



HAL
open science

Unsupervised Clustering of Immunohistochemical Markers to Define High-Risk Endometrial Cancer

Enora Laas, Marcos Ballester, Annie Cortez, Olivier Graesslin, Emile Darai

► **To cite this version:**

Enora Laas, Marcos Ballester, Annie Cortez, Olivier Graesslin, Emile Darai. Unsupervised Clustering of Immunohistochemical Markers to Define High-Risk Endometrial Cancer. *Pathology and Oncology Research*, 2019, 25 (2), pp.461-469. 10.1007/s12253-017-0335-y . hal-02336287

HAL Id: hal-02336287

<https://hal.sorbonne-universite.fr/hal-02336287>

Submitted on 28 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised clustering of immunohistochemical markers to define high-risk endometrial cancer.

Enora Laas, MD ^{1,2}, Marcos Ballester, MD PhD ^{1,2,3}, Annie Cortez, MD ^{2,4},

Olivier Graesslin, MD PhD⁶, Emile Daraï, MDPHD^{1,2,5}.

¹ Service de Gynécologie-Obstétrique, Hôpital Tenon, AP-HP, 4 rue de la Chine 75020 Paris, France

² Institut Universitaire de Cancérologie, Université Pierre et Marie Curie, Paris 6, France.

³ UMRS-938, Université Pierre et Marie Curie, Paris 6, France.

⁴ Service d'Anatomie Pathologie, Hôpital Tenon, AP-HP, 4 rue de la Chine 75020, Paris, France.

⁵ UMRS 938, Université Pierre et Marie Curie, Paris 6, France.

⁶ Service de gynécologie-obstétrique, Hôpital Alix de Champagne, CHU de Reims, 45 rue Cognacq-Jay 51100 Reims, France.

Correspondance:

Enora Laas, MD

Service de Gynécologie-Obstétrique, Hôpital Tenon, AP-HP, Paris, France.

Tel: +33156017318

Email: enora.laas@tnn.aphp.fr

Abstract

Objective: Considerable heterogeneity exists in outcomes of early endometrial cancer (EC) according to the type but also the histological grading.

Our goal was to describe the immunohistochemical profiles of type I EC according to grades and type II EC, to identify groups of interacting proteins using principal component analysis (PCA) and unsupervised clustering.

Methods: We studied 13 immunohistochemical markers (steroid receptors, pro/anti-apoptotic proteins, metalloproteinases (MMP) and tissue inhibitor of metalloproteinase (TIMP), and CD44 isoforms known for their role in endometrial pathology. Co-expressed proteins associated with the type, grade and outcome of EC were determined by PCA and unsupervised clustering.

Results: PCA identified three functional groups of proteins from 43 tissue samples (38 type I and 5 type II EC): the first was characterized by p53 expression; the second by MMPs, bcl-2, PR B and CD44v6; and the third by ER alpha, PR A, TIMP-2 and CD44v3.

Unsupervised clustering found two main clusters of proteins, with both type I grade 3 and type II EC exhibiting the same cluster profile.

Conclusion: PCA and unsupervised clustering of immunohistochemical markers in EC contribute to a better comprehension and classification of the disease.

Keywords: endometrial cancer, high-risk endometrial cancer, immunohistochemistry, unsupervised clustering, principal component analysis

Background

Endometrial cancer (EC) is the most common malignant tumour of the female genital tract and mainly affects postmenopausal women. Based on clinico-pathological, immunohistochemical, and molecular genetic properties, EC tumours have been classified into two types. Type I, corresponding to endometrioid carcinoma, accounts for 85% of EC and is mainly of low or intermediate grade, estrogen-related and is preceded or coexists with atypical and complex endometrial hyperplasia. Type II, corresponding to serous, clear cell or carcinosarcoma, accounts for approximately 15% of EC, and exhibits an aggressive behaviour which is unrelated to oestrogen stimulation, associated with P53 mutations and occasionally arises from endometrial polyps or precancerous lesions in atrophic endometria [1, 2]. Type I EC has a good prognosis with an overall 5-year survival over 85% while type II EC has a poor prognosis with an overall 5-year survival of 35% as well as being responsible for about 50% of EC recurrences [3, 4].

In addition to EC type, histological grading is a major prognostic factor and a determinant for adjuvant therapy. For type I EC, grade 1 (G1) and 2 (G2) are associated with good prognosis when restricted to the uterus. Type I Grade 3 (G3) EC is an aggressive tumour conveying a higher risk of lymph node metastasis. Type II ECs are high-grade tumours by definition and carry a higher mortality than type I [3, 4].

In addition to the International Federation of Gynecology and Obstetrics (FIGO) classification, the European Society of Medical Oncology (ESMO) recurrence risk classification defines patients as being at low (type I, G1 or G2 EC with less than 50% of myometrial invasion), intermediate (type I G3 EC with less than 50% of myometrial invasion and type I G1 or 2 EC with more than 50% of myometrial invasion), and high-risk (type I G3 EC with more than 50% of myometrial invasion and type II EC). To improve patient risk stratification, the presence of lymphovascular space involvement (LVSI) [5] has been incorporated into international classifications such as in France (INCa recommendations). Despite these modifications, a considerable heterogeneity in outcomes of early EC, especially for the intermediate risk group, has been observed [6, 7]. Moreover, controversies exist on whether type I G3 EC should be assimilated to type II EC [8, 9]. Voss *et al.* demonstrated that type I G3 EC and type II EC had similar clinical, immunohistochemical profiles and survival outcomes [8] while Zannoni *et al.* reported different immunohistological profiles for p53, oestrogen and progesterone receptors between these two entities [10].

To better define risk groups of recurrence with a view to adapting the therapeutic strategy, there is a need for new tools. Principal component analysis (PCA) and unsupervised clustering are machine learning methods

which analyse explanatory data in an attempt to identify hidden structures in unlabeled data. These methods are widely used in genomics to reduce the dimensionality of high dimensional genomic data.

We decided to study a panel of previously assessed EC immunohistochemical markers by PCA and unsupervised clustering to describe the immunohistochemical profiles of type I EC according to grade and type II EC and thereby to identify groups of interacting proteins involved in the physiopathology of the disease.

Material and Methods

Patients

We obtained tissue samples from 43 women who underwent total hysterectomy from 1999 to 2004 in the Gynecology Departments of Tenon Hospital, Paris, France and at the Institut Mère-Enfant Alix de Champagne, Reims, France. These tissue samples included 43 EC: 38 type I EC and 5 type II EC. Two pathologists reviewed all the tissue samples to confirm the diagnosis. None of the women with EC had received chemotherapy or radiotherapy before surgery. Patients previously treated with tamoxifen were excluded from the study. The histological grade of EC was assessed according to the WHO classification. All the samples were obtained with full and informed patient consent. Only patients with strict follow-up in the two institutions were included. The patients gave their consent and the research protocol was approved by the Consultative Committee for Protection of Persons in Biomedical Research of Paris 6 (France).

Immunohistochemistry

We used a selection of the 13 most studied biological markers which are known to play a role in EC. Tissues were immediately fixed in formalin (10%) and then processed as paraffin blocks. Four-micron-thick sections of formalin-fixed tissues were deparaffinised in a xylene substitute (EZ Prep w 1x ref 950102) and rehydrated through a graded series of ethanol solutions. The sections were immunostained using the Ventana Benchmark XTw automated immunohistochemistry system (Ultra-View™, Universal DAB-Ventana™). Immunohistochemical analysis was performed on hot spots on ten fields at x40 magnification.

Oestrogen receptor (ER), Progesteron receptor (PR), Ki67, bcl-2 and p53 immunostaining

Rabbit monoclonal antibodies directed against ER (prediluted, SP1, Ventana Medical Systems, USA), PR (prediluted, 1E2, Ventana Medical Systems, USA), mouse monoclonal antibody against ki67 (1/150, Mib 1, DAKO) and p53 (1/75, DO7, DAKO), and an anti-human-bcl-2 rabbit polyclonal antibody (1/50, Santa-Cruz Biotechnology, Santa-Cruz, CA, USA) were used. An antigen retrieval procedure was performed (Dako Target Retrieval Solution; 98°C, 20 min) prior to ER, PR, ki67, p53 and bcl-2 staining. This automated procedure is based on an indirect biotin-avidin system. A universal biotinylated immunoglobulin was used as a secondary antibody, diaminobenzidine as the substrate and hematoxylin as the counterstain. Positive controls were sections

of human breast tissue for ER and PR, colon adenocarcinoma for ki67, serous ovarian carcinoma for p53 and lymph node for bcl-2 according to the manufacturer's instructions. For negative controls, the primary antibody was replaced by an irrelevant antibody of the same IgG subtype.

Metalloproteinases MMP-2, -7 and -9 and tissue inhibitor of metalloproteinase TIMP-1 and -2 immunostaining

Purified mouse monoclonal antibodies against human MMP-2 (1/10, gelatinase-A, Ab-2, clone 17B11, Novocastra Laboratories Ltd, Newcastle, United Kingdom), human MMP-7 (1/50, matrilysin, Ab-3, clone ID2, Calbiochem, Oncogene Research Products, Cambridge, United Kingdom), human MMP-9 (1/10, gelatinase-B, Ab-3, clone 56 2A4, Calbiochem, Oncogene Research Products), human TIMP-1 (1/50 Ab-4, clone 3A4, Novocastra Laboratories Ltd) and human TIMP-2 (Ab-2, clone 6F6a, Novocastra Laboratories Ltd) were used as primary antibodies.

Prior to MMP-2, MMP-9, TIMP-1 and TIMP-2 staining, an antigen retrieval step was performed combined with a high-temperature antigen-unmasking technique (DAKO® Target Retrieval Solution; 100°C, 30 minutes). For MMP-7, antigen unmasking was achieved with proteinase K. The automated procedure is based on an indirect biotin-avidin system with a universal biotinylated immunoglobulin as secondary antibody, diaminobenzidine as substrate, and hematoxylin as counterstain. For MMP-7 and MMP-9, a Ventana amplification kit was used (Ventana Medical Systems®) in addition to the automated procedure. Positive controls for MMP-2, -7, -9, TIMP-1, -2 were sections of bronchial tissue, ovarian cancer, breast cancer and placenta, respectively. For negative controls, the primary antibody was replaced by an irrelevant non-immune mouse antibody of the same IgG subtype (Dako Laboratories, Glostrup, Denmark).

CD44 isoforms and variants v3 and v6 immunostaining

The following monoclonal antibodies (R&D Systems, Abingdon, UK) were used: 2C5, binding to CD44s and all the variants encoded by exons 3 to 10, 3G5, specific for CD44v3 and 2F10, specific for CD44v6.

For the antigen unmasking procedure, first the sections were incubated in citrate buffer 10-mM pH6 in a microwave oven for periods of 5 minutes each at 500 watts. After washing, sections were incubated for 1 hour at room temperature with the primary antibody in a 1/100 dilution. The revelation was performed by using the avidin-biotin technique (Vector Laboratories, Burlingame, UK). Peroxidase activity was detected according to the method of Graham and Karnovsky. Positive controls were uterine myomas as previously published by our

team [11] Negative controls were obtained by substituting the primary antibody by either PBS or isotopic immunoglobins.

Analysis of immunohistochemical results

For semi-quantitative analysis of ER, PR, Ki67, p53, bcl-2 and CD44 isoforms, the percentage of stained cells was calculated by two independent observers. For MMPs and TIMPs, the intensity of staining was analysed using HSCORE: $\sum P_i \cdot (i+1)$, where i is the intensity of the staining varying from 1 to 3 and P_i is the percentage of stained cells.

Statistical analysis

Quantitative expression of protein is expressed as mean values with standard deviation. Mean values in both groups were compared using t-test for normal distribution or non-parametric Wilcoxon test, as appropriate.

To take into account the complexity of biological networks of a large number of proteins, we used a statistical method of unsupervised clustering as has already been used in genomic research. PCA is a classic technique used to reduce the dimensionality of a dataset by transforming to a new set of variables (the principal components (PCs)), to summarize the features of the data. PCs are uncorrelated and the k th PC has the k th largest variance among all the PCs. The first PC can be interpreted as the direction that maximizes the variation of the projection of the data points. The second PC must be orthogonal to the first PC, ... The traditional approach is to use the first few PCs in the data analysis since they capture most of the variation in the original dataset.

Quantitative data of protein immunostaining can then be summarized in coherent patterns, related to their correlation with the PCs. Such dimensionality reduction can be a very useful step for visualizing and processing high-dimensional datasets, while retaining as much of the variance in the dataset as possible. EC type was used as a supplementary variable to help interpret the dimensions of variability. For the PCA, the expression values of each protein were standardized to zero mean and unit variance. Types of EC were then colorized on the dimensional space (the factorial map) and confidence ellipses were constructed.

In a second step, we grouped the different EC types into clusters, using the k -means method. This constitutes one of the simplest unsupervised learning algorithms that regroup data points into disjoint groups such that data points belonging to same cluster are similar while data points belonging to different clusters are dissimilar. The procedure is a simple way of classifying a given data set through k clusters, fixed *a priori*. The

main idea is to define k centroids, one for each cluster. In the first step, k initial "means" (centroid) are randomly generated within the data domain. The best choice is to place them as far away from each other as possible. In the next step, k clusters are created by associating each observation with the nearest centroid. Once no point is outstanding, an early groupage is done. At this point, k new centroids are recalculated as barycenters of the clusters resulting from the previous step and the centroid of each of the k clusters becomes the new centroid. After forming these k new centroids, steps 2 and 3 are repeated until convergence has been reached.

To know the optimal number of clusters (k), we plotted the inter-cluster distance function of the number of clusters. We chose the k which corresponded to the inflexion of the curve.

Once the high dimensional data were transformed into lower dimensional data via the PCA, we used the factorial map to project k-means clusters into this lower dimensional subspace. We could then confront the clusters to the real type of EC. Data were managed with an Excel database (Microsoft, Redmond, WA) and analysed using R 2.15.2 software with the *FactoMineR* package, available online.

Results

Immunohistochemical results according to histological type and grade of EC. Among the 43 specimens, 38 were type I EC (17 of G1, 13 of G2 and 8 of G3) and 5 were type II EC.

Table 1 and Figure 1 show the distribution of immunohistochemical protein expressions in type I G1, 2 or 3 and type II EC using univariate analysis. A difference in p53 ($p<0.001$), ER alpha ($p<0.001$), PRA ($p=0.008$) and TIMP-2 ($p<0.001$) expressions was noted between the groups. No differences in bcl-2, PRB, MMP-2, TIMP-1, MMP-2, MMP-7, CD44s, CD44v3, CD44v6 were observed between the groups while a trend for a difference was noted for MMP-9.

Results of principal component analysis

The first two dimensions of the PCA are shown in Figure 2. These two dimensions represented 43% of the total variance of the proteins (28.2% for dimension 1 and 14.6% for dimension 2). With three axes, more than half (54.1%) of the variance was reached. Table 2 summarizes the proteins with a significant contribution to each of the three axes.

Dimension 1 of the PCA separated proteins into pro-apoptotic (p53) and 'others'. Dimension 2 separated proteins into 2 groups: one comprising CD44v3, ER alpha, TIMP-2 and PR A and the other p53, CD44s and v6, MMP-2, -7 and -9, TIMP-1, bcl-2 and PR B. Schematically, this PCA identified three functional groups of proteins: the first characterized by p53 expression; the second by MMPs, bcl-2, PR B and CD44v6; and the third by ER alpha, PR A, TIMP-2 and CD44v3.

The *K*-means clustering analysis was performed on the 38 patients with no missing data on proteins. The *K*-means clustering found two main clusters of proteins. The characteristics of these clusters are summarized in Table 3. Distribution of histological types was significantly different: within cluster 1 there was 21.7% of type I G3 EC and 17.4% of type II EC; and in cluster 2 only 7.1% of type I G3 EC and no type II EC ($p=0.02$). LVSI was significantly higher in cluster 1 (63.6% in cluster 1 vs. 14.3% in cluster 2, $p=0.004$). Recurrence rate and lymph node involvement were higher in cluster 1 without reaching statistical significance.

The two clusters are represented on the factor map (dimension 1 and 2) (Figure 3). Cluster 1 is on the left of axis 2 (corresponding to the pro-apoptotic p53 protein in the PCA), and cluster 2 on the right side of this axis.

When the types and grades of EC were also represented on the factor map, an overlap in the confidence ellipses was observed among type I EC between G1 and 2. Among type 1 G2 EC, a heterogeneity was observed

with some tumours corresponding more to cluster 1 profile and others exhibiting cluster 2 profile. Finally, both type 1 G3 EC and type II EC exhibited cluster 1 profile.

Discussion

Using unsupervised clustering, the present study distinguishes two clusters of protein expression in EC. A panel of ten proteins were found to discriminate these two clusters by PCA. The first cluster corresponds to type I G3 and type II EC and the second to type I G1 and 2 EC.

From a pathological point of view, our results suggest that a continuum exists from type I G1 to type II EC involving additional protein alterations. Among the various immunohistochemical markers, PCA analysis found that the co-expression of steroid receptors (PR A et B, ER alpha), TIMP-1 and -2, MMP-7, CD44v6, bcl-2 and p53 were the most predictive, confirming their crucial role in endometrial carcinogenesis. These results are in agreement with those of previous studies focusing on the transition between simple hyperplasia and atypical endometrial hyperplasia [12] on one hand and between atypical endometrial hyperplasia and type I G1 EC on the other, with similar proteins involved in the processes [13]. Several retrospective studies have shown that ER and PR status in EC are independent prognostic markers [14–16]. Both PR A and PR B have been shown to modulate ER alpha and oestrogen-dependent gene expression [17]. High levels of expression are common in well-differentiated endometrioid EC and their presence is associated with a survival benefit compared with ER/PR negative tumours [18, 19]. Lapi Nska-Szcumicz *et al.* reported that the ER+/PR+/HER2- subtype was associated with the most favourable prognosis [20] whereas a loss of PR expression may lead to the development of a more invasive phenotype [21]. Saegusa and Okayasu reported that ER alpha shows a stepwise decrease from normal endometrium or low-grade EC to high-grade EC [22] while decreased expression in ER and PR was observed in tumours of poor prognosis such as advanced stages [23], G3 EC [24], clear cell carcinoma [25] and carcinosarcomas [26] as well as serous carcinoma that are mainly ER/PR negative [27]. Finally, in accordance with previous studies [24, 28], Voss *et al.*, reported no significant difference in overall rates of expression of ER/PR between G3 EC and type II [8].

In the current study, p53 expression was a key factor to distinguish various biological profiles of EC. We also found that bcl-2 strongly contributed to the dimension of the PCA. Our results are in agreement with previous studies showing that point mutations in p53 are found in 90% of type 2 EC but in only 10–20% of G3 type I EC [4, 29, 30]. The human TP53 gene encodes a nuclear protein that induces growth arrest or apoptosis in response to endogenous and exogenous stress. The functional inactivation of the p53 protein plays a crucial role in the malignant transformation process by increasing the capacity for tumour cell division and proliferation, and is consequently associated with poor survival [19, 32, 33]. The bcl-2 gene codes for a protein which inhibits

apoptotic cell death and is involved in the normal cell death program [34]. An imbalance in the bcl-2/bax ratio is known to convey apoptosis resistance in cancer cells [35]. However, important controversies exist on bcl-2 expression in EC [36, 37]. Bcl-2 persistence is frequently associated with EC, and failure to inactivate its expression is probably related to the development of EC [38]. BAG-1, bcl-2 and steroid receptor binding protein was shown to stain more frequently in high-grade cancers indicating an association between expression of BAG-1 and prognosis [39]. Moreover, our results are in accordance with those of the Cancer Genome Atlas (TCGA) highlighting the heterogeneity of EC by identifying 4 distinct genomic classes: 1) POLE ultra-mutated, 2) MSI hypermutated, 3) copy-number low microsatellite stable (MSS), and 4) copy-number high serous-like [40] displaying increasing TP53 and somatic copy-number alterations. The POLE ultra-mutated defined a subset of patients characterized by mutations in the exonuclease domain of the polymerase epsilon (POLE) gene, a high mutation load and an excellent prognosis.[41] Sixty percent of POLE ultra-mutated EC corresponded to high grade endometrioid EC with 35% exhibiting TP53 mutations.

In addition to steroid receptors and pro- and anti-apoptotic proteins, PCA analysis underlines that TIMP-1 and -2 and MMP-7 expression play a crucial role in EC pathophysiology. Low TIMP-2 expression has been linked to invasiveness in various malignancies [42]. Graesslin *et al.* showed that TIMP-2 expression was lower in type 2 EC (serous and clear cell) than in type 1 EC, and that low TIMP-2 expression was correlated with aggressiveness of the tumour [43]. A relation was also found between TIMP-2 downregulation in G3 EC compared to G1 and 2 but without difference between these two latter grades. These results are in agreement with those using microarray gene expression demonstrating that TIMP-2 is downregulated in high grade tumours [44]. Moreover, Graesslin *et al.*, reported a relation between low TIMP-2 expression and myometrial invasion, LVSI and lymph node metastasis, whereas Moser *et al.* found no relation between TIMP-2 expression and histological prognostic factors or overall survival [45]. Few data are available on the possible role of MMP-7 in EC. A previous study showed a relation between high MMP-7 expression and lymph node metastasis. Wang *et al.* demonstrated that MMP-7 activates progelatinases (proMMP-2 and proMMP-9) and thus facilitates tumour invasion and metastasis [46]. These results are consistent with our study showing that TIMP-1 and 2, as well as MMP-7, were selected as essential variables in the progression from type I G1 EC to type II EC.

There are limited and controversial reports on CD44 expression in EC [47, 48]. Our PCA analysis supports the contribution of CD44 in EC. Ayhan *et al.* [49] and Hoshimoto *et al.* [50] reported a relation between CD44v6 expression and poor prognosis while Gun *et al.* found no relation [51] underlining the need for further investigations.

Some limits of the present study have to be underlined. First, despite the use of the most relevant biological markers, we did not evaluate PTEN and E-cadherin, that have been recently suggested to be involved in the carcinogenesis of EC [30]. Type I ECs have been shown to display mutations in PTEN, PIK3CA, KRAS, and β -catenin, and microsatellite instability; while type II demonstrate genetic alterations in p53, HER-2/neu, p16, and E-cadherin [3].

Second, the low number of patients with type 2 EC could be a source of bias. However, it is important to underline that type 2 EC represent only about 20% of all EC. Third, even though PCA is a valuable tool to reduce dimensionality of protein expression data sets for visualization, caution should be exercised when interpreting cluster structures observed in the reduced dimensional subspace of the PCs. Yeung *et al.* showed that clustering with the PCs enhances cluster quality only when the right number of components or when the right set of PCs is chosen, and that using different data analysis techniques and different clustering algorithms to analyse the same data set can lead to very different conclusions [52]. Finally, several scoring systems (H-Score, Allred scoring system, ASCO/CAP criterion) have been advocated to evaluate the ER and PR positivity in endometrial cancer. To avoid bias with others markers evaluated in the current study, we chose a classic evaluation of ER and PR on hot spots without using Allred scoring system that has been demonstrated to be more sensitive and specific in breast cancer [53, 54].

In conclusion, using an unsupervised clustering method, our results indicate that there are sequential alterations in protein expression in EC contributing to a better understanding of the physiopathology of EC.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

1. Prat J (2004) Prognostic parameters of endometrial carcinoma. *Hum Pathol* 35:649–662.
2. Pansare V, Munkarah AR, Schimp V, et al (2007) Increased expression of hypoxia-inducible factor 1alpha in type I and type II endometrial carcinomas. *Mod Pathol* 20:35–43. doi: 10.1038/modpathol.3800718
3. Llobet D, Pallares J, Yeramian A, et al (2009) Molecular pathology of endometrial carcinoma: practical aspects from the diagnostic and therapeutic viewpoints. *J Clin Pathol* 62:777–785. doi: 10.1136/jcp.2008.056101
4. Lax SF (2004) Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. *Virchows Arch* 444:213–223. doi: 10.1007/s00428-003-0947-3
5. Colombo N, Preti E, Landoni F, et al (2011) Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 22 Suppl 6:vi35–39. doi: 10.1093/annonc/mdr374
6. Ballester M, Dubernard G, Lécuru F, et al (2011) Detection rate and diagnostic accuracy of sentinel-node biopsy in early stage endometrial cancer: a prospective multicentre study (SENTI-ENDO). *Lancet Oncol* 12:469–476. doi: 10.1016/S1470-2045(11)70070-5
7. Nugent EK, Bishop EA, Mathews CA, et al (2012) Do uterine risk factors or lymph node metastasis more significantly affect recurrence in patients with endometrioid adenocarcinoma? *Gynecol Oncol* 125:94–98. doi: 10.1016/j.ygyno.2011.11.049
8. Voss MA, Ganesan R, Ludeman L, et al (2012) Should grade 3 endometrioid endometrial carcinoma be considered a type 2 cancer—a clinical and pathological evaluation. *Gynecol Oncol* 124:15–20. doi: 10.1016/j.ygyno.2011.07.030
9. Alvarez T, Miller E, Duska L, Oliva E (2012) Molecular profile of grade 3 endometrioid endometrial carcinoma: is it a type I or type II endometrial carcinoma? *Am J Surg Pathol* 36:753–761. doi: 10.1097/PAS.0b013e318247b7bb
10. Zannoni GF, Vellone VG, Arena V, et al (2010) Does high-grade endometrioid carcinoma (grade 3 FIGO) belong to type I or type II endometrial cancer? A clinical–pathological and immunohistochemical study. *Virchows Archiv* 457:27–34. doi: 10.1007/s00428-010-0939-z
11. Poncelet C, Walker F, Madelenat P, et al (2001) Expression of CD44 standard and isoforms V3 and V6 in uterine smooth muscle tumors: a possible diagnostic tool for the diagnosis of leiomyosarcoma. *Hum Pathol* 32:1190–1196.
12. Laas E, Ballester M, Cortez A, et al (2014) Supervised clustering of immunohistochemical markers to distinguish atypical and non-atypical endometrial hyperplasia. *Gynecol Endocrinol* 1–4. doi: 10.3109/09513590.2014.989981
13. Laas E, Ballester M, Cortez A, et al (2014) Supervised clustering of immunohistochemical markers to distinguish atypical endometrial hyperplasia from grade 1 endometrial cancer. *Gynecol Oncol*. doi: 10.1016/j.ygyno.2014.02.018

14. Zhang G, Wu L, Li B, et al (2013) Retrospective analysis of prognostic variables and clinical outcomes in surgically staged intermediate risk endometrial carcinoma. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 169:309–316. doi: 10.1016/j.ejogrb.2013.02.025
15. Fukuda K, Mori M, Uchiyama M, et al (1998) Prognostic significance of progesterone receptor immunohistochemistry in endometrial carcinoma. *Gynecol Oncol* 69:220–225. doi: 10.1006/gyno.1998.5023
16. Singh P, Smith CL, Cheetham G, et al (2008) Serous carcinoma of the uterus-determination of HER-2/neu status using immunohistochemistry, chromogenic in situ hybridization, and quantitative polymerase chain reaction techniques: its significance and clinical correlation. *Int J Gynecol Cancer* 18:1344–1351. doi: 10.1111/j.1525-1438.2007.01181.x
17. Giangrande PH, McDonnell DP (1999) The A and B isoforms of the human progesterone receptor: two functionally different transcription factors encoded by a single gene. *Recent Prog Horm Res* 54:291-313; discussion 313-314.
18. Wik E, Ræder MB, Krakstad C, et al (2013) Lack of Estrogen Receptor- α Is Associated with Epithelial-Mesenchymal Transition and PI3K Alterations in Endometrial Carcinoma. *Clin Cancer Res*. doi: 10.1158/1078-0432.CCR-12-3039
19. Jongen V, Briët J, de Jong R, et al (2009) Expression of estrogen receptor-alpha and -beta and progesterone receptor-A and -B in a large cohort of patients with endometrioid endometrial cancer. *Gynecol Oncol* 112:537–542. doi: 10.1016/j.ygyno.2008.10.032
20. Lapi Nska-Szumczyk S, Supernat A, Majewska H, et al (2014) HER2-Positive Endometrial Cancer Subtype Carries Poor Prognosis. *Clin Transl Sci*. doi: 10.1111/cts.12207
21. Hanekamp EE, Gielen SC, Smid-Koopman E, et al (2003) Consequences of loss of progesterone receptor expression in development of invasive endometrial cancer. *Clinical cancer research* 9:4190–4199.
22. Saegusa M, Okayasu I (2000) Changes in expression of estrogen receptors alpha and beta in relation to progesterone receptor and pS2 status in normal and malignant endometrium. *Jpn J Cancer Res* 91:510–518.
23. Sho T, Hachisuga T, Nguyen TT, et al (2014) Expression of estrogen receptor- α as a prognostic factor in patients with uterine serous carcinoma. *Int J Gynecol Cancer* 24:102–106. doi: 10.1097/IGC.0000000000000029
24. Soslow RA, Shen PU, Chung MH, et al (2000) Cyclin D1 expression in high-grade endometrial carcinomas--association with histologic subtype. *Int J Gynecol Pathol* 19:329–334.
25. Vang R, Whitaker BP, Farhood AI, et al (2001) Immunohistochemical analysis of clear cell carcinoma of the gynecologic tract. *Int J Gynecol Pathol* 20:252–259.

26. Ansink AC, Cross PA, Scorer P, et al (1997) The hormonal receptor status of uterine carcinosarcomas (mixed müllerian tumours): an immunohistochemical study. *J Clin Pathol* 50:328–331.
27. Demopoulos RI, Mesia AF, Mittal K, Vamvakas E (1999) Immunohistochemical comparison of uterine papillary serous and papillary endometrioid carcinoma: clues to pathogenesis. *Int J Gynecol Pathol* 18:233–237.
28. Reid-Nicholson M, Iyengar P, Hummer AJ, et al (2006) Immunophenotypic diversity of endometrial adenocarcinomas: implications for differential diagnosis. *Mod Pathol* 19:1091–1100. doi: 10.1038/modpathol.3800620
29. Tashiro H, Isacson C, Levine R, et al (1997) p53 gene mutations are common in uterine serous carcinoma and occur early in their pathogenesis. *Am J Pathol* 150:177–185.
30. Matias-Guiu X, Davidson B (2014) Prognostic biomarkers in endometrial and ovarian carcinoma. *Virchows Archiv* 464:315–331. doi: 10.1007/s00428-013-1509-y
31. Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. *Nature* 408:307–310. doi: 10.1038/35042675
32. Alkushi A, Lim P, Coldman A, et al (2004) Interpretation of p53 immunoreactivity in endometrial carcinoma: establishing a clinically relevant cut-off level. *Int J Gynecol Pathol* 23:129–137.
33. Dupont J, Wang X, Marshall DS, et al (2004) Wilms Tumor Gene (WT1) and p53 expression in endometrial carcinomas: a study of 130 cases using a tissue microarray. *Gynecol Oncol* 94:449–455. doi: 10.1016/j.ygyno.2004.05.014
34. Mitselou A, Ioachim E, Kitsou E, et al (2003) Immunohistochemical study of apoptosis-related Bcl-2 protein and its correlation with proliferation indices (Ki67, PCNA), tumor suppressor genes (p53, pRb), the oncogene c-erbB-2, sex steroid hormone receptors and other clinicopathological features, in normal, hyperplastic and neoplastic endometrium. *In Vivo* 17:469–477.
35. Zhang M, Zhang P, Zhang C, et al (2009) Prognostic significance of Bcl-2 and Bax protein expression in the patients with oral squamous cell carcinoma. *J Oral Pathol Med* 38:307–313. doi: 10.1111/j.1600-0714.2008.00689.x
36. Miturski R, Semczuk A, Tomaszewski J, Jakowicki J (1998) bcl-2 protein expression in endometrial carcinoma: the lack of correlation with p53. *Cancer Lett* 133:63–69.
37. Zhang R, He Y, Zhang X, et al (2012) Estrogen receptor-regulated microRNAs contribute to the BCL2/BAX imbalance in endometrial adenocarcinoma and precancerous lesions. *Cancer Lett* 314:155–165. doi: 10.1016/j.canlet.2011.09.027
38. Zheng W, Cao P, Zheng M, et al (1996) p53 overexpression and bcl-2 persistence in endometrial carcinoma: comparison of papillary serous and endometrioid subtypes. *Gynecol Oncol* 61:167–174. doi: 10.1006/gyno.1996.0120
39. Moriyama T, Littell RD, Debernardo R, et al (2004) BAG-1 expression in normal and neoplastic endometrium. *Gynecol Oncol* 94:289–295. doi: 10.1016/j.ygyno.2004.04.026

40. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, et al (2013) Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67–73. doi: 10.1038/nature12113
41. Hussein YR, Weigelt B, Levine DA, et al (2014) Clinicopathological analysis of endometrial carcinomas harboring somatic POLE exonuclease domain mutations. *Mod Pathol*. doi: 10.1038/modpathol.2014.143
42. Uzan C, Cortez A, Dufournet C, et al (2004) Eutopic endometrium and peritoneal, ovarian and bowel endometriotic tissues express a different profile of matrix metalloproteinases-2, -3 and -11, and of tissue inhibitor metalloproteinases-1 and -2. *Virchows Arch* 445:603–609. doi: 10.1007/s00428-004-1117-y
43. Graesslin O, Cortez A, Uzan C, et al (2006) Endometrial tumor invasiveness is related to metalloproteinase 2 and tissue inhibitor of metalloproteinase 2 expressions. *Int J Gynecol Cancer* 16:1911–1917. doi: 10.1111/j.1525-1438.2006.00717.x
44. Ferguson SE, Olshen AB, Viale A, et al (2004) Gene expression profiling of tamoxifen-associated uterine cancers: evidence for two molecular classes of endometrial carcinoma. *Gynecol Oncol* 92:719–725. doi: 10.1016/j.ygyno.2003.10.038
45. Moser PL, Kieback DG, Hefler L, et al (1999) Immunohistochemical detection of matrix metalloproteinases (MMP) 1 and 2, and tissue inhibitor of metalloproteinase 2 (TIMP 2) in stage IB cervical cancer. *Anticancer Res* 19:4391–4393.
46. Wang F-Q, So J, Reierstad S, Fishman DA (2005) Matrilysin (MMP-7) promotes invasion of ovarian cancer cells by activation of progelatinase. *Int J Cancer* 114:19–31. doi: 10.1002/ijc.20697
47. Katsura M, Furumoto H, Nishimura M, et al (1998) Overexpression of CD44 variants 6 and 7 in human endometrial cancer. *Gynecol Oncol* 71:185–189. doi: 10.1006/gyno.1998.5169
48. Tempfer C, Haeusler G, Kaider A, et al (1998) The prognostic value of CD44 isoform expression in endometrial cancer. *British journal of cancer* 77:1137.
49. Ayhan A, Tok EC, Bildirici I, Ayhan A (2001) Overexpression of CD44 Variant 6 in Human Endometrial Cancer and Its Prognostic Significance. *Gynecologic Oncology* 80:355–358. doi: 10.1006/gyno.2000.6014
50. Hoshimoto K, Yamauchi N, Takazawa Y, et al (2003) CD44 variant 6 in endometrioid carcinoma of the uterus: its expression in the adenocarcinoma component is an independent prognostic marker. *Pathology-Research and Practice* 199:71–77.
51. Gun BD, Bahadir B, Bektas S, et al (2012) Clinicopathological significance of fascin and CD44v6 expression in endometrioid carcinoma. *Diagn Pathol* 7:15.
52. Yeung KY, Ruzzo WL (2001) Principal component analysis for clustering gene expression data. *Bioinformatics* 17:763–774.
53. Wang Y, Ma X, Xi C, et al (2013) [Correlation between estrogen receptor status and clinicopathologic parameters in endometrial cancer: a comparative study by

immunohistochemistry using different scoring systems]. Zhonghua Bing Li Xue Za Zhi 42:509–514.

54. Qureshi A, Pervez S (2010) Allred scoring for ER reporting and it's impact in clearly distinguishing ER negative from ER positive breast cancers. J Pak Med Assoc 60:350–353.

Figure caption

Fig1 Distribution of protein expressions according to endometrial cancer type and grade

1a:Pro/antiapoptotic proteins and steroid receptors

G : grade, EC endometrial cancer; ER Estrogen receptor; PR progesterone receptor

1b: Metalloproteinases and tissue inhibitors of metalloproteinases

G : grade, EC :endometrial cancer, MMP: Metalloproteinase ; TIMP: tissue inhibitor of metalloproteinase

1c: CD44s, CD44v3 and v6

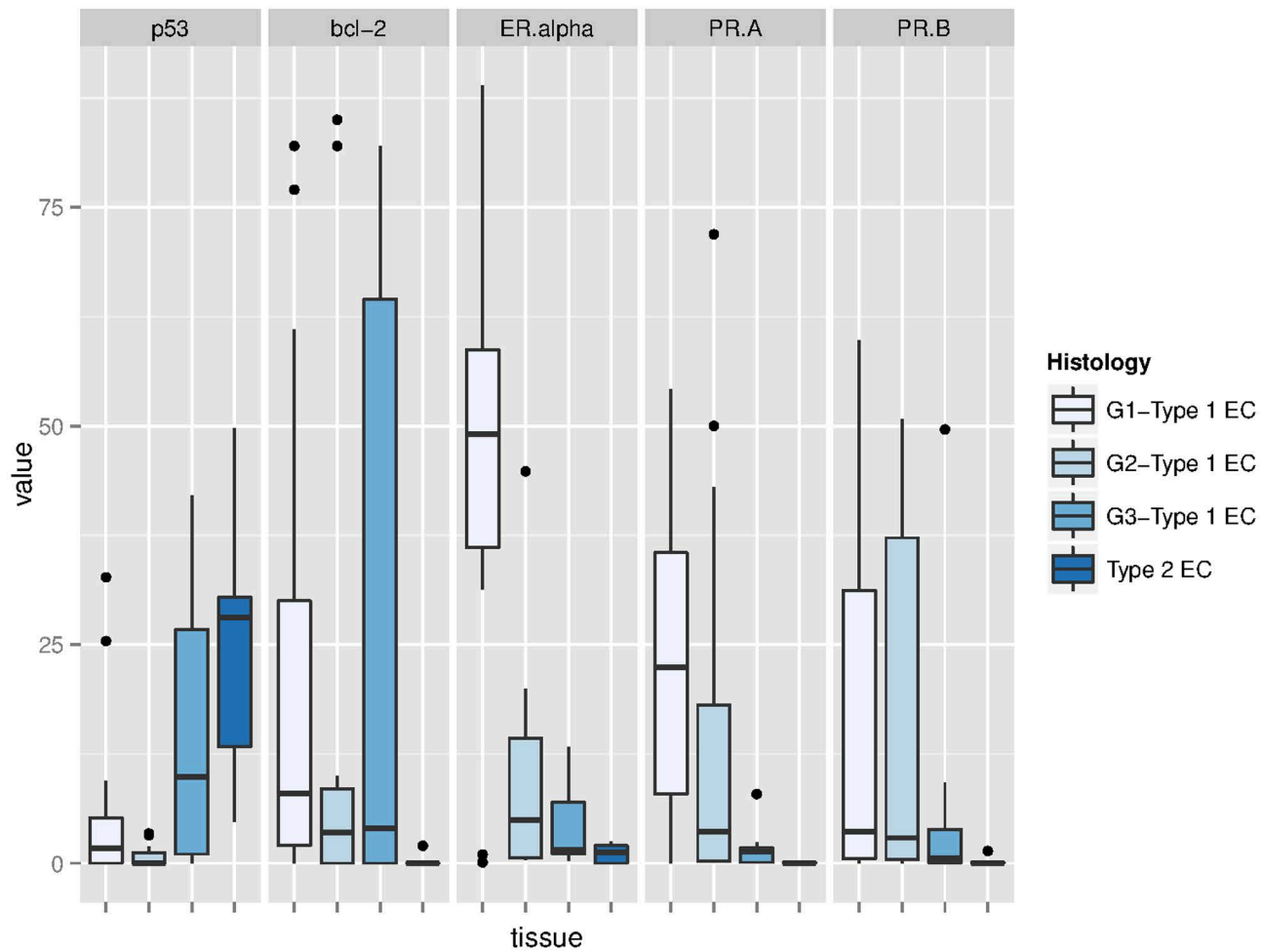
G : grade, EC endometrial cancer

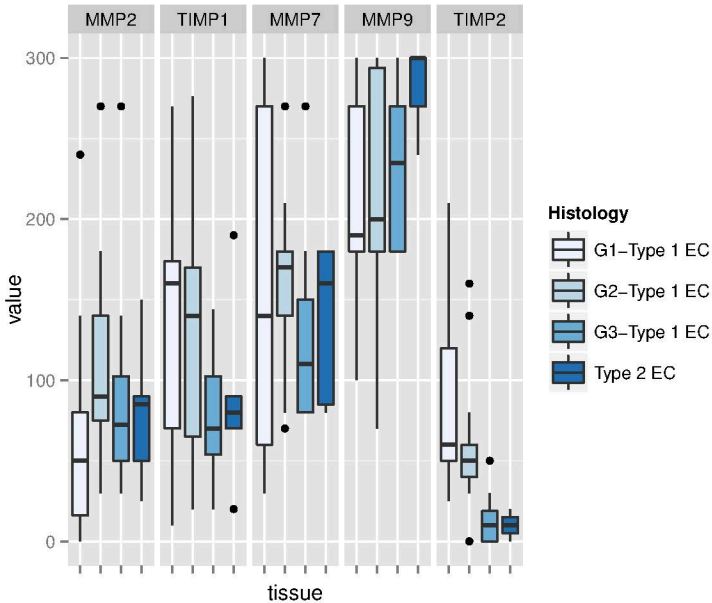
Fig2 Principal component analysis: variable factor map

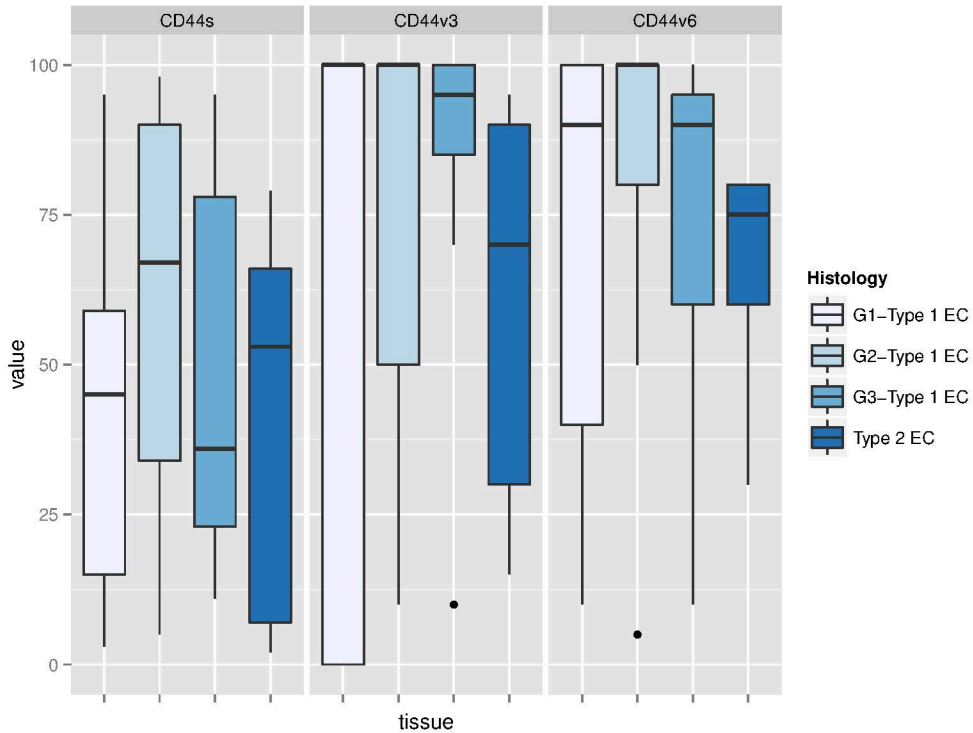
Dim : dimension; ER: Estrogen receptor; PR: progesterone receptor; MMP: Metalloproteinase ; TIMP: tissue inhibitor of metalloproteinase

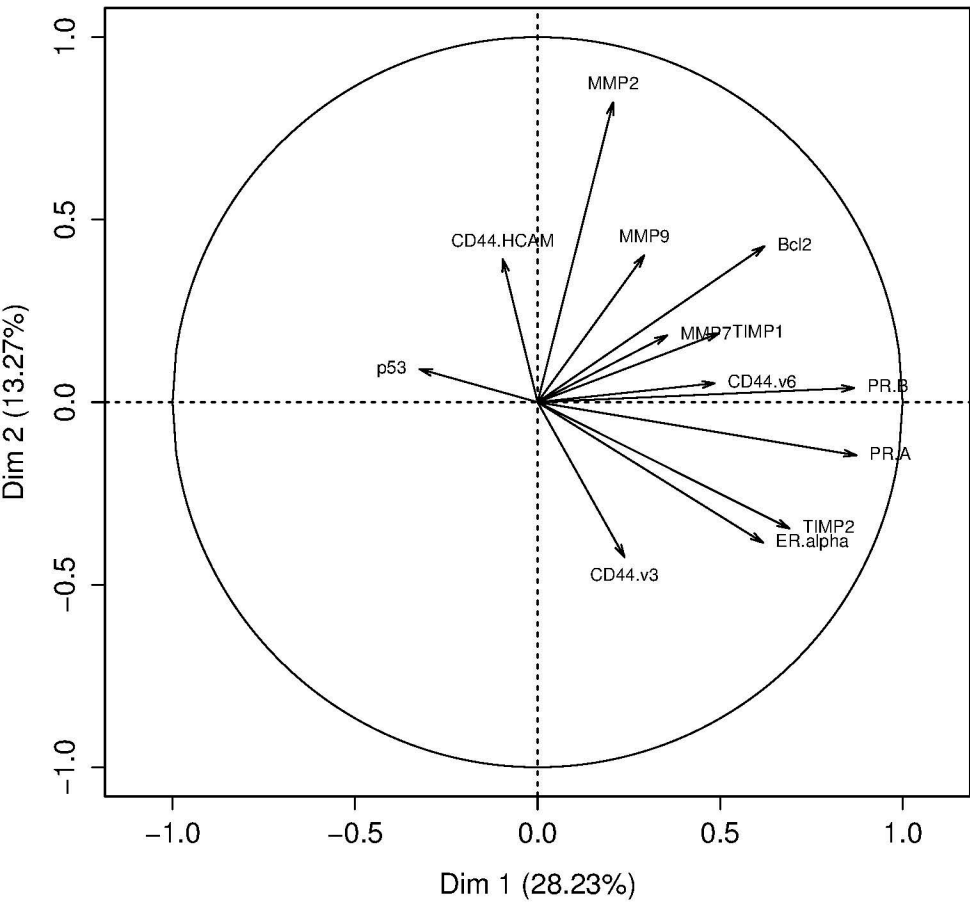
Fig3 Principal component analysis: individual factor map with the clusters (defined by the K-means method) and the confidence ellipses according to types and grades of EC.

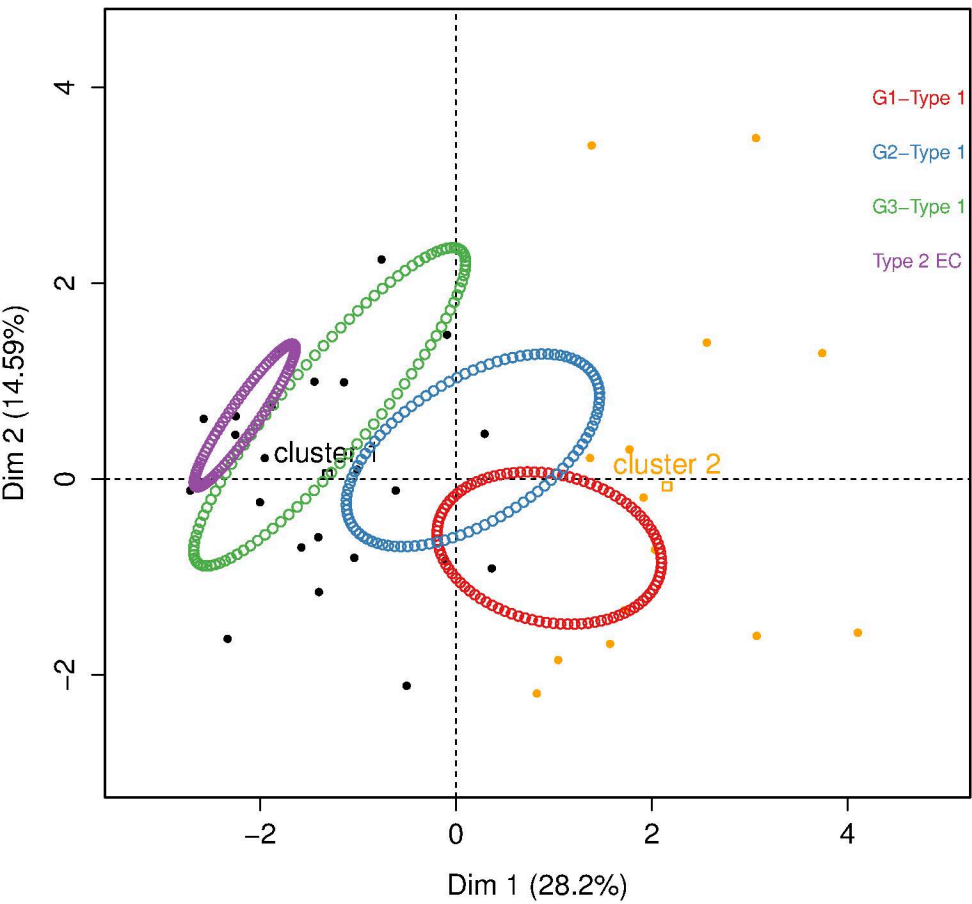
Dim : dimension; G: grade; EC: endometrial cancer











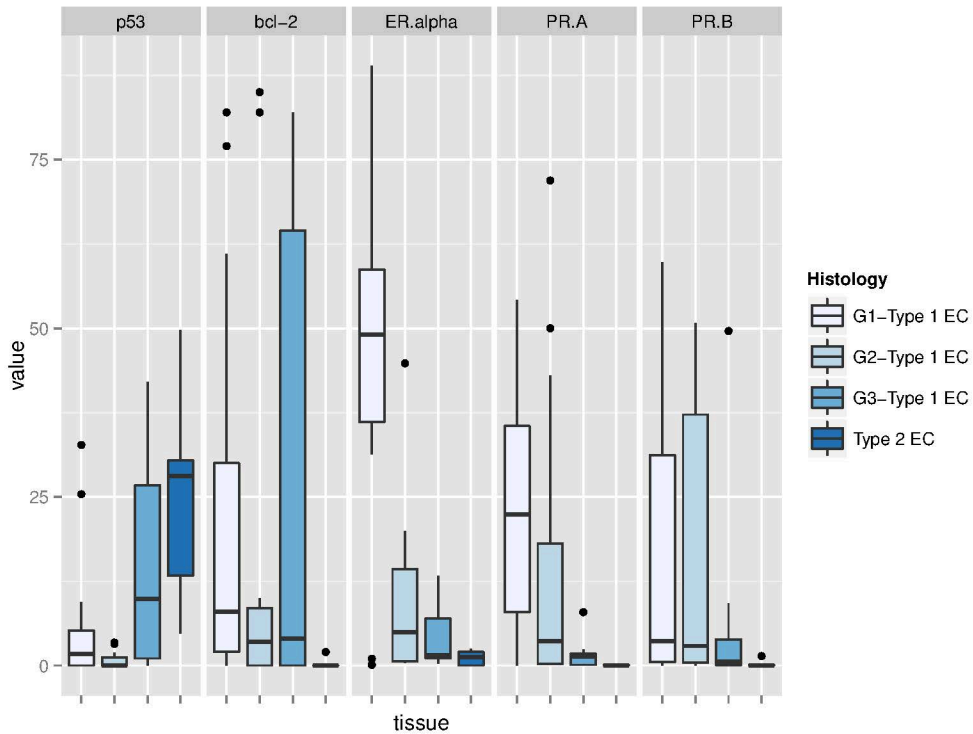


Table 1. Proteins immunostaining in endometrial cancer according the histological types and grades.

	G1-Type I EC (n=17)	G2-Type I EC (n=13)	G3-Type I EC (n=8)	Type II EC (n=5)	p-value
p53	5.4 (9.4)	0.8 (1.25)	15.7 (17.5)	25.2 (17.3)	<0.001
bcl-2	22.5 (27.7)	16.7 (31.5)	28 (36.9)	0.4 (0.894)	0.4
ER alpha	45.1 (24.9)	9.1 (12.6)	4.2 (4.7)	1.1 (1.14)	<0.001
PR A	23.9 (16.3)	16.1 (23.6)	1.8 (2.59)	0.06 (0.09)	0.008
PR B	16.3 (19)	16.6 (21)	7.8 (17.2)	0.3 (0.6)	0.3
MMP-2	61.9 (61.9)	107.3 (65.2)	96.9 (77.7)	80 (47.3)	0.3
TIMP-1	130.5 (72.7)	120.9 (74.8)	79.9 (43.6)	90 (62)	0.3
MMP-7	157.9 (97.7)	166.9 (61.8)	131.2 (67.9)	137 (50.4)	0.7
MMP-9	208.6 (54.5)	208 (74.7)	231.2 (50.8)	282 (26.8)	0.09
TIMP-2	87.3 (55.4)	61.1 (43.8)	14.4 (17.6)	10 (7.9)	<0.001
CD44s	44.7 (29.6)	59.9 (35.6)	49.1 (34)	41.4 (35)	0.6
CD44v3	64.8 (49.1)	78.8 (29.9)	82.5 (31.1)	60 (35.9)	0.6
CD44v6	70 (34.8)	84.6 (27.9)	72.9 (36.7)	65 (23.8)	0.6

G=Grade; EC= Endometrial cancer; PR Progesteron Receptor; ER Estrogen receptor; MMP Metalloproteinase;

TIMP : tissue inhibitor of metalloproteinase

Table 2. Contribution of proteins to the three first principal components

DIMENSION 1			DIMENSION 2			DIMENSION 3		
PR A	0.86	<0.001	MMP-2	0.83	<0.001	p53	0.6	<0.001
PR B	0.84	<0.001	bcl-2	0.52	<0.001	MMP-9	0.55	<0.001
TIMP-2	0.68	<0.001	CD44s	0.52	<0.001	CD44v6	0.55	<0.001
ER alpha	0.62	<0.001	MMP-9	0.40	0.01	CD44s	-0,35	0.03
TIMP-1	0.60	<0.001	CD44v3	-0,33	0.04	MMP-7	-0,39	0.01
bcl-2	0.52	<0.001	ER alpha	-0,39	0.01			
CD44v6	0.44	<0.01						
MMP-7	0.33	0.04						
p53	-0,5	0.01						

PR Progesteron Receptor; ER Estrogen receptor; MMP Metalloproteinase; TIMP : tissue inhibitor of metalloproteinase

Table 3. Characteristics of the clusters defined by the K-means method*

		Cluster 1 (n=23)	Cluster 2 (n=14)	p-value
Histology	G1-type I EC	5 (21.7)	10 (71.4)	0.02
	G2-type I EC	9 (39.1)	3 (21.4)	
	G3-type I EC	5 (21.7)	1 (7.1)	
	Type II EC	4 (17.4)	0 (0)	
Stage	IA	7 (30.4)	6(42.9)	0.9
	IB	6 (26.1)	2 (14.3)	
	II	3 (13)	3 (21.4)	
	III	6 (26.1)	3 (21.4)	
	IV	1 (4.3)	0 (0)	
Myometrial infiltration	None	5 (21.7)	1 (7.1)	0.3
	<50%	6 (26.1)	7 (50)	
	≥50%	12 (52.2)	6 (42.9)	
LVSI		14 (63.6)	2 (14.3)	0.004
Lymph node involvement		7 (46.7)	1 (9.1)	0.1
Recurrence		3 (13)	0 (0)	0.3

*Patients with no missing data for protein (n=38). G=Grade ; EC= Endometrial cancer, LVSI

= Lymphovascular space involvement

