12 (3) (2006) 781–790.
[14] E. Korsching, J. Packeisen, K. Agelopoulos, M. Eisenacher, R. Voss, J. Isola, P. van Diest, B. Brandt, W. Boecker, H. Buerger, Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis, Lab. Invest. 82 (2002) 1525–1533.
[15] G. Callagy, E. Cattaneo, Y. Daigo, L. Happerfield, L. Bobrow, P. Pharoah, C. Caldas, Molecular classification of breast carcinomas using tissue microarrays, Diag. Mol. Pathol. 12 (2003) 27–34.
[16] N. Makretsov, D. Huntsman, T. Nielsen, E. Yorida, M. Peacock, M. Cheang, S. Dunn, M. Hayes, M. van de Rijn, C. Bajdik, C. Blake Gilks, Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma, Clin. Cancer Res. 10 (2004) 6143–6151.

[17] J. Jacquemier, C. Ginestier, J. Rougemont, V.-J. Bardou, E. Charafe-Jauffret, J. Geneix, J. Adélaïde, A. Koki, G. Houvenaeghel, J. Hassoun, D. Maraninchi, P. Viens, D. Birnbaum, F. Bertucci, Protein expression profiling identifies subclasses of breast cancer and predicts prognosis, Cancer Res. 65 (2005) 767–779.
[18] R. Diallo-Danebrock, E. Ting, O. Gluz, A. Herr, S. Mohrmann, H. Geddert, A. Rody, K. Schaefer, S. Baldus, A. Hartmann, P. Wild, M. Burson, H. Gabbert, U. Nitz, C. Poremba, Protein expression profiling in high-risk breast cancer patients treated with high-dose or conventional dose-dense chemotherapy, Clin. Cancer Res. 13 (2007) 488–497.
[19] M. Dolled-Filhart, L. Rydén, M. Cregger, K. Jirström, M. Harigopal, R. Camp, D. Rimm, Classification of breast cancer using genetic algorithms and tissue microarrays, Clin. Cancer Res. 12 (2006) 6459–6468.
[20] S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data, Mach. Learn. 52 (2003) 91–118.
[21] S. Swift, A. Tucker, V. Vinciotti, N. Martin, C. Orengo, X. Liu, P. Kellam, Consensus clustering and functional interpretation of gene-expression data, Genome Biol. 5 R94, 2004.
[22] V. Filkov, S. Skiena, Integrating microarray data by consensus clustering, in: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, 2003, pp. 418–426.
[23] P. Kellam, X. Liu, N. Martin, C. Orengo, S. Swift, A. Tucker, Comparing, contrasting and combining clusters in viral gene expression data, in: Proceedings of the 6th Workshop on Intelligent Data Analysis in Medicine, 2001.
[24] J. Maindonald, W. Braun, Data Analysis and Graphics Using R—An Example-Based Approach, Cambridge University Press, Cambridge, 2003.
[25] M. Al-Daoud, S. Roberts, New methods for the initialisation of clusters, Pattern Recognition Letters 17 (5) (1996) 451–455.
[26] L. Kaufman, P. Rousseeuw, Finding Groups in Data: an Introduction to Cluster Analysis, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, New York, 1990.
[27] G. Carpenter, S. Grossberg, ART2: Stable self-organization of pattern recognition codes for analogue input patterns, Appl. Opt. 26 (1987) 4919–4930.
[28] L. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[29] J. Bezdek, Cluster validity with fuzzy sets, J. Cybern. 3 (3) (1974) 58–73.
[30] U. Maulik, S. Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1650–1654.
[31] J. Hartigan, Clustering Algorithms, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, New York, 1975.
[32] A. Scott, M. Symons, Clustering methods based on likelihood ratio criteria, Biometrics 27 (2) (1971) 387–397.
[33] F. Marriot, Practical problems in a method of cluster analysis, Biometrics 27 (3) (1971) 501–514.
[34] A. Edwards, L. Cavalli-Sforza, A method for cluster analysis, Biometrics 21 (2) (1965) 362–375.
[35] H. Friedman, J. Rubin, On some invariant criteria for grouping data, J. Am. Statist. Assoc. 62 (320) (1967) 1159–1178.
[36] A. Weingessel, E. Dimitriadou, S. Dolnicar, An examination of indexes for determining the number of clusters in binary data sets, Working Paper No. 29, 1999.
[37] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.
[38] J. Jackson, A User's Guide to Principal Components, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, New York, 1991.
[39] P. Velleman, D. Hoaglin, Applications, Basics and Computing of Exploratory Data Analysis, Duxbury Press, Boston, MA, 1981.
[40] T. Etchells, P. Lisboa, Rule extraction from neural networks: a practical and efficient approach, IEEE Trans. Neural Networks 17 (2) (2006) 374–384.
[41] S. Michiels, S. Koscielny, C. Hill, Prediction of cancer outcome with microarrays: a multiple random validation strategy, The Lancet 365 (9458) (2005) 488–492.
[42] B. Matharoo-Ball, L. Ratcliffe, L. Lancashire, S. Ugurel, A. Miles, D. Weston, R. Rees, D. Schadendorf, G. Ball, C. Creaser, Diagnostic biomarkers differentiating metastatic melanoma patients from healthy controls identified by an integrated MALDI-TOF mass spectrometry/bioinformatic approach, Proteomics Clin. Appl. 1 (6) (2007) 605–620.
[43] R. McClelland, P. Finlay, K. Walker, D. Nicholson, J. Robertson, R. Blamey, R. Nicholson, Automated quantitation of immunocytochemically localized estrogen receptors in human breast cancer, Cancer Res. 50 (1990) 3545–3550.
[44] S. Detre, G. Saclani Jotti, M. Dowsett, A 'quickscore' method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas, J. Clin. Pathol. 48 (1995) 876–878.
[45] X. Wang, J. Garibaldi, A comparison of fuzzy and non-fuzzy clustering techniques in cancer diagnosis, in: Proceedings of Second International Conference in Computational Intelligence in Medicine and Healthcare, 2005, pp. 250–256.