SUPPLEMENTARY MATERIALS

Association between proportion of nuclei with high chromatin entropy and prognosis in gynecological cancers

Birgitte Nielsen1,3, Andreas Kleppe1,2,3, Tarjei Sveinsgjerd Hveem1,3, Manohar Pradhan1,3, Rolf Anders Syvertsen1,3, John Arne Nesheim1,3, Gunnar Balle Kristensen1,4, Jone Trovik5,6, David Kerr7, Fritz Albregtsen1,2, Håvard Emil Danielsen1,2,3,7

1Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo, Norway
2Department of Informatics, University of Oslo, Oslo, Norway
3Center for Cancer Biomedicine, University of Oslo, Oslo, Norway
4Department of Gynecologic Oncology, Oslo University Hospital, Oslo, Norway
5Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway
6Center for Cancer Biomarkers, Department of Clinical Science, University of Bergen, Bergen, Norway
7Nuffield Division of Clinical Laboratory Sciences, University of Oxford, United Kingdom

Correspondence to: Prof. Håvard Emil Danielsen, Institute for Cancer Genetics and Informatics, Oslo University Hospital, NO-0424 Oslo, Norway. Tel. +47 22782320. e-mail: hdaniels@ifi.uio.no
Supplementary materials and methods

Sample preparation and imaging

Curettage specimens taken from endometrial tumors and tumor tissue from surgically removed ovarian cancer and uterine sarcomas were formalin-fixed and paraffin-embedded. Haematoxylin and eosin-stained sections from each tumor were evaluated to identify representative regions with tumor tissue. For each tumor, isolated nuclei were prepared from one or more 50 µm sections and stained with Feulgen-Schiff according to an established protocol (1).

A DNA Ploidy System, which consisted of a Zeiss Axioplan microscope equipped with a 546 nm green filter and a monochrome high-resolution digital camera (AxioCam MRm, Zeiss, Jena, Germany or C4742-95, Hamamatsu Photonics, Japan), was used to capture images. The grey level resolution was 10 bits/pixel and the physical resolution was 166 nm/pixel for the uterine and ovarian specimens and 162 nm/pixel for the endometrial specimens.

Cell nuclei were automatically segmented and the cell type of the nuclei was assigned by an automated cell classification system. Trained personnel verified the selected nuclei such that only tumor nuclei, and only epithelial nuclei in the carcinoma data sets, were included in the analyses. Necrotic nuclei, doublets, cut and folded nuclei were discarded. In order to compensate for technical variation in sample preparation and imaging setup across the three materials, the size and grey levels of the nuclear images were normalized based on automatic detection of diploid cells in each sample (2).

Nuclear texture analysis by grey level entropy matrices

A grey level entropy matrix (GLEM) (3,4) was computed from each nuclear image. Such a matrix contains information about the bi-variate distribution of pixel grey level intensities and corresponding grey level entropy values measured locally around the pixels.

The GLEM element value \( P(i,j) \) is the normalized frequency of the first order Shannon grey level entropy value \( j \) within a window of size \( w \times w \) centered on a pixel with grey level value \( i \) (3). The grey level entropy is defined as

\[
 j = -\sum_i P(i) \times \log P(i), \ 0 < P(i),
\]

where \( P(i) \) is the normalized frequency of grey level \( i \) within the window and \( \log \) is the natural logarithm. The entropy measures the grey level heterogeneity within the window. Homogeneous chromatin structures will give low entropy values whereas disorganized chromatin structures will give high entropy values. The number of grey levels in the nuclear images was reduced by uniform re-quantification to 64 before computation of the GLEMs and the window size was set to 15×15 pixels just as in (5).

Measuring the compactness and center of mass of a GLEM

In previous studies of early ovarian cancer (5), uterine sarcoma (6) and endometrial cancer (7), we observed that the (grey level, entropy)-values in nuclei from good prognosis patients were (on average) more concentrated to a smaller fraction of the GLEM compared to values in nuclei from poor prognosis cases, and in a study for assessment of dysplasia (8), we found that (grey level, entropy)-values in nuclei from dysplastic cases showed more variation compared to nuclei from non-dysplastic cases. In all these studies we also observed a shift in the mean of the GLEM distributions that was related to the concentration/variation of the (grey level, entropy)-values.

Classical texture features extracted from a matrix, such as e.g. a grey level co-occurrence matrix (GLCM) (9) or a GLEM(3), may be seen as a weighted sum of the matrix element values, where the weighting applied to each element is based on a given weighting function. By varying this weighting function, different types of information about the texture can be extracted (4). In previous studies, we have used adaptive weighting functions that are learned from a discovery set and are based on average differences between clinical classes (2,5-8,10). In the present study, we propose a new way of extracting features from a GLEM that is not based on computation of a weighted sum. We instead calculate three novel features (defined below) that measure the compactness and the location (mean) of the GLEM probability distribution. The aim of the study was to evaluate if these novel features could be useful for (i) capturing the concentration/variation in a GLEM, (ii) identifying a
small subgroup of nuclei with high chromatin entropy and (iii) classifying nuclei and cancer patients according to prognosis.

Center of mass

The center of mass in a GLEM, i.e. the center of mass in the bi-variate distribution of (grey level, entropy)-values, was computed as the mean grey level and the mean entropy value.

Relative matrix area

The probability values in the GLEM were sorted in descending order, and then the values were summed until the total sum was greater than or equal to 0.25. The GLEM concentration was measured as the number of matrix elements included in the summation divided by the total number of matrix elements (i.e., the relative matrix area of the summed elements). This measure is a dimensionless relative spread feature, invariant to location, shape and rotation.

Sarcoma discovery and validation sets

The uterine sarcomas were divided into a discovery set and a validation set by sorting the sarcoma patients by date of diagnosis and then assigning every second sarcoma (starting with the first one) to the discovery set and the other sarcomas to the validation set. The two data sets were subsequently balanced for histology type by moving seven leiomyosarcomas from the validation set to the discovery set, and five endometrial stromal sarcomas and two adenosarcomas from the discovery to the validation set (6). To move the leiomyosarcomas, the 129 leiomyosarcomas in the validation set were sorted by date of diagnosis and every eighteenth leiomyosarcoma was moved to the discovery set. Similarly, the 46 endometrial stromal sarcomas in the discovery set were sorted by date of diagnosis and every ninth endometrial stromal sarcoma was moved to the validation set, and the 14 adenosarcomas were sorted by date of diagnosis and every seventh adenosarcoma was moved to the validation set.

Prognostic grouping of early censored patients

In the present study, the design and evaluation of the chromatin entropy marker were based on a grouping of patients into good and poor prognosis classes (Supplementary Tables 4-5). Due to censoring, this prognostic grouping of patients may cause problems for early censored patients. The 184 uterine sarcoma patients who survived for at least five years were defined as good prognosis, whereas the 170 patients who died within five years were defined as poor prognosis. All patients still alive at the end of the uterine sarcoma study were followed up for more than five years, so in this material early censored patients caused no problems.

In the ovarian cancer material, good prognosis was defined as no recurrence of disease (n=173), whereas poor prognosis was defined as recurrence of disease (n=73). The good prognosis patients had a medium follow-up time of 12.9 years (IQR, 11.3-14.4), whereas the poor prognosis patients had a medium follow-up time of 1.9 years (IQR, 0.8-4.4). Among the poor prognosis patient, 72 (of 73) patients experienced recurrence within 10 years and only one patient experienced recurrence after 10 years, specifically this recurrence was detected 11.1 years after surgery. Among the good prognosis patients, only 12 (of 173) patients were followed up for less than 10 years. In this cohort, early censoring will thus cause uncertainty about the prognostic grouping in only a very few patients.

In the endometrial cancer material, good prognosis was defined as no endometrial cancer death (n=707), whereas poor prognosis was defined as cancer-specific death (n=84). The good prognosis patients had a median follow-up time of 37.0 months (IQR, 19.0-55.0), whereas the poor prognosis patients had a median follow-up time of 14.0 months (IQR, 6.0-28.8). In this cohort, the relatively short follow-up time caused uncertainty in the prognostic grouping of the early censored patients that were defined as good prognosis.

Chromatin entropy marker

We have previously shown that nuclear texture varies with nuclear area and that prognostication can be enhanced by grouping the nuclear images according to their area (5,6,11). Analyses in the discovery set of uterine sarcoma showed that nuclei with 2,000-4,999 pixels had sufficiently similar texture with respect to the three described features to be analyzed together, and the new texture biomarker was therefore computed using only these nuclei. After applying the normalization method (2), 2,000-4,999 pixels correspond to a nuclear area between 54 \( \mu m^2 \) and 134 \( \mu m^2 \).
**k-means clustering**

$k$-means clustering was used to partition the data points in the three-dimensional feature space (comprised of relative matrix area and center of mass in grey level and entropy) into $k=5$ clusters. The squared Euclidean distance was used to measure distance between data points. The data points represented all nuclei (with nuclear area 2,000-4,999 pixels) in the discovery set, in total 158,868 nuclei from uterine sarcoma (Supplementary Figure 2A). Prior to clustering, each of the three features was standardized by subtracting the mean and dividing by the standard deviation (SD) of the feature values of the 158,868 nuclei. The five clusters were sorted and named cluster 1 to 5, such that increasing cluster number corresponded to increasing relative matrix area of cluster centroid. Supplementary Figure 2B shows the average GLEM computed from all nuclei in each of the five clusters, while Supplementary Figure 2C shows a boxplot of the relative matrix areas in the clusters.

For each patient in the discovery set, the proportion of nuclei in each of the five clusters was computed. The average proportion of nuclei in each cluster was computed based on the discovery cases in the two prognostic classes. The proportion of nuclei in cluster 5 (corresponding to nuclei with high relative matrix area and defined as nuclei with high chromatin entropy) was statistically significantly higher ($P<0.001$, Mann-Whitney $U$ test) for patients with a poor prognosis compared to good prognosis patients (16.6% [SD=15.3] versus 9.3% [SD=10.5] high chromatin entropy nuclei). The average proportion of nuclei in clusters 3 and 4 was also higher for poor prognosis patients compared to good prognosis patients, but the difference was not statistically significant (25.4% [SD=14.9] versus 22.4% [SD=10.8], $P=0.23$ in cluster 3 and 20.4% [SD=12.4] versus 19.2% [SD=12.5], $P=0.47$ in cluster 4), while the average proportion of nuclei in clusters 1 and 2 was lower for poor prognosis than for good prognosis patients (15.3% [SD=13.9] versus 17.2% [SD=11.5], $P=0.07$ for cluster 1 and 22.3% [SD=12.9] versus 31.9% [SD=14.3], $P<0.001$ for cluster 2).

**Patient classification**

We chose to focus on the cluster corresponding to the nuclei with highest entropy, i.e. cluster 5, and wanted to select a threshold on the proportion of nuclei with high chromatin entropy (the proportion of nuclei in cluster 5). Tumors with a proportion of high chromatin entropy nuclei less than the threshold value were to be classified as low chromatin entropy whereas tumors with a proportion of high chromatin entropy nuclei more or equal to the threshold were to be classified as high chromatin entropy. Correct classification rate, sensitivity, and specificity on the uterine sarcoma discovery set were computed for all different threshold values for the proportion of high chromatin entropy nuclei ($0\% - 100\%$, Supplementary Figure 3). The correct classification rate varied between 58% and 62% in a large interval of threshold values (thresholds from 8% to 27%), and then started to decrease with increasing threshold values. In that interval of threshold values, the specificity increased and the sensitivity decreased with increasing threshold value (Supplementary Figure 3B). Martingale residuals in the null Cox model and in multivariate analysis with established clinicopathological variables (where chromatin entropy is not included in the model) plotted against the proportion of high chromatin entropy nuclei indicated a piecewise linear relation (Supplementary Figure 4). Supplementary Figure 5 shows hazard ratios obtained in univariate and multivariate 5-year overall survival analyses among the 175 uterine sarcomas in the discovery set for different threshold values for the proportion of high chromatin entropy nuclei. Considering the same interval of thresholds (8% to 27%), chromatin entropy was statistically significant in univariate analysis for all thresholds, with highest hazard ratios corresponding to threshold values around 20% and 25%, and lower hazard ratios in the interval defined by average patient values (of proportion of high chromatin entropy nuclei) in the two prognostic groups (i.e., 9.3% [SD=10.5] and 16.6% [SD=15.3], Supplementary Table 5). In multivariate analysis with clinically relevant variables (the same as in Table 1), the chromatin entropy marker was statistically significant only for threshold values 24-26. Therefore, we chose to set the threshold as high as 25%, resulting in a sensitivity of 25.3% (95% CI=16.4-36.0), a specificity of 91.3% (95% CI=83.6-96.2) (Supplementary Table 4), and hazard ratios of 2.13, 95% CI=1.30-3.49, in univariate and 1.91, 95% CI=1.04-3.50, in multivariate analyses on the uterine sarcoma discovery set.

**Validation of the chromatin entropy marker**

For each nuclear image with 2,000-4,999 pixels in the uterine sarcoma validation set, in total 167,381 nuclear images, we calculated the squared Euclidean distance between the three-dimensional data point (comprised of the relative matrix area and the center of mass feature values) and each of the five cluster centroids. The nucleus was assigned to the cluster with the nearest centroid.

In the same way, each data point in the ovarian cancer set (corresponding to 53,027 nuclei) and in the endometrial cancer set (623,595 nuclei) was assigned to one of the five clusters. The feature values computed from all three validation sets were normalized using the mean and standard deviation computed from the uterine sarcoma validation set.
sarcoma discovery set. Statistics on the proportion of nuclei/patient in cluster 5 for good and poor prognosis patients in the three data sets are shown in Supplementary Table 5.
References

1. Pradhan M, Abeler VM, Danielsen HE, Trope CG, Risberg B. Image cytometry DNA ploidy correlates with histological subtypes in endometrial carcinomas. *Mod Pathol.* 2006;19(9):1227-1235.

2. Hveem TS, Kleppe A, Vlatkovic L, et al. Chromatin changes predict recurrence after radical prostatectomy. *Br J Cancer.* 2016;114(1):1243-1250.

3. Yogesan K, Jørgensen T, Albregtsen F, Tveter KJ, Danielsen HE. Entropy-based texture analysis of chromatin structure in advanced prostate cancer. *Cytometry A.* 1996;24(3):268-276.

4. Nielsen B, Albregtsen F, Danielsen HE. Statistical nuclear texture analysis in cancer research: a review of methods and applications. *Crit Rev Oncog.* 2008;14(2-3):89-164.

5. Nielsen B, Albregtsen F, Kildal W, Abeler VM, Kristensen GB, Danielsen HE. The prognostic value of adaptive nuclear texture features from patient grey level entropy matrices in early stage ovarian cancer. *Anal Cell Pathol.* 2012;35(4):305-314.

6. Nielsen B, Hveem TS, Kildal W, et al. Entropy-based adaptive nuclear texture features are independent prognostic markers in a total population of uterine sarcomas. *Cytometry A.* 2015;87(4):315-325.

7. Hveem TS, Njolstad TS, Nielsen B, et al. Changes in chromatin structure in curettage specimens identifies high-risk patients in endometrial cancer. *Cancer Epidemiol Biomarkers Prev.* 2017;26(1):61-67.

8. Dunn JM, Hveem T, Pretorius M, et al. Comparison of nuclear texture analysis and image cytometric DNA analysis for the assessment of dysplasia in Barrett’s esophagus. *Br J Cancer.* 2011;105(8):1218-1223.

9. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybernet.* 1973;3(6):610-621.

10. Nielsen B, Albregtsen F, Danielsen HE. Low dimensional adaptive texture feature vectors from class distance and class difference matrices. *IEEE Trans Med Imaging.* 2004;23(1):73-84.

11. Nielsen B, Danielsen HE. Prognostic value of adaptive textural features—the effect of standardizing nuclear first-order gray level statistics and mixing information from nuclei having different area. *Cell Oncol.* 2006;28(3):85-95.

12. Abeler VM, Røyne O, Thoresen S, Danielsen HE, Nesland JM, Kristensen GB. Uterine sarcomas in Norway. A histopathological and prognostic survey of a total population from 1970 to 2000 including 419 patients. *Histopathology.* 2009;54(3):355-364.

13. Pecorelli S. FIGO Committee on Gynecologic Oncology: Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynecol Obstet.* 2009; (105):103-104.
Supplementary Figure 1. CONSORT diagrams showing the origin of the different cohorts. A) Uterine sarcoma, B) ovarian carcinoma, and C) endometrial carcinoma cohorts.
Supplementary Figure 2. Clustering of the 158,868 nuclei in the discovery set of 175 uterine sarcomas into five subgroups. A) The five clusters of nuclei (seen from different viewpoints, Cluster 1, 25,357 nuclei [16.0%], Cluster 2, 47,851 nuclei [30.1%], Cluster 3, 38,737 nuclei [24.4%], Cluster 4, 29,784 nuclei [18.8%] and Cluster 5, 17,139 nuclei [10.8%]) were obtained by k-means clustering (k=5) of the 3D feature space comprised of center of mass of pixel grey level (CMG), center of mass of pixel grey level entropy (CME) and relative matrix area. B) The average GLEM computed from all nuclei in each of the five clusters. C) Boxplot of the variation/concentration feature relative matrix area in each cluster.
Supplementary Figure 3. A) Receiver operating characteristic (ROC) curve obtained from the uterine sarcoma discovery set (175 patients) by moving the classification threshold for the proportion of high chromatin entropy nuclei (percent nuclei in cluster 5) from 100% to 0%. The dots represent the sensitivity and specificity obtained for different threshold values, 100%, 99%, 98%,…,0%. The biomarker classified each patient as low chromatin entropy or high chromatin entropy based on the proportion of high chromatin entropy nuclei (< classification threshold versus ≥ classification threshold). The red and blue dotted lines highlight the selected threshold of 25%. The black dotted line is the line of no discrimination corresponding to random guessing. B) Correct classification rate (CCR), sensitivity and specificity plotted versus classification threshold on proportion of nuclei with high chromatin entropy. The black dotted line highlights the selected threshold of 25%.
Supplementary Figure 4. Martingale residuals in the null Cox model in univariate A) 5-year overall survival analysis and B) overall survival analysis of the 175 uterine sarcomas in the discovery set plotted versus the proportion of high chromatin entropy nuclei. Martingale residuals in multivariate C) 5-year overall survival analysis and D) overall survival analysis of the established clinicopathological variables (histological subtype, mitotic index, tumor extent, tumor size, tumor margins, cellular atypia, tumor necrosis, hyaline necrosis, vascular invasion [see Table 1], where the marker of high chromatin entropy nuclei is not included in the model) plotted versus the proportion of high chromatin entropy nuclei. Abbreviation: LOWESS, locally weighted scatterplot smoothing.
**Supplementary Figure 5.** Hazard ratios (blue points) and 95% confidence intervals (grey regions) in A) univariate and B) multivariate 5-year overall survival analysis obtained from evaluating the uterine sarcoma discovery set (175 patients) for different classification thresholds for the proportion of high chromatin entropy nuclei (percent nuclei in cluster 5) versus the proportion of high chromatin entropy nuclei. The biomarker classified each patient as low chromatin entropy or high chromatin entropy based on the proportion of high chromatin entropy nuclei (< classification threshold versus ≥ classification threshold). In addition to chromatin entropy, the following variables were included in the multivariate analysis: histological subtype, mitotic index, tumor extent, tumor size, tumor margins, cellular atypia, tumor necrosis, hyaline necrosis, and vascular invasion (see Table 1).
Supplementary Figure 6. Forest plot of the high chromatin entropy marker for the uterine sarcoma patients in analysis of 5-year overall survival. The $P$ values were calculated using the two-sided Mantel-Cox log-rank test. Other sarcomas include ten sarcoma not otherwise specified, four rhabdomyosarcoma, two giant cell tumors with/without leiomyosarcoma and one PEComa (12). Abbreviation: HR, hazard ratio; CI, confidence interval.
Supplementary Figure 7. Kaplan-Meier curves according to the chromatin entropy marker in analysis of overall survival of 354 uterine sarcoma patients. The $P$ values were calculated using the two-sided Mantel-Cox log-rank test. Abbreviation: HR, hazard ratio; LCE, low chromatin entropy; HCE, high chromatin entropy.
Supplementary Figure 8. Forest plot of the high chromatin entropy marker for the ovarian cancer patients in analysis of time to recurrence. High risk ovarian carcinoma was defined as either clear cell histology, poorly differentiated tumor or the combination of moderately differentiated tumor and FIGO, International Federation of Gynecology and Obstetrics, stage Ib or Ic; otherwise, the risk was assessed as low (well differentiated tumor and FIGO stage Ia) or medium (well differentiated tumor and FIGO stage Ib or Ic, or moderately differentiated tumor and FIGO stage Ia). The P values were calculated using the two-sided Mantel-Cox log-rank test. Abbreviation: HR, hazard ratio; CI, confidence interval; n/a, not applicable because all patients with poor prognosis had the same classification.
Supplementary Figure 9. Forest plot of the high chromatin entropy marker for the endometrial cancer patients in analysis of cancer-specific survival. Preoperative curettage histological risk classification was: low risk if benign, hyperplasia, or endometrioid grades 1-2; high risk if non-endometrioid or endometrioid grade 3. FIGO, International Federation of Gynecology and Obstetrics, stage was according to 2009 criteria (13). Histological grade was assessed using hysterectomy specimens. The P values were calculated using the two-sided Mantel-Cox log-rank test. Abbreviation: HR, hazard ratio; CI, confidence interval; n/a, not applicable because all patients with poor prognosis had the same classification.
### Supplementary Table 1. Patient characteristics of the uterine sarcoma cohort.

| Characteristics                                      | All (n=354) | Death ≥ 5y (n=184) | Death < 5y (n=170) |
|------------------------------------------------------|-------------|--------------------|--------------------|
| Median follow-up time; months (IQR)                  | 64.0 (14.8-144.3) | 141.0 (96.3-193.0) | 14.0 (8.0-29.5)    |
| Histological subtype - no. (%)                       |             |                    |                    |
| Leiomyosarcoma                                       | 222         | 99 (44.6)          | 123 (55.4)         |
| Endometrial stromal sarcoma                          | 78          | 58 (75.4)          | 20 (25.6)          |
| Adenosarcomas                                        | 21          | 15 (71.4)          | 6 (28.6)           |
| Undifferentiated uterine sarcoma                     | 16          | 8 (50.0)           | 8 (50.0)           |
| Other sarcomas*                                      | 17          | 4 (23.5)           | 13 (76.5)          |
| Mitotic index, high-power field† - no. (%)           |             |                    |                    |
| ≤10                                                  | 207         | 139 (67.1)         | 68 (32.9)          |
| >10                                                  | 143         | 43 (30.1)          | 100 (69.9)         |
| Tumor extent - no. (%)                               |             |                    |                    |
| Confined to the uterus                               | 267         | 158 (59.2)         | 109 (40.8)         |
| Spread outside the uterus                             | 87          | 26 (29.9)          | 61 (70.1)          |
| Tumor size, cm† - no. (%)                            |             |                    |                    |
| ≤10                                                  | 260         | 156 (60.0)         | 104 (40.0)         |
| >10                                                  | 75          | 23 (30.7)          | 52 (69.3)          |
| Tumor margins† - no. (%)                             |             |                    |                    |
| Pushing                                              | 75          | 45 (60.0)          | 30 (40.0)          |
| Infiltrating                                         | 263         | 129 (49.0)         | 134 (51.0)         |
| Cellular atypia† - no. (%)                           |             |                    |                    |
| Mild                                                 | 106         | 82 (77.4)          | 24 (22.6)          |
| Moderate                                             | 130         | 53 (40.8)          | 77 (59.2)          |
| Severe                                               | 112         | 46 (41.1)          | 66 (58.9)          |
| Tumor necrosis† - no. (%)                            |             |                    |                    |
| Absent                                               | 86          | 62 (72.1)          | 24 (27.9)          |
| Present                                              | 264         | 118 (44.7)         | 146 (55.3)         |
| Hyaline necrosis† - no. (%)                          |             |                    |                    |
| Absent                                               | 174         | 101 (58.0)         | 73 (42.0)          |
| Present                                              | 168         | 76 (45.2)          | 92 (54.8)          |
| Vascular invasion† - no. (%)                         |             |                    |                    |
| Absent                                               | 186         | 104 (55.9)         | 82 (44.1)          |
| Present                                              | 147         | 69 (46.9)          | 78 (53.1)          |
| Chromatin entropy - no. (%)                          |             |                    |                    |
| Low chromatin entropy                                | 293         | 167 (57.0)         | 126 (43.0)         |
| High chromatin entropy                               | 61          | 17 (27.9)          | 44 (72.1)          |

*Other sarcomas include ten sarcoma not otherwise specified, four rhabdomyosarcoma, two giant cell tumors with/without leiomyosarcoma and one PEComa (12).
†Missing data: mitotic index 4, tumor size 19, tumor margins 16, cellular atypia 6, tumor necrosis 4, hyaline necrosis 12, vascular invasion 21.
Abbreviation: IQR, interquartile range.
**Supplementary Table 2.** Patient characteristics of the ovarian cancer cohort.

| Characteristics                        | All (n=246) | No recurrence (n=173) | Recurrence (n=73) |
|----------------------------------------|-------------|-----------------------|-------------------|
| Median follow-up time; years (IQR)     | 11.5 (5.8-13.6) | 12.9 (11.3-14.4) | 1.9 (0.8-4.4) |
| FIGO stage* - no. (%)                  |             |                       |                   |
| Ia                                     | 86          | 72 (83.7)             | 14 (16.3)         |
| Ib                                     | 13          | 6 (46.2)              | 7 (53.8)          |
| Ic                                     | 147         | 95 (64.6)             | 52 (35.4)         |
| Histological grade - no. (%)           |             |                       |                   |
| 1                                      | 106         | 97 (91.5)             | 9 (8.5)           |
| 2                                      | 36          | 27 (75.0)             | 9 (25.0)          |
| 3                                      | 46          | 22 (47.8)             | 24 (52.2)         |
| Not graded (clear cell)                | 58          | 27 (46.6)             | 31 (53.4)         |
| Histological type - no. (%)            |             |                       |                   |
| Mucinous                               | 65          | 60 (92.3)             | 5 (7.7)           |
| Endometrioid                           | 49          | 42 (85.7)             | 7 (14.3)          |
| Mixed                                  | 8           | 6 (75.0)              | 2 (25.0)          |
| Serous                                 | 49          | 30 (61.2)             | 19 (38.8)         |
| Unclassified                           | 15          | 8 (53.3)              | 7 (46.7)          |
| Clear cell                             | 58          | 27 (46.6)             | 31 (53.4)         |
| Small cell                             | 2           | 0 (0.0)               | 2 (100.0)         |
| Dense adhesions† - no. (%)             |             |                       |                   |
| Absent                                 | 157         | 116 (73.9)            | 41 (26.1)         |
| Present                                | 85          | 54 (63.5)             | 31 (36.5)         |
| Rupture† - no. (%)                     |             |                       |                   |
| Absent                                 | 128         | 95 (74.2)             | 33 (25.8)         |
| Present                                | 115         | 75 (65.2)             | 40 (34.8)         |
| Chromatin entropy - no. (%)            |             |                       |                   |
| Low chromatin entropy                  | 210         | 157 (74.8)            | 53 (25.2)         |
| High chromatin entropy                 | 36          | 16 (44.4)             | 20 (55.6)         |

*The FIGO stage was reviewed according to 1988 criteria, although para-aortic and pelvic lymphadenectomy was not routinely performed.
†Missing data: dense adhesions 4, rupture 3.
Abbreviation: IQR, interquartile range; FIGO, International Federation of Gynecology and Obstetrics.
### Supplementary Table 3. Patient characteristics of the endometrial cancer cohort.

| Characteristics                                                                 | All (n=791) | No cancer death (n=707) | Cancer death (n=84) |
|--------------------------------------------------------------------------------|-------------|-------------------------|--------------------|
| **Median follow-up time; months (IQR)**                                         | 35.0 (17.0-53.0) | 37.0 (19.0-55.0) | 14.0 (6.0-28.8) |
| **Preoperative available parameters:**                                          |             |                         |                    |
| Age at diagnosis, years - no. (%)                                               |             |                         |                    |
| <66                                                                               | 385         | 363 (94.3)              | 22 (5.7)           |
| ≥66                                                                               | 406         | 344 (84.7)              | 62 (15.3)          |
| Curettage histology*† - no. (%)                                                 |             |                         |                    |
| Low-risk                                                                          | 610         | 37 (6.1)                | 573 (93.9)         |
| High-risk                                                                         | 175         | 47 (26.9)               | 128 (73.1)         |
| Chromatin entropy - no. (%)                                                      |             |                         |                    |
| Low chromatin entropy                                                            | 726         | 661 (91.0)              | 65 (9.0)           |
| High chromatin entropy                                                           | 65          | 46 (70.8)               | 19 (29.2)          |
| **Postoperative available parameters:**                                          |             |                         |                    |
| FIGO stage‡ - no. (%)                                                            |             |                         |                    |
| I                                                                                 | 617         | 592 (95.9)              | 25 (4.1)           |
| II                                                                                | 55          | 48 (87.3)               | 7 (12.7)           |
| III                                                                               | 90          | 59 (65.6)               | 31 (34.4)          |
| IV                                                                                | 29          | 8 (27.6)                | 21 (72.4)          |
| Histological subtype - no. (%)                                                  |             |                         |                    |
| Endometrioid                                                                      | 665         | 623 (93.7)              | 42 (6.3)           |
| Non-endometrioid                                                                 | 126         | 84 (66.7)               | 42 (33.3)          |
| Histological grade§† - no. (%)                                                  |             |                         |                    |
| 1                                                                                 | 290         | 286 (98.6)              | 4 (1.4)            |
| 2                                                                                 | 265         | 240 (90.6)              | 25 (9.4)           |
| 3                                                                                 | 233         | 178 (76.4)              | 55 (23.6)          |
| Myometrial infiltration† - no. (%)                                               |             |                         |                    |
| <50%                                                                              | 457         | 440 (96.3)              | 17 (3.7)           |
| ≥50%                                                                              | 250         | 208 (83.2)              | 42 (16.8)          |
| Pathologic node (N) stage† - no. (%)                                            |             |                         |                    |
| N0                                                                                | 517         | 494 (95.6)              | 23 (4.4)           |
| N1/2                                                                              | 72          | 50 (69.4)               | 22 (30.6)          |

*Preoperative curettage histological risk classification; low risk if benign, hyperplasia, or endometrioid grades 1-2, high risk if non-endometrioid or endometrioid grade 3.  
†Missing data: curettage histology 6, histological grade 3, myometrial infiltration 84, pathologic node (N) stage 202.  
‡The FIGO stage was reviewed according to 2009 criteria (13).  
§Histological grade from hysterectomy specimens.  
Abbreviation: IQR, interquartile range; FIGO, International Federation of Gynecology and Obstetrics.
**Supplementary Table 4.** Classification results of the chromatin entropy marker.*

| Patient cohort       | CCR (95% CI) | Sens. (95% CI) | Spec. (95% CI) | PPV (95% CI) | NPV (95% CI) |
|----------------------|--------------|----------------|----------------|--------------|--------------|
| Sarcoma, Discovery   | 60.0 (52.3-67.3) | 25.3 (16.4-36.0) | 91.3 (83.6-96.2) | 72.4 (52.8-87.3) | 57.5 (49.1-65.7) |
| Sarcoma, Validation  | 59.2 (51.6-66.5) | 26.4 (17.6-37.0) | 90.2 (82.2-95.4) | 71.9 (53.3-86.3) | 56.5 (48.1-64.6) |
| Sarcoma, Complete    | 59.6 (54.3-64.8) | 25.9 (19.5-33.2) | 90.8 (85.6-94.5) | 72.1 (59.2-82.9) | 57.0 (51.1-62.7) |
| Ovarian cancer       | 72.0 (65.9-77.5) | 27.4 (17.6-39.1) | 90.8 (85.4-94.6) | 55.6 (38.1-72.1) | 74.8 (68.3-80.5) |
| Endometrial cancer   | 86.0 (83.4-88.3) | 22.6 (14.2-33.1) | 93.5 (91.4-95.2) | 29.2 (18.6-41.8) | 91.0 (88.7-93.0) |

*The uterine sarcoma patients who survived for at least five years were defined as good prognosis, whereas the patients who died within five years were defined as poor prognosis. In the ovarian cancer material, good prognosis was defined as no recurrence of disease, whereas poor prognosis was defined as recurrence of disease. In the endometrial cancer material, good prognosis was defined as no endometrial cancer death, whereas poor prognosis was defined as cancer-specific death.

Abbreviation: CCR, correct classification rate; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval.
**Supplementary Table 5.** Proportion of high chromatin entropy nuclei in good and poor prognosis patients.*

| Patient Cohort                | Average (SD), median (IQR) in good prognosis† | Average (SD), median (IQR) in poor prognosis* | P‡   |
|------------------------------|-----------------------------------------------|-----------------------------------------------|------|
| Uterine sarcoma, Discovery set| 9.3 (10.5), 4.9 (2.5-14.2)                     | 16.6 (15.3), 10.5 (5.1-24.5)                  | <0.001 |
| Uterine sarcoma, Validation set| 9.7 (12.5), 5.4 (2.3-9.8)                      | 18.0 (15.5), 14.1 (6.7-26.1)                  | <0.001 |
| Uterine sarcoma, Complete cohort| 9.5 (11.5), 5.3 (2.5-11.8)                  | 17.3 (15.4), 12.5 (5.7-26.0)                  | <0.001 |
| Ovarian cancer                | 9.4 (12.6), 4.1 (2.2-12.2)                     | 18.6 (19.6), 9.9 (5.8-27.4)                   | <0.001 |
| Endometrial cancer            | 8.4 (9.5), 5.3 (2.9-9.5)                       | 15.8 (12.9), 11.6 (5.9-22.2)                  | <0.001 |

*The average, standard deviation (SD), median and interquartile range (IQR) of the proportion of high chromatin entropy nuclei (i.e., nuclei in cluster 5) for a patient computed in different prognostic groups of patients with uterine sarcoma, ovarian cancer and endometrial cancer.

†The uterine sarcoma patients who survived for at least five years were defined as good prognosis, whereas the patients who died within five years were defined as poor prognosis. In the ovarian cancer material, good prognosis was defined as no recurrence of disease, whereas poor prognosis was defined as recurrence of disease. In the endometrial cancer material, good prognosis was defined as no endometrial cancer death, whereas poor prognosis was defined as cancer-specific death.

‡The P values were calculated using the two-sided Mann-Whitney U test.
**Supplementary Table 6.** Multivariate 5-year overall survival analysis of the final model after stepwise Cox regression analysis among 354 uterine sarcomas.*

| FEATURE                        | HR (95% CI)          | P†   |
|--------------------------------|----------------------|------|
| Chromatin entropy              |                      |      |
| Low chromatin entropy          | Ref                  |      |
| High chromatin entropy         | 1.64 (1.14-2.36)     | 0.008|
| Histological subtype           |                      |      |
| Leiomyosarcoma                  | Ref                  |      |
| Endometrial stromal sarcoma    | 0.54 (0.30-0.94)     | 0.003|
| Adenocarcinoma                  | 0.83 (0.33-2.07)     |      |
| Undifferentiated uterine sarcoma| 0.73 (0.34-1.58)    |      |
| Other sarcomas                  | 2.43 (1.36-4.36)     |      |
| Mitotic index, high-power field|                      |      |
| ≤10                            | Ref                  |      |
| >10                            | 2.56 (1.80-3.63)     | <0.001|
| Tumor extent                    |                      |      |
| Confined to the uterus          | Ref                  |      |
| Spread outside the uterus       | 2.48 (1.75-3.51)     | <0.001|
| Tumor size, cm                  |                      |      |
| ≤10                            | Ref                  |      |
| >10                            | 1.81 (1.26-2.60)     | 0.001|

*Starting with the chromatin entropy marker and established clinicopathological variables (see Table 1), backward selection was used in stepwise Cox regression analysis (exclusion criterion P>0.05) and all excluded variables were eventually tested for model inclusion.

†The P values were calculated using the two-sided Wald $\chi^2$ test.

Abbreviation: HR, hazard ratio; CI, confidence interval.
**Supplementary Table 7.** Multivariate overall survival analysis (without truncating at 5 years) of the chromatin entropy marker and established clinicopathological variables among 354 uterine sarcomas.

| FEATURE                                      | HR (95% CI)          | \( \* \) |
|----------------------------------------------|----------------------|---------|
| Chromatin entropy                            |                      | <0.001 |
| Low chromatin entropy                        | Ref                  |         |
| High chromatin entropy                       | 1.98 (1.40-2.81)     | 0.001  |
| Histological subtype                         |                      |         |
| Leiomyosarcoma                               | Ref                  |         |
| Endometrial stromal sarcoma                  | 0.59 (0.34-1.03)     |         |
| Adenosarcoma                                 | 1.53 (0.75-3.12)     |         |
| Undifferentiated uterine sarcoma             | 0.78 (0.38-1.58)     |         |
| Other sarcomas                               | 2.80 (1.57-4.99)     |         |
| Mitotic index, high-power field              |                      | <0.001 |
| \( \leq 10 \)                               | Ref                  |         |
| \( >10 \)                                   | 2.07 (1.52-2.81)     |         |
| Tumor extent                                 |                      | <0.001 |
| Confined to the uterus                       | Ref                  |         |
| Spread outside the uterus                    | 2.25 (1.60-3.17)     |         |
| Tumor size, cm                               |                      | <0.001 |
| \( \leq 10 \)                               | Ref                  |         |
| \( >10 \)                                   | 1.88 (1.34-2.63)     |         |
| Tumor margins                                |                      | 0.38   |
| Pushing                                      | Ref                  |         |
| Infiltrating                                 | 1.17 (0.83-1.65)     |         |
| Cellular atypia                              |                      | 0.01   |
| Mild                                         | Ref                  |         |
| Moderate                                     | 1.73 (1.10-2.70)     |         |
| Severe                                       | 1.19 (0.73-1.94)     |         |
| Tumor necrosis                               |                      | 0.25   |
| Absent                                       | Ref                  |         |
| Present                                      | 1.28 (0.85-1.92)     |         |
| Hyaline necrosis                             |                      | 0.18   |
| Absent                                       | Ref                  |         |
| Present                                      | 1.23 (0.91-1.65)     |         |
| Vascular invasion                            |                      | 0.43   |
| Absent                                       | Ref                  |         |
| Present                                      | 1.13 (0.83-1.54)     |         |

*The \( P \) values were calculated using the two-sided Wald \( \chi^2 \) test.

Abbreviation: HR, hazard ratio; CI, confidence interval.
**Supplementary Table 8.** Multivariate cancer-specific survival analysis of the chromatin entropy marker and established clinicopathological variables among 354 uterine sarcomas.

| FEATURE                        | HR (95% CI) | P*    |
|--------------------------------|-------------|-------|
| Chromatin entropy              |             | 0.001 |
| Low chromatin entropy          | Ref         |       |
| High chromatin entropy         | 1.87 (1.30-2.70) |       |
| Histological subtype           | <0.001      |       |
| Leiomyosarcoma                 | Ref         |       |
| Endometrial stromal sarcoma    | 0.53 (0.29-0.97) |       |
| Adenosarcoma                   | 1.33 (0.61-2.92) |       |
| Undifferentiated uterine sarcoma| 0.69 (0.33-1.45) |       |
| Other sarcomas                 | 2.97 (1.66-5.30) |       |
| Mitotic index, high-power field| <0.001      |       |
| ≤10                            | Ref         |       |
| >10                            | 2.10 (1.52-2.90) |       |
| Tumor extent                   | <0.001      |       |
| Confined to the uterus         | Ref         |       |
| Spread outside the uterus      | 2.57 (1.80-3.65) |       |
| Tumor size, cm                 | 0.002       |       |
| ≤10                            | Ref         |       |
| >10                            | 1.73 (1.21-2.47) |       |
| Tumor margins                  | 0.36        |       |
| Pushing                        | Ref         |       |
| Infiltrating                   | 1.19 (0.82-1.71) |       |
| Cellular atypia                | 0.02        |       |
| Mild                           | Ref         |       |
| Moderate                       | 1.82 (1.11-2.99) |       |
| Severe                         | 1.30 (0.76-2.20) |       |
| Tumor necrosis                 | 0.18        |       |
| Absent                         | Ref         |       |
| Present                        | 1.35 (0.87-2.09) |       |
| Hyaline necrosis               | 0.78        |       |
| Absent                         | Ref         |       |
| Present                        | 1.05 (0.76-1.43) |       |
| Vascular invasion              | 0.33        |       |
| Absent                         | Ref         |       |
| Present                        | 1.18 (0.85-1.62) |       |

*The P values were calculated using the two-sided Wald χ² test. Abbreviation: HR, hazard ratio; CI, confidence interval.
Supplementary Table 9. Multivariate cancer-specific survival analysis of the chromatin entropy marker and established variables among 246 stage I ovarian cancer patients.

| FEATURE            | HR (95% CI)          | P*  |
|---------------------|----------------------|-----|
| Chromatin entropy   |                      | 0.02|
| Low chromatin entropy | Ref                  |     |
| High chromatin entropy | 1.91 (1.12-3.27)    |     |
| FIGO stage†         |                      | 0.01|
| Ia                  | Ref                  |     |
| lb-c                | 2.27 (1.21-4.26)     |     |
| Histological grade  |                      | <0.001|
| 1-2                 | Ref                  |     |
| 3 or not graded (clear cell) | 5.61 (3.06-10.27) |     |

*The P values were calculated using the two-sided Wald $\chi^2$ test.
†The FIGO stage was reviewed according to 1988 criteria, although para-aortic and pelvic lymphadenectomy was not routinely performed.
Abbreviation: HR, hazard ratio; CI, confidence interval; FIGO, International Federation of Gynecology and Obstetrics.