THE LANCET
Oncology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Kleppe A, Albregtsen F, Vlatkovic L, et al. Chromatin organisation and cancer prognosis: a pan-cancer study. *Lancet Oncol* 2017; published online Feb 2. http://dx.doi.org/10.1016/S1470-2045(17)30899-9.

Appendix

Chromatin organisation and cancer prognosis: a pan-cancer study

Andreas Kleppe, Ph.D., Prof. Fritz Albregtsen, Cand.Real., Ljiljana Vlatkovic, M.D., Manohar Pradhan, Ph.D., Birgitte Nielsen, Ph.D., Tarjei S. Hveem, Ph.D., Hanne A. Askautrud, Ph.D., Gunnar B. Kristensen, Ph.D., Prof. Arild Nesbakken, Ph.D., Prof. Jone Trovik, Ph.D., Håkon Wæhre, M.D., Prof. Ian Tomlinson, Ph.D., Prof. Neil A. Shepherd, D.M., Prof. Marco Novelli, Ph.D., Prof. David J. Kerr, M.D., and Prof. Håvard E. Danielsen, Ph.D.

Table of contents:

Search criteria used in “Evidence before this study”

In the “Research in context” panel, the PubMed search outlined under “Evidence before this study” was:

(nuclear[All Fields] OR ("chromatin"[MeSH Terms] OR "chromatin"[All Fields]) OR ("cells"[MeSH Terms] OR "cells"[All Fields] OR "cell"[All Fields])) AND (texture[All Fields] OR (("chromatin"[MeSH Terms] OR "chromatin"[All Fields]) AND ("organisation"[All Fields] OR "organization"[All Fields]) AND "changes"[All Fields])) AND (("analysis"[Subheading] OR "analysis"[All Fields]) AND ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]))

Nuclear texture analysis

The GLEM4D representation of a patient

The GLEM of a nuclear image is a bivariate probability mass function describing the relation between pixel grey level and how disordered the surrounding chromatin organisation appears, measured by the entropy of the grey levels in a square subregion. By changing the size of the subregion, the chromatin organisation can be analysed at different scales. With small subregions an organisational disorder between small chromatin compartments could be detected more accurately, while large subregions facilitate evaluation of large chromatin compartments and the overall chromatin disorder in the nucleus. To simultaneously allow examinations at multiple scales, we computed the GLEM for subregions of 3x3, 5x5 and every odd number up to 31x31 pixels. The resulting GLEMs were dividing by the number of scales and concatenated to form a three-way table, here termed GLEM3D. The GLEM3D is thus a trivariate probability mass function with size of subregion as the third axis, and grey level and entropy as the first two axes (as in the GLEM).

We could characterise the patient by the element-wise mean GLEM3D of its nuclei, but measurements of disordered chromatin organisation and DNA content (i.e. the entropy and the grey level) have been shown to greatly depend on the nuclear size.[1] The nuclei were therefore stratified into 11 groups according to their size; the first group consisted of nuclei with 1 to 999 pixels, the second with 1000 to 1999 pixels, and so on up to the group of nuclei with 10000 pixels or more. This stratification was implemented by including the nuclear size group as a fourth axis, giving a four-way table known as the GLEM4D.[2] Each patient was represented by the element-wise mean GLEM4D of its nuclei.

A patient's GLEM4D represents the associations between disordered chromatin organisation and DNA density found in its nuclei when differences in size of chromatin compartments and nuclei are accounted for. Each possible chromatin pattern will be associated with a specific GLEM4D element. Considering a specific element in the GLEM4D representation of a patient, its value will describe the relative frequency (i.e. probability) in which the corresponding chromatin pattern occurs in the nuclei of that patient. Computing the GLEM4D for a patient is thus a way of embodying information about the observed chromatin patterns.

Discovery algorithm

The algorithm used to create the Nucleotyping classifier constitutes of the following steps, all of which were applied solely to the discovery cohort:

1. The GLEM4D for each patient in the discovery cohort was computed after uniform re-quantification of the nuclear images to 64 grey levels (original pixel depth was 10 bits which gives 1024 grey levels). The entropy value was computed using the natural logarithm and uniformly quantified using 12·5 levels per integer entropy. The entropy axis was not limited, but technically set by the theoretical maximum entropy value for the given number of grey levels (64), i.e. log(64)≈4·16, which was multiplied by the number of levels per integer entropy (12·5) and rounded down to the nearest integer to obtain the maximum index of the entropy axis when given in a zero-indexed format. The GLEM4D was thus a 64x52x15x11 matrix, i.e. it had 64 elements along the grey level axis, 52 along the entropy axis, 15 along the subregion size axis and 11 along the nuclear size axis.

2. The patients in the discovery cohort were divided in two prognosis groups by separating those who suffered from those who did not suffer a recurrence of their colorectal cancer.

3. In each GLEM4D element, the statistical discrepancy between the two prognosis groups was estimated as:

$$t = \frac{m(\omega_g) - m(\omega_p)}{s}$$

$m(\omega_g)$ and $m(\omega_g)$ denotes the mean value in a specific GLEM4D element among discovery patients with good and poor prognosis, respectively. Defining $s^2(\omega_g)$ and $s^2(\omega_g)$ as the corresponding unbiased

variance estimates (also in the same GLEM4D element), the pooled standard deviation was estimated as:

$$s = \sqrt{\frac{s^2(\omega_g) + s^2(\omega_g)}{2}}$$

The applied measure of statistical discrepancy is thus $\sqrt{2/n}$ times the *t*-statistic in the two-sample *t*-test which assumes equal variances and sample sizes (denoted n), under the null hypothesis that the expected GLEM4D element value in the two prognosis groups was equal (the null hypothesis was thus that we expect the chromatin pattern to be equally frequent in both prognosis groups). The constant scaling of $\sqrt{2/n}$ affects the range of all quantities (including the chromatin value), but is irrelevant for the final assessment of chromatin heterogeneity. The magnitude of *t* reflects the associated GLEM4D element's ability to discriminate between poor and good prognosis and the sign designates which prognosis class the element indicated; GLEM4D elements of aberrant chromatin patterns received a negative value and GLEM4D elements of common chromatin patterns received a positive value.

4. For each patient, the weighted sum of the GLEM4D elements was calculated using the just-described measure of statistical discrepancy as weights. If disregarding that averaging were performed on two levels (first in computing the GLEM4D for each nucleus from the observed chromatin patterns on different scales and then in computing the GLEM4D for the patient from the GLEM4Ds of its nuclei), this would amount to simply summing the contribution of each observed chromatin pattern, i.e. summing the estimated influence of each observed chromatin pattern on determining chromatin heterogeneity. The calculated weighted sum is a single continuous value termed the chromatin value for the patient and corresponds to the difference between the positive and negative adaptive feature described by Nielsen et al.[3]

5. The minimum Euclidean distance classification method, recognised to be robust, simple and accurate,[4,5] was utilised to dichotomise the chromatin value. The method, which is also known as *nearest centroid classifier* and *nearest prototype classifier*, computes a threshold for the chromatin value as the value with equal distance to the mean chromatin value of discovery cohort patients with either good or poor prognosis. Specifically, the method first calculates the mean chromatin value for the patients in the discovery cohort who suffered colorectal cancer recurrence, and similarly for patients who did not suffer a recurrence, and then defines the average of these two means as the threshold for dichotomisation. Using the discovery cohort, the threshold for the chromatin value was computed as 0.044. If the chromatin value of a new patient was below the 0.044, the patient was classified as chromatin heterogeneous (CHE), otherwise the chromatin value was above the threshold and the patient was classified as chromatin homogeneous (CHO).

Classifying a new patient

To assess the chromatin heterogeneity of a new patient, e.g. a patient in one of the six validation cohorts, the following steps were performed:

1. The GLEM4D representation of the patient was computed (the technical details are the same as described in step 1 of the discovery algorithm).

2. The chromatin value for the patient was calculated (see step 4 of the discovery algorithm). Note that the applied estimates only depend on the discovery cohort, not on the new patient.

3. The chromatin value was dichotomised using the threshold computed in step 5 in the discovery algorithm, giving the chromatin heterogeneity classification, either CHO or CHE. Again we could point out that the threshold only depends on the discovery cohort, not on the new patient.

Appendix references

1. 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.
2. Nielsen B, Albregtsen F, Kildal W, Abeler VM, Kristensen GB, Danielsen HE. The prognostic value of adaptive nuclear texture features from patient gray level entropy matrices in early stage ovarian cancer. *Anal Cell Pathol (Amst)* 2012; **35**(4): 305-14.
3. 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.
4. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans Pattern Anal Mach Intell* 1991; **13**(3): 252-64.
5. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005; **365**(9458): 488-92.

Follow-up time (years)	11·7 (8·3-13·6)
Age at surgery (years)	57 (49-66)
≤60	141 (57%)
>60	105 (43%)
FIGO stage	
Ia	86 (35%)
Ib	13 (5%)
Ic	147 (60%)
Histological grade	
1	106 (43%)
2	36 (15%)
3	46 (19%)
Not graded (clear cell)	58 (24%)
Histological subtype	
Mucinous	65 (26%)
Endometrioid	49 (20%)
Serous	49 (20%)
Clear cell	58 (24%)
Small cell	2 (1%)
Mixed	8 (3%)
Unclassifiable	15 (6%)
Dense adhesions	
Absent	157 (65%)
Present	85 (35%)
Rupture	
Absent	128 (53%)
Present	115 (47%)
Chromatin heterogeneity	
Homogeneous (CHO)	169 (69%)
Heterogeneous (CHE)	77 (31%)

Data are median (interquartile range [IQR]) or number (%).
FIGO=International Federation of Gynecology and Obstetrics.

Table A1: **Characteristics of the 246 ovarian carcinoma patients**

Follow-up time (years)	5·0 (1·3-5·0)
Age at surgery (years)	56 (48-65)
≤60	224 (63%)
>60	130 (37%)
Histological subtype	
Leiomyosarcoma	222 (63%)
Endometrial stromal sarcoma	78 (22%)
Adenosarcoma	21 (6%)
Undifferentiated uterine sarcomas	16 (5%)
Other sarcomas	17 (5%)
Mitotic index (mitoses per high-power field)	
≤10	207 (59%)
>10	143 (41%)
Tumour extent	
Confined to the uterus	267 (75%)
Spread outside the uterus	87 (25%)
Tumour size (cm)	
≤10	260 (78%)
>10	75 (22%)
Tumour margins	
Pushing	75 (22%)
Infiltrating	263 (78%)
Cellular atypia	
Mild	106 (30%)
Moderate	130 (37%)
Severe	112 (32%)
Tumour necrosis	
Absent	86 (25%)
Present	264 (75%)
Hyaline necrosis	
Absent	174 (51%)
Present	168 (49%)
Vascular invasion	
Absent	186 (56%)
Present	147 (44%)
Chromatin heterogeneity	
Homogeneous (CHO)	201 (57%)
Heterogeneous (CHE)	153 (43%)

Data are median (interquartile range [IQR]) or number (%).

Table A2: **Characteristics of the 354 uterine sarcoma patients**

Follow-up time (years)	10·3 (7·2-14·0)
Age at surgery (years)	63 (58-67)
≤65	192 (63%)
>65	115 (37%)
Preoperative PSA (ng/ml)	
≤6	79 (26%)
>6 to ≤10	62 (20%)
>10 to ≤20	93 (31%)
>20	70 (23%)
Gleason grade	
≤6	18 (6%)
3+4	118 (38%)
4+3	88 (29%)
≥8	83 (27%)
Surgical margins	
Negative	119 (39%)
Positive	188 (61%)
Extracapsular extension	
Absent	78 (26%)
Present	226 (74%)
Seminal vesicle invasion	
Absent	238 (78%)
Present	69 (22%)
Pathologic node (N) stage	
N0/x	292 (95%)
N1	15 (5%)
CAPRA-S risk group*	
Low	46 (15%)
Intermediate	108 (36%)
High	147 (49%)
Chromatin heterogeneity	
Homogeneous (CHO)	252 (82%)
Heterogeneous (CHE)	55 (18%)

Data are median (interquartile range [IQR]) or number (%). CAPRA-S=Cancer of the Prostate Risk Assessment Postsurgical. PSA=prostate-specific antigen. *The CAPRA-S score was categorised to give three CAPRA-S risk groups: Low risk if score 0 to 2; Intermediate risk if score 3 to 5; High risk if score 6 to 12.

Table A3: **Characteristics of the 307 prostate carcinoma patients**

Follow-up time (years)	3·0 (1·5-4·4)
Age at surgery (years)	66 (59-74)
≤70	507 (64%)
>70	284 (36%)
Curettage histology classification*	
Low risk	610 (78%)
High risk	175 (22%)
FIGO stage	
I	617 (78%)
II	55 (7%)
III	90 (11%)
IV	29 (4%)
Histological grade	
1	290 (37%)
2	265 (34%)
3	233 (30%)
Histological subtype	
Endometrioid carcinoma	665 (84%)
Serous carcinoma	53 (7%)
Clear cell carcinoma	30 (4%)
Carcinosarcoma	29 (4%)
Other	14 (2%)
Myometrial invasion	
<50%	457 (65%)
≥50%	250 (35%)
Pathologic node (N) stage	
N0	517 (88%)
N1/2	72 (12%)
Chromatin heterogeneity	
Homogeneous (CHO)	673 (85%)
Heterogeneous (CHE)	118 (15%)

Data are median (interquartile range [IQR]) or number (%). FIGO=International Federation of Gynecology and Obstetrics. *Curettage histology classification was: Low risk if benign, hyperplasia or endometrioid grade 1 or 2; High risk if non-endometrioid or endometrioid grade 3.

Table A4: **Characteristics of the 791 endometrial carcinoma patients**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	1·7	1·2-2·5	0·0056	1·7	1·1-2·5	0·0096
Age>72 *vs* ≤72 years	1·9	1·3-2·8	0·0010	1·8	1·2-2·7	0·0029
Stage II *vs* I	2·3	1·4-3·8	0·0008	2·0	1·2-3·3	0·011
Histological grade			0·49			0·62
2 *vs* 1	1·48	0·69-3·20	0·31	1·26	0·57-2·76	0·57
3 *vs* 1	1·77	0·69-4·57	0·24	1·61	0·61-4·28	0·34
Acute* *vs* elective surgery	2·6	1·5-4·3	0·0002	2·1	1·2-3·6	0·0065

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. *Acute surgery was performed due to obstruction or perforation of the bowel at presentation.

Table A5: **Analysis of cancer-specific survival in the colorectal cancer discovery cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	1·8	1·0-3·0	0·033	1·9	1·1-3·2	0·026
Age>72 *vs* ≤72 years	0·98	0·57-1·68	0·95	0·88	0·51-1·52	0·64
Stage II *vs* I	14·0	1·9-101·3	0·0006	8·8	1·2-65·5	0·034
Histological grade			0·0008			0·020
2 *vs* 1	2·7	1·1-6·4	0·027	1·86	0·77-4·51	0·17
3 *vs* 1	5·2	2·0-13·4	0·0005	3·55	1·36-9·29	0·010
Acute* *vs* elective surgery	2·8	1·6-4·9	0·0003	2·0	1·1-3·6	0·021

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. *Acute surgery was defined as either urgent or emergency surgery.

Table A6: **Analysis of cancer-specific survival in the Gloucester validation cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	2·2	1·1-4·5	0·027	2·6	1·2-5·6	0·016
Age>72 *vs* ≤72 years	1·13	0·43-2·94	0·81	1·46	0·54-3·94	0·45
Histological grade			0·91			0·90
2 *vs* 1	1·26	0·17-9·32	0·82	1·23	0·16-9·17	0·84
3 *vs* 1	1·02	0·11-9·13	0·99	0·98	0·11-8·94	0·98

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio.

Table A7: **Analysis of cancer-specific survival in the QUASAR 2 validation cohort**

Characteristic	CHO (n=787)	CHE (n=436)	Spearman's correlation ρ (95% CI)	p
Follow-up time (years)	4·8 (3·0-6·3)	4·8 (2·8-6·8)	0·01 (-0·05 to 0·06)	0·81
Age at surgery (years)	70 (61-77)	68 (61-75)	-0·05 (-0·11 to 0·01)	0·078
≤72	463 (59%)	290 (67%)	-0·08 (-0·13 to -0·02)	0·0081
>72	324 (41%)	146 (33%)		
Gender			-0·04 (-0·10 to 0·01)	0·13
Female	371 (47%)	186 (43%)		
Male	416 (53%)	250 (57%)		
Stage			-0·00 (-0·06 to 0·05)	0·94
I	125 (16%)	70 (16%)		
II	662 (84%)	366 (84%)		
Histological grade			-0·07 (-0·12 to -0·01)	0·022
1	110 (14%)	62 (15%)		
2	532 (69%)	322 (76%)		
3	127 (17%)	40 (9%)		
Pathologic tumour (T) stage			-0·02 (-0·08 to 0·04)	0·49
T1	26 (3%)	11 (3%)		
T2	98 (13%)	59 (14%)		
T3	437 (56%)	250 (58%)		
T4	216 (28%)	109 (25%)		
Microsatellite stability*			0·23 (0·16 to 0·29)	<0·0001
Unstable (MSI)	106 (24%)	19 (7%)		
Stable (MSS)	336 (76%)	270 (93%)		
Location			0·09 (0·04 to 0·15)	0·0010
Rectum	184 (24%)	109 (25%)		
Distal colon	241 (31%)	179 (42%)		
Proximal colon	353 (45%)	140 (33%)		
Surgery type†			0·07 (0·01 to 0·14)	0·033
Elective	477 (89%)	243 (84%)		
Acute	56 (11%)	45 (16%)		

Data are median (interquartile range [IQR]) or number (%), unless otherwise indicated. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. *Microsatellite stability data was not available for the Gloucester validation cohort. †Surgery type data was not available for the QUASAR 2 validation cohort.

Table A8: **Association between chromatin heterogeneity and other patient characteristics in the three colorectal carcinoma cohorts**

Patient cohort	Subgroup	All	CHO	CHE	HR (95% CI)*
Discovery					
	All		24% (40 of 164)	44% (50 of 114)	2·0 (1·3-3·0)
	MSI	26% (14 of 53)	23% (10 of 44)	44% (4 of 9)	2·9 (0·9-9·4)
	MSS	36% (73 of 205)	27% (28 of 105)	45% (45 of 100)	1·8 (1·1-2·8)
	HR (95% CI)†	1·5 (0·8-2·6)	1·3 (0·7-2·8)	0·8 (0·3-2·2)	
QUASAR 2 validation					
	All		5% (13 of 244)	12% (17 of 147)	2·2 (1·1-4·5)
	MSI	5% (3 of 62)	4% (2 of 54)	13% (1 of 8)	3·4 (0·3-37·7)
	MSS	9% (27 of 306)	6% (11 of 170)	12% (16 of 136)	1·9 (0·9-4·0)
	HR (95% CI)†	1·8 (0·6-6·1)	1·7 (0·4-7·8)	1·0 (0·1-7·9)	
Combined					
	All		13% (53 of 408)	26% (67 of 261)	2·1 (1·4-2·9)
	MSI	15% (17 of 115)	12% (12 of 98)	29% (5 of 17)	2·9 (1·0-8·4)
	MSS	20% (100 of 511)	14% (39 of 275)	26% (61 of 236)	1·8 (1·2-2·7)
	HR (95% CI)†	1·4 (0·9-2·4)	1·3 (0·7-2·4)	0·8 (0·3-2·0)	

Data are % (number of cancer-specific deaths of number of patients in subgroup), unless otherwise indicated. Microsatellite stability data was not available for the Gloucester validation cohort. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. MSI=microsatellite unstable. MSS=microsatellite stable. *Survival analysis of CHE *vs* CHO in the specified subgroup. †Survival analysis of MSS *vs* MSI in the specified subgroup.

Table A9: **Cancer-specific mortality rates and survival analyses in subgroups of microsatellite stability and Nucleotyping for stage II colorectal cancer**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	3·1	1·9-5·0	<0·0001	1·8	1·1-3·0	0·022
FIGO stage Ib or Ic *vs* Ia	2·8	1·5-5·2	0·0008	2·3	1·2-4·2	0·011
Grade 3 or clear cell *vs* grade 1 or 2	6·7	3·7-12·2	<0·0001	5·4	2·9-9·9	<0·0001

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. FIGO=International Federation of Gynecology and Obstetrics. HR=hazard ratio.

Table A10: **Analysis of cancer-specific survival in the ovarian carcinoma cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	2·5	1·8-3·4	<0·0001	1·6	1·0-2·4	0·038
Histological subtype			<0·0001			0·021
Endometrial stromal sarcoma *vs* LMS	0·35	0·21-0·57	<0·0001	0·63	0·29-1·37	0·25
Adenosarcoma *vs* LMS	0·38	0·15-0·92	0·033	1·08	0·37-3·16	0·88
Undifferentiated uterine sarcoma *vs* LMS	0·96	0·47-1·96	0·90	0·75	0·31-1·84	0·53
Other sarcomas *vs* LMS	2·26	1·27-4·01	0·0056	2·79	1·44-5·39	0·0023
Mitotic index (>10 *vs* ≤10 mitoses per high-power field)	2·9	2·1-3·9	<0·0001	2·4	1·6-3·5	<0·0001
Tumour spread outside *vs* confined to the uterus	2·6	1·9-3·5	<0·0001	2·6	1·7-3·9	<0·0001
Tumour size (>10 cm *vs* ≤10)	2·2	1·6-3·1	<0·0001	1·7	1·1-2·5	0·010
Tumour margins (infiltrating *vs* pushing)	1·46	0·98-2·19	0·062	1·21	0·77-1·90	0·41
Cellular atypia			<0·0001			0·34
Moderate *vs* mild	3·9	2·4-6·3	<0·0001	1·46	0·75-2·82	0·26
Severe *vs* mild	3·7	2·3-6·1	<0·0001	1·16	0·57-2·36	0·68
Tumour necrosis (present *vs* absent)	2·6	1·6-4·0	<0·0001	1·62	0·94-2·79	0·082
Hyaline necrosis (present *vs* absent)	1·32	0·97-1·82	0·075	0·99	0·69-1·44	0·97
Vascular invasion (present *vs* absent)	1·34	0·97-1·84	0·070	1·39	0·96-2·01	0·083

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. LMS=leiomyosarcoma.

Table A11: **Analysis of 5-year cancer-specific survival in the uterine sarcoma cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	2·3	1·2-4·6	0·012	1·43	0·68-2·99	0·34
Age (years)*	1·01	0·96-1·07	0·69	0·98	0·92-1·05	0·62
Preoperative PSA (ng/ml)			0·0019			0·25
>6 to ≤10 *vs* ≤6	1·64	0·33-8·15	0·54	1·46	0·27-7·90	0·66
>10 to ≤20 *vs* ≤6	1·83	0·48-6·91	0·37	0·94	0·24-3·78	0·94
>20 *vs* ≤6	5·43	1·60-18·48	0·0067	2·15	0·59-7·90	0·25
Gleason grade			<0·0001			0·0073
≤6 *vs* 3+4		n/a			n/a	
4+3 *vs* 3+4	2·64	0·68-10·21	0·16	0·81	0·16-4·18	0·80
≥8 *vs* 3+4	10·82	3·26-35·89	0·0001	3·90	0·87-17·41	0·075
Positive *vs* negative surgical margins	2·3	1·0-5·4	0·038	1·38	0·47-4·09	0·56
Extracapsular extension (present *vs* absent)	4·7	1·1-19·5	0·020	0·79	0·15-4·28	0·79
Seminal vesicle invasion (present *vs* absent)	5·5	2·8-10·8	<0·0001	3·9	1·7-9·0	0·0015
Pathologic node stage N1 *vs* N0/x	3·5	1·4-9·0	0·0059	1·11	0·39-3·19	0·84

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. n/a=not applicable because none of the patients died of prostate cancer. PSA=prostate-specific antigen. *Continuous variable.

Table A12: **Analysis of cancer-specific survival in the prostate carcinoma cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	4·3	2·8-6·8	<0·0001	1·9	1·1-3·1	0·013
Age (years)*	1·06	1·04-1·09	<0·0001	1·05	1·03-1·07	<0·0001
Histology classification (high *vs* low risk)†	5·6	3·6-8·7	<0·0001	3·6	2·2-6·0	<0·0001

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. *Continuous variable. †Curettage histology classification was: Low risk if benign, hyperplasia or endometrioid grade 1 or 2; High risk if non-endometrioid or endometrioid grade 3.

Table A13: **Analysis of cancer-specific survival in the endometrial cancer cohort**

Variable	Univariable analysis			Multivariable analysis		
	HR	95% CI	p	HR	95% CI	p
Chromatin heterogeneity (CHE *vs* CHO)	10·9	4·8-24·8	<0·0001	4·6	1·8-11·3	0·0010
Age (years)*	1·08	1·04-1·13	0·0001	1·06	1·02-1·10	0·0048
Histology classification (high *vs* low risk)†	10·2	4·5-23·2	<0·0001	4·5	1·8-11·3	0·0011

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio. *Continuous variable. †Curettage histology classification was: Low risk if benign, hyperplasia or endometrioid grade 1 or 2; High risk if non-endometrioid or endometrioid grade 3.

Table A14: **Analysis of cancer-specific survival in endometrial cancer patients without adjuvant treatment**

Patient cohort	Time to censoring (years)	Sens. (95% CI)	Spec. (95% CI)	PPV (95% CI)	NPV (95% CI)	CCR (95% CI)
CRC, discovery	7·9 (6·2-11·3)	52% (42%-62%)	65% (59%-70%)	36% (29%-44%)	78% (72%-83%)	61% (56%-66%)
CRC, Gloucester validation	3·8 (2·0-5·6)	42% (29%-56%)	71% (67%-76%)	17% (11%-25%)	90% (86%-93%)	68% (63%-72%)
CRC, QUASAR 2 validation	4·9 (4·0-5·1)	57% (37%-75%)	64% (59%-69%)	12% (7%-18%)	95% (91%-97%)	63% (58%-68%)
Ovarian carcinoma	12·9 (11·2-14·3)	55% (42%-67%)	77% (70%-83%)	47% (35%-58%)	82% (76%-88%)	71% (65%-77%)
Uterine sarcoma	5·0 (5·0-5·0)	60% (52%-67%)	71% (64%-77%)	63% (55%-71%)	68% (61%-74%)	66% (61%-71%)
Prostate carcinoma	10·8 (7·4-14·3)	40% (24%-58%)	85% (80%-89%)	25% (15%-39%)	92% (88%-95%)	80% (75%-84%)
Endometrial carcinoma	3·1 (1·6-4·7)	38% (28%-49%)	88% (85%-90%)	27% (19%-36%)	92% (90%-94%)	83% (80%-85%)

Data are median (interquartile range [IQR]) or % (95% CI). CCR=correct classification rate. CI=confidence interval. CRC=colorectal cancer. NPV=negative predictive value. PPV=positive predictive value.

Table A15: **The ability of Nucleotyping to assess cancer-specific survival outcome**

Patient cohort	CHO	CHE	ρ or HR (95% CI)	p
CRC, discovery				
Spearman's correlation coefficient			-0·06 (-0·16 to 0·04)	0·25
Grade 1	25 (11%)	12 (8%)		
Grade 2	179 (77%)	136 (88%)		
Grade 3	28 (12%)	6 (4%)		
CHE in multivariable analysis with grade			1·7 (1·1-2·5)	0·0082
CRC, Gloucester validation				
Spearman's correlation coefficient			-0·11 (-0·20 to -0·02)	0·021
Grade 1	76 (25%)	44 (33%)		
Grade 2	180 (58%)	77 (57%)		
Grade 3	52 (17%)	13 (10%)		
CHE in multivariable analysis with grade			2·1 (1·2-3·6)	0·0074
CRC, QUASAR 2 validation				
Spearman's correlation coefficient			-0·06 (-0·16 to 0·04)	0·25
Grade 1	9 (4%)	6 (4%)		
Grade 2	173 (76%)	109 (80%)		
Grade 3	47 (21%)	21 (15%)		
CHE in multivariable analysis with grade			2·5 (1·2-5·4)	0·019
Ovarian carcinoma				
Spearman's correlation coefficient			0·35 (0·23 to 0·45)	<0·0001
Grade 1	91 (54%)	15 (19%)		
Grade 2	25 (15%)	11 (14%)		
Grade 3	25 (15%)	21 (27%)		
Not graded (clear cell)	28 (17%)	30 (39%)		
CHE in multivariable analysis with grade			1·8 (1·1-3·0)	0·021
Uterine sarcoma*				
Spearman's correlation coefficient			0·46 (0·37 to 0·54)	<0·0001
Mild	93 (47%)	13 (9%)		
Moderate	70 (36%)	60 (40%)		
Severe	34 (17%)	78 (52%)		
CHE in multivariable analysis with grade			1·9 (1·4-2·7)	0·0002
Prostate carcinoma†				
Spearman's correlation coefficient			0·27 (0·16 to 0·37)	<0·0001
≤ 6	18 (7%)			
3+4	107 (42%)	11 (20%)		
4+3	71 (28%)	17 (31%)		
≥ 8	56 (22%)	27 (49%)		
CHE in multivariable analysis with grade			1·43 (0·72-2·86)	0·31
Endometrial carcinoma				
Spearman's correlation coefficient			0·31 (0·24 to 0·37)	<0·0001
Grade 1	275 (41%)	15 (13%)		
Grade 2	240 (36%)	25 (21%)		
Grade 3	155 (23%)	78 (66%)		
CHE in multivariable analysis with grade			2·4 (1·5-3·9)	0·0004

Data are number (%), unless otherwise indicated. Spearman's correlation coefficient is denoted as ρ. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. CRC=colorectal cancer. HR=hazard ratio. *Cellular atypia is analysed instead of histological grade. †Gleason grade is analysed instead of histological grade.

Table A16: **Association between chromatin heterogeneity and histological grade, and the ability of chromatin heterogeneity to predict cancer-specific survival in multivariable analysis with histological grade**

Patient cohort	Non-diploid CHO *vs* diploid CHO			CHE *vs* diploid CHO		
	HR	95% CI	p	HR	95% CI	p
CRC, discovery	1·7	1·0-3·0	0·046	2·2	1·4-3·6	0·0012
CRC, Gloucester validation	1·32	0·66-2·65	0·43	2·0	1·1-3·8	0·031
CRC, QUASAR 2 validation	3·9	1·1-14·0	0·040	5·1	1·5-17·5	0·0090
Ovarian carcinoma	12·4	4·3-35·7	<0·0001	15·2	5·4-42·7	<0·0001
Uterine sarcoma	3·3	2·0-5·6	<0·0001	4·8	3·0-7·8	<0·0001
Prostate carcinoma	1·78	0·75-4·20	0·19	2·9	1·3-6·5	0·0078
Endometrial carcinoma	1·30	0·61-2·80	0·50	4·6	2·9-7·4	<0·0001

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CI=confidence interval. HR=hazard ratio.

Table A17: **Analysis of cancer-specific survival by chromatin heterogeneity and DNA ploidy**

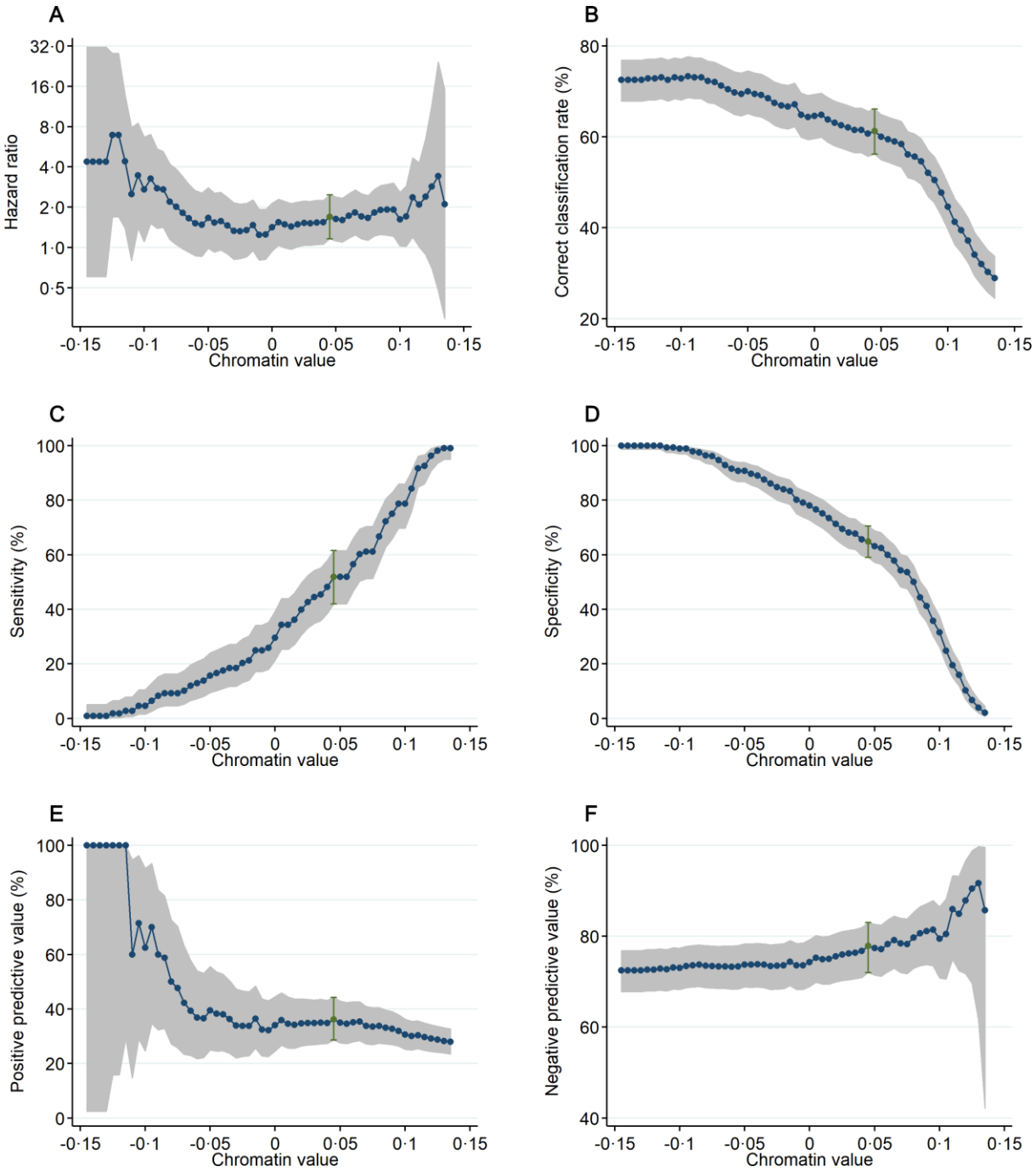

Figure A1: **Ability of markers created by thresholding the chromatin value to predict cancer-specific survival in the colorectal cancer discovery cohort**

Dichotomous markers were created by thresholding the chromatin value at -0.145, 0.14, …, 0.135, which was the range of chromatin values in the discovery cohort where each prognostic group contained at least one event. The estimated accuracy of each marker is shown as a blue point and the corresponding 95% confidence interval as a grey area. The performance estimates of the marker obtained by using the threshold 0.045 is highlighted in green to indicate that this corresponds to the chromatin heterogeneity marker (none of the discovery patients had a chromatin value between 0.045 and the threshold computed using the minimum Euclidean distance classification method, which was 0.044).

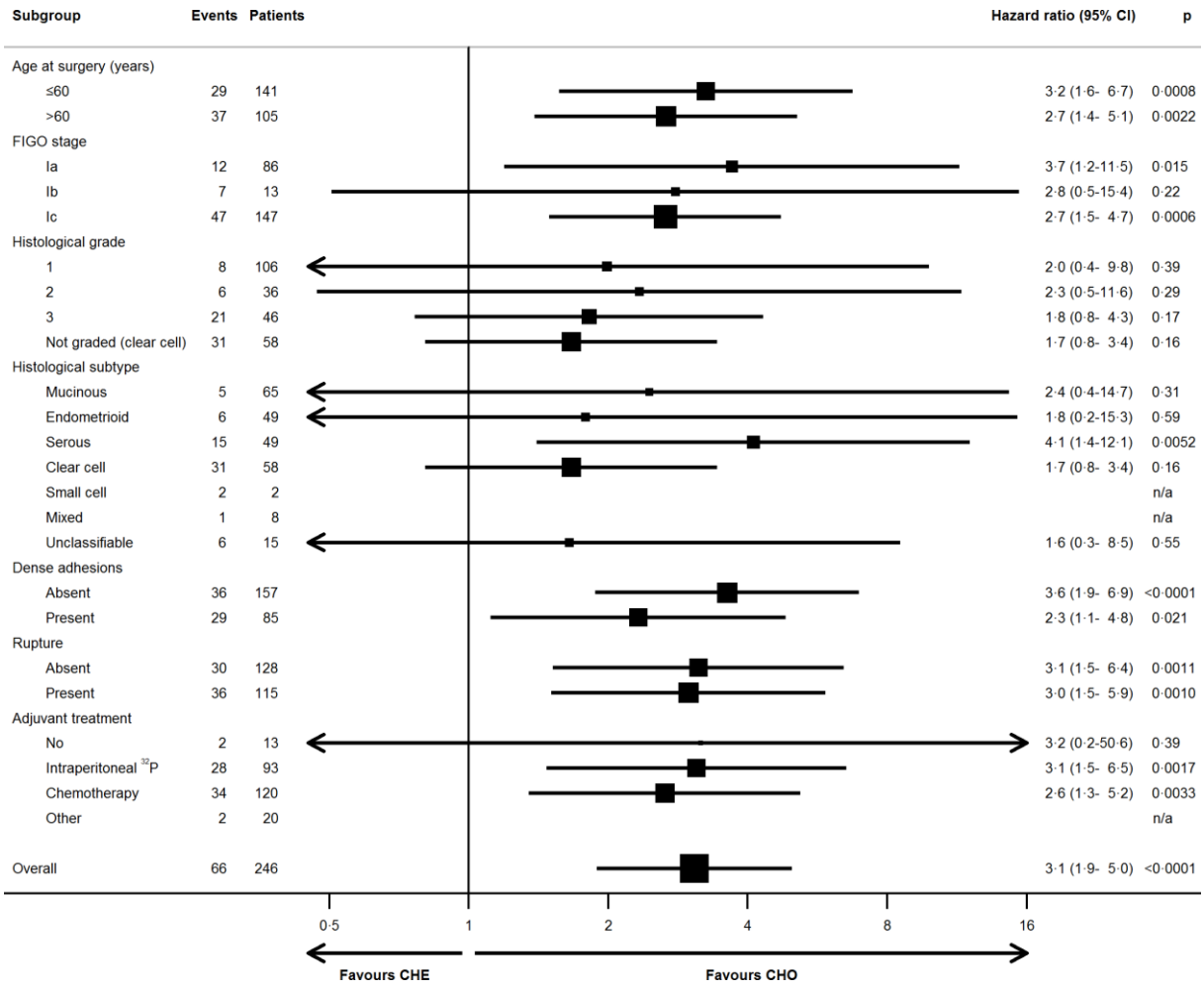

Figure A2: **Forest plot of the chromatin heterogeneity marker for the ovarian carcinoma patients in analysis of cancer-specific survival**

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CSD=cancer-specific death. FIGO=International Federation of Gynecology and Obstetrics. n/a=not applicable because all patients who suffered CSD had the same classification.

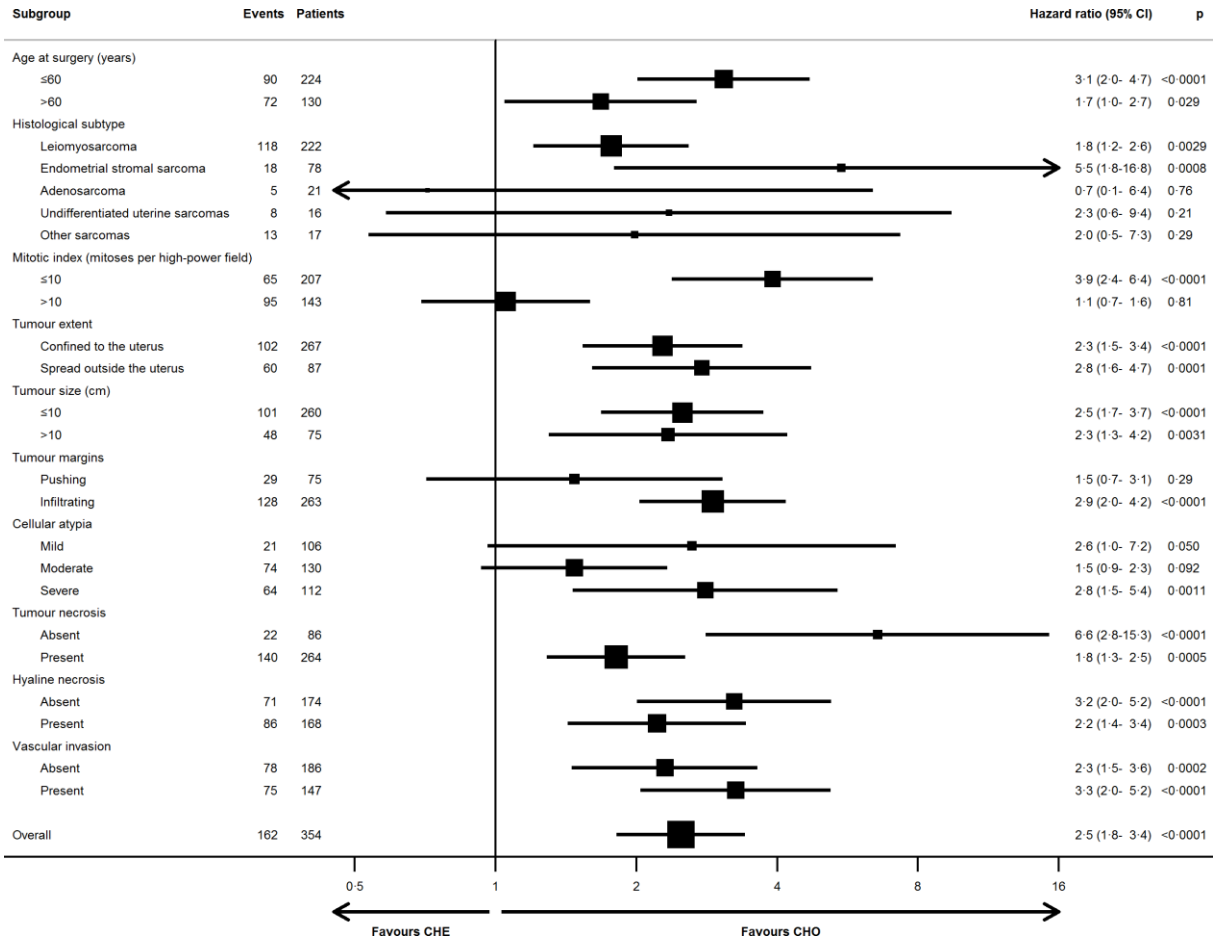

Figure A3: **Forest plot of the chromatin heterogeneity marker for the uterine sarcoma patients in analysis of 5-year cancer-specific survival**

CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CSD=cancer-specific death.

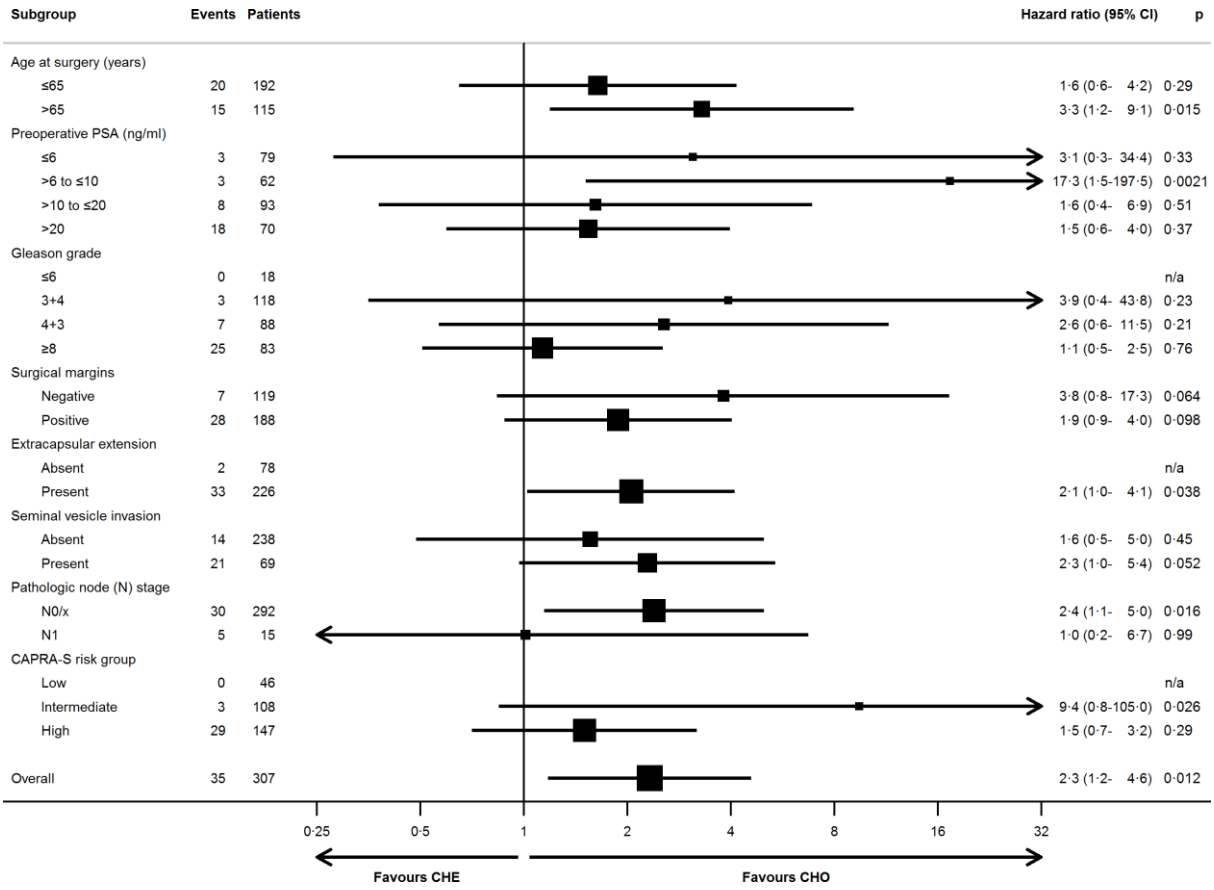

Figure A4: **Forest plot of the chromatin heterogeneity marker for the prostate cancer patients in analysis of cancer-specific survival**

The CAPRA-S score was categorised to give three CAPRA-S risk groups: Low risk if score 0 to 2; Intermediate risk if score 3 to 5; High risk if score 6 to 12. CAPRA-S=Cancer of the Prostate Risk Assessment Postsurgical. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CSD=cancer-specific death. n/a=not applicable because all patients who suffered CSD had the same classification. PSA=prostate-specific antigen.

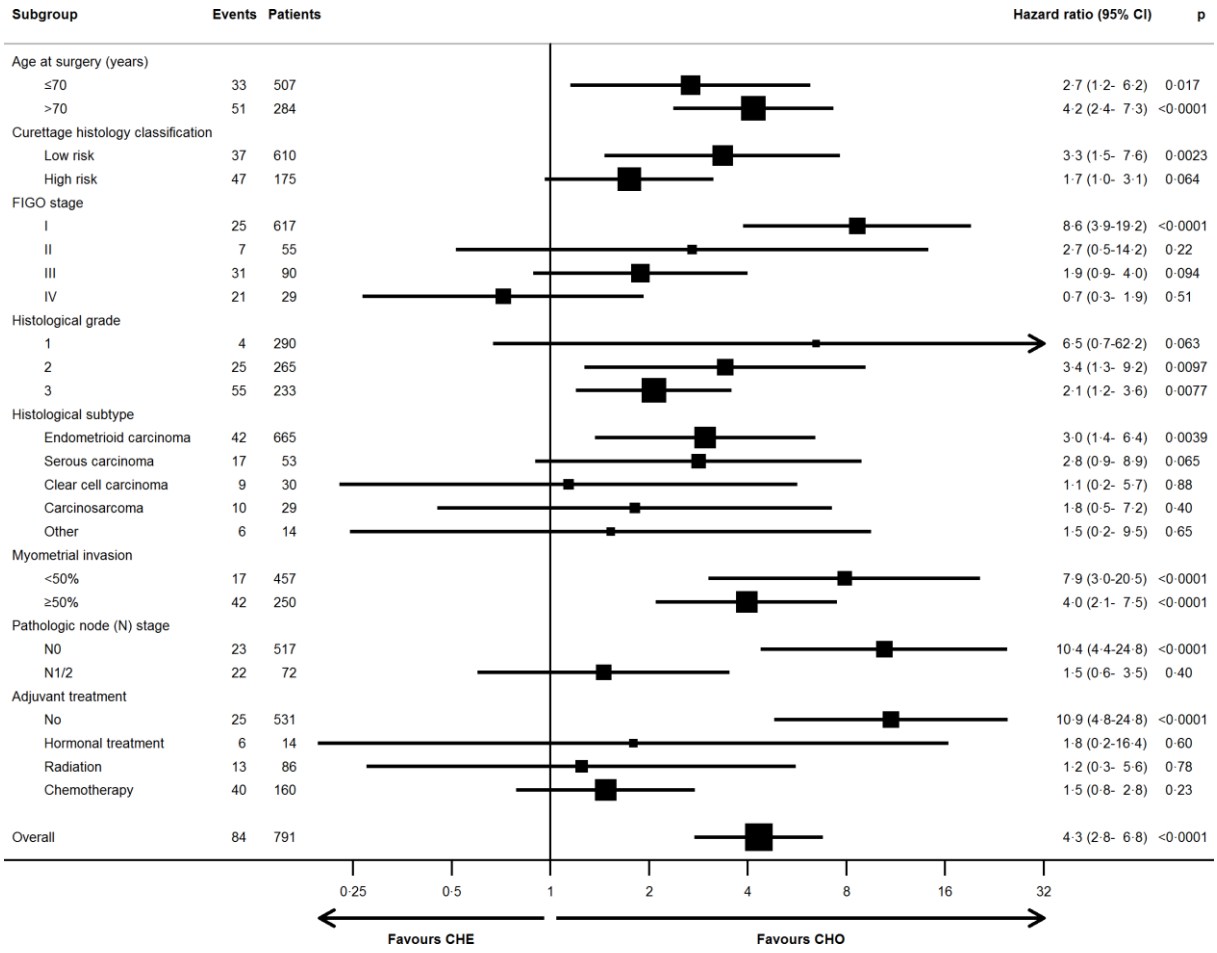

Figure A5: **Forest plot of the chromatin heterogeneity marker for the endometrial carcinoma patients in analysis of cancer-specific survival**

Curettage histology classification was: Low risk if benign, hyperplasia or endometrioid grade 1 or 2; High risk if non-endometrioid or endometrioid grade 3. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CSD=cancer-specific death. FIGO=International Federation of Gynecology and Obstetrics.

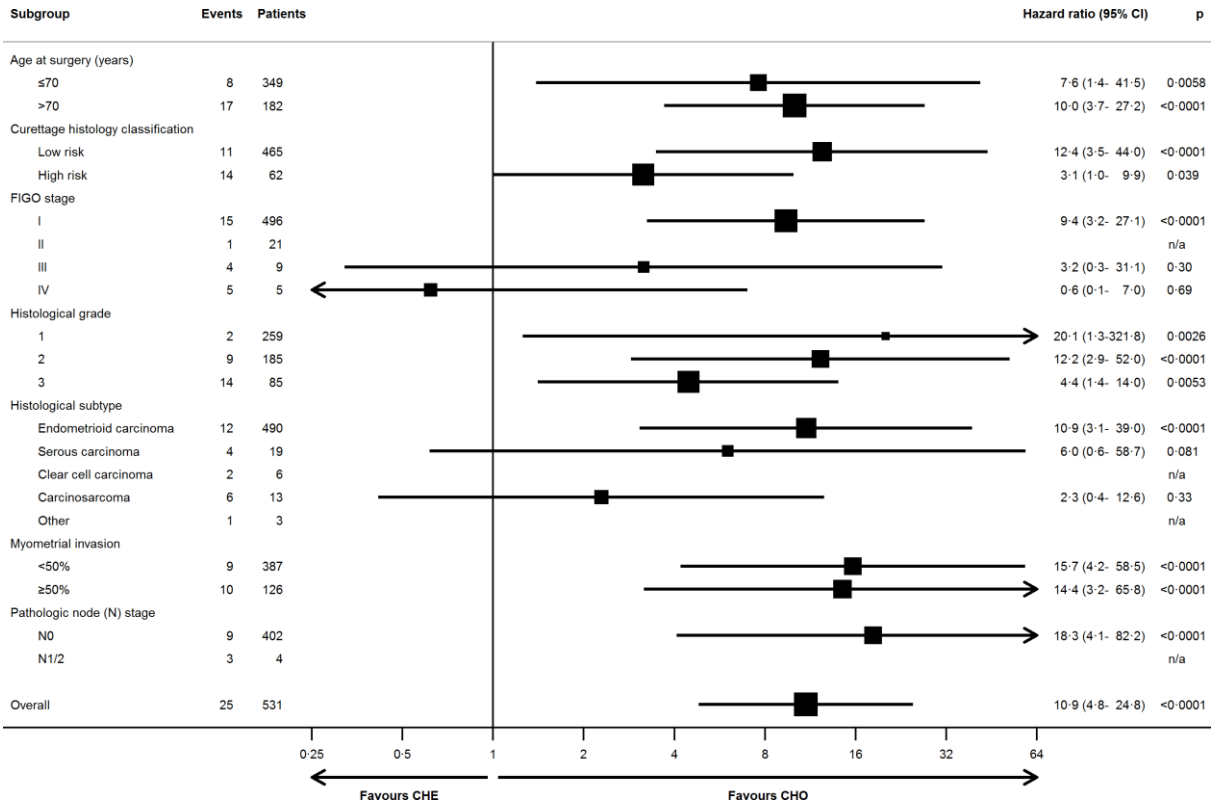

Figure A6: **Forest plot of the chromatin heterogeneity marker for the endometrial carcinoma patients without adjuvant treatment in analysis of cancer-specific survival**

Curettage histology classification was: Low risk if benign, hyperplasia or endometrioid grade 1 or 2; High risk if non-endometrioid or endometrioid grade 3. CHE=chromatin heterogeneous. CHO=chromatin homogeneous. CSD=cancer-specific death. FIGO=International Federation of Gynecology and Obstetrics. n/a=not applicable because all patients who suffered CSD had the same classification.