Supporting Information

Salvesen et al. 10.1073/pnas.0806514106

SI Methods

Selecting Gene Subsets with Good Combined Discriminatory Power. As our objective function to minimize for determination of maximal predictive power, we used the sum of squared residuals between the relative probability of the correct class label and that given by a diagonal linear discriminant analysis (DLDA) classifier. The relative probability given by a DLDA classifier is the probability density for the correct class divided by the sum of probability densities over both classes. Ideally a classifier assigns a relative probability of 1 to the correct class label in all cases, but in practice classifiers often do not perform this well.

We then tested increasing numbers of genes using a forward feature subset selection method (1) and found a 29-gene predictor gave the best results. These genes therefore were included in a gene set for validation with QRT-PCR.

Immunohistochemical Staining. Five-μm tissue microarray sections of paraffin-embedded tissue were stained as previously reported (2), using antigen retrieval for 10 min at 750 W and for 15 min at 350 W in citrate buffer (pH = 6). Slides were incubated 1 h at room temperature with polyclonal STMN1 antibody (#3352, Cell Signaling) diluted 1:50. A staining index was calculated as the product of staining intensity (0–3) and area of positive tumor cells (1 = < 10%, 2 = 10%–50%, 3 = > 50%). Values in the upper quartile (which corresponded to indices of 6 and 9) were considered positive.

SI Text Note 1. The association between 3q26.32 amplification and tumor recurrence suggests a causal relationship, with its functional effects leading to the aggressive phenotype. An alternative model would be that both 3q26.23 amplification and the aggressive phenotype are caused by a prior event, such as generalized aneuploidy in the cell, leading to an association but no direct causal link. Although this possibility cannot be ruled out, when we measured aneuploidy in 59 of the tumors with SNP array data, we did find that 3q26.32 amplification remains significantly associated with recurrence-free survival after adjustment for the impact of ploidy ($P = 0.03$). Amplification of 3q26.32 therefore seems to have an association with poor survival independent of the overall level of copy-number changes in the cell.

Ploidy was determined from DNA histograms based on measurement of 10^4-10^5 cells by flow cytometry as previously described (3), using fresh tumors and adjacent hematoxylin eosin (HE) sections to confirm malignant histology.

1. Bo T, Jonassen, I (2002) New feature subset selection procedures for classification of expression profiles. *Genome Biology* 3:1-0017.11.
2. Engelsen IB, et al. (2008) Low BMI-1 expression is associated with an activated BMI-1-driven signature, vascular invasion, and hormone receptor loss in endometrial carcinoma. *Br J Cancer* 98:1662–1669.
3. Salvesen HB, Iversen OE, Akslen LA (1998) Identification of high-risk patients by assessment of nuclear Ki-67 expression in a prospective study of endometrial carcinomas. *Clin Cancer Res* 4:2779–2785.

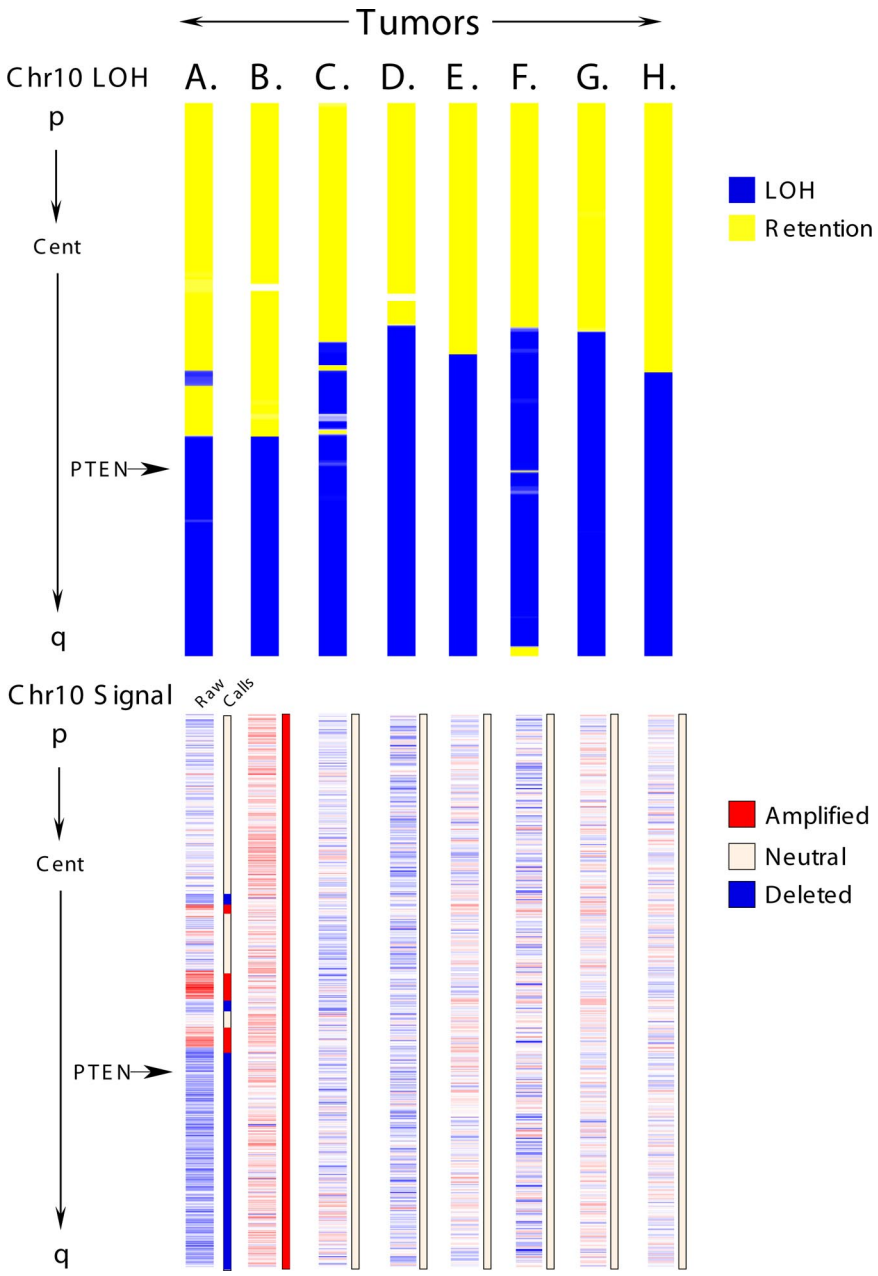

Fig. S1. LOH at the *PTEN* locus is rarely associated with deletions. The top panel displays loss of heterozygosity (LOH, *blue*) and retention of heterozygosity (*yellow*) along chromosome 10 for 8 endometrial carcinomas (labeled A–H) with LOH at the *PTEN* locus. The bottom panel displays signal intensities (red = high, blue = low, white = neutral) and copy-number calls (red bar = amplified, blue bar = deleted) for those endometrial carcinomas. LOH at *PTEN* is associated with deletion only in tumor A. It is associated with amplification in tumor B and neutral copy numbers in tumors C–H. Across our dataset, copy loss is seen in only 4 of 18 endometrial carcinomas observed to have LOH at the *PTEN* locus.

a Genes most over- and under-expressed in 3q26.32 amplified compared to unamplified samples*

Overexpressed		
NCBP2	I_1110064	SUV39H2
HPN	TAF4	PRKR
CCNE1	TRC8	AK3
ACVR2B	MRPS23	ASNA1
DDEF1	GGA3	LOC135763
UBE2C	I_930381	IFIT1
PCL1	OASL	AK2
UBL4	WBP11	PTD011
NMU	PLAGL2	GTF2E1
I_1100438	MGC15397	Apg4B
MX1	I_1100479	IFI27
LOC55831	PRRG1	ZNF148
EHHADH	FLJ12960	LOC90355
OPTN	MGC14595	PARL
TTK	UBE2E1	SLC2A4RG
I_961605	FLJ10511	C20orf1
FLJ14751	OAS3	

Underexpressed		
KLK11	GALNT7	TGIF
PIG3	RPL7A	FLJ13119
KIAA0776	NOL3	GOLPH2
ANXA5	EIF3S6IP	CPR8
ALDH2	USP8	I_1100588
ZNF143	NEIL1	FLJ20343
AES	AGR2	FLJ20136
TRAF5	TSPAN-5	I_959144
I_1109401	SDR1	SPRY2
LOC51103	ITGA6	I_1100131
SLC1A1	I_959279	CRYZ
SLC6A14	ARL4	HIRIP5
SPINK5	TOM1L1	TCN1
FLJ21868	ABCC3	SAT
HRY	DKFZp547O146	FLJ21439
FKSG2	IQGAP1	FLJ14950
TANK	FLJ21817	

b

Rank	Name of compound	Known target	N	p-value*
1	LY-294002	PI3-Kinase	17	0.003
2	Colforsin	adenylate cyclases	2	0.005
3	17-AG†	Hsp90	18	0.01
4	5224221		2	0.01
5	Resveratrol		5	0.02
6	Sirolimus	mTOR	10	0.04
7	carbamazepine	Sodium channel	3	0.05

N=number of active instances; *see methods; †17-allylamino-geldanamycin

c

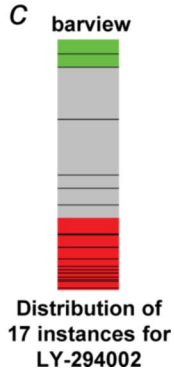

Fig. S2. The signatures of 3q26 amplification and a PI3K inhibitor are associated. (*A*) Genes with the highest (most overexpressed) and lowest (most underexpressed) *t* test statistic in the comparison between 3q26.32 amplified and unamplified samples. (*B*) Comparison of global expression patterns using the Connectivity Map (1) identifies the PI3K inhibitor LY-294002 as the most significantly anticorrelated with the 3q26 amplification signature among the 164 small molecules tested. Each of the N instances in which the small molecule was tested in the Connectivity Map was scored according to this signature. The p-value for each small molecule represents the distribution of these scores compared with the distribution of scores among all small molecules, using a permutation test as described by Lamb and colleagues (1). (*C*) The rankings of the 3q26 amplification signature scores for the 17 instances of LY-294002 are displayed as black bars. Rankings within the green and red regions represent scores that are respectively aligned with and divergent from the 3q26 amplification signature.

1. Lamb J, et al. (2006) The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935.

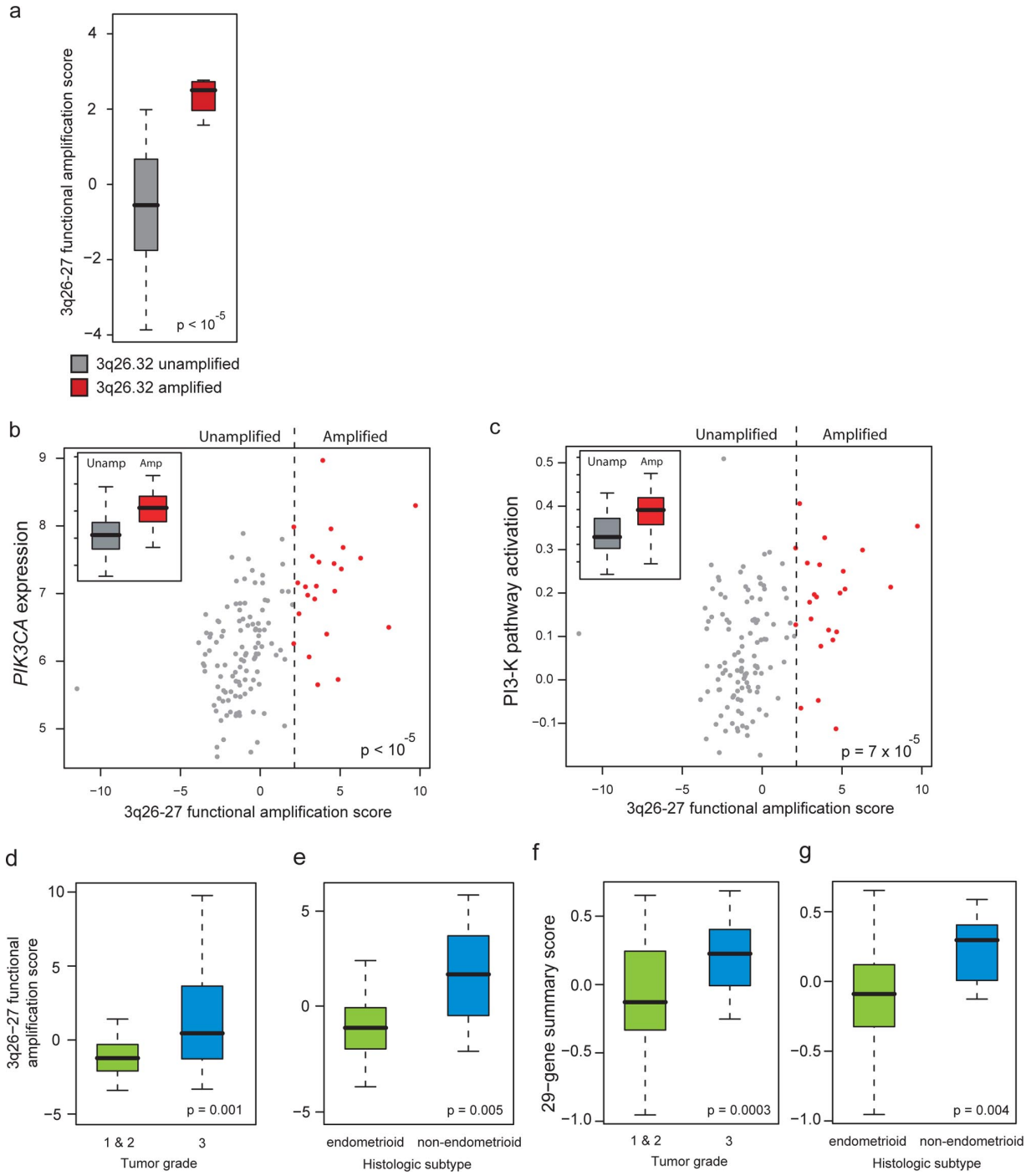

Fig. S3. Endometrial carcinomas with 3q26–27 amplification tend to overexpress *PIK3CA* and have high PI3K scores and aggressive features. (*A*) Functional amplification scores for 3q26–27 are significantly higher among samples with 3q26.32 amplification in the primary investigation set. Functional amplification scores were generated from genome-wide expression data as previously described (1). Amplification of 3q26.32 was determined by SNP array analysis (see *Methods*). This correlation validates the use of these functional amplification scores as a measure of 3q26–27 amplification. (*B*) Expression of *PIK3CA* and (*C*) PI3K scores are higher among tumors in the validation set with high 3q26–27 functional amplification scores. The vertical lines represent the minimum threshold of functional amplification score typically seen in 3q26.32-amplified tumors in (*A*). The box plot insets represent *PIK3CA* expression levels and PI3K scores in samples with functional amplification scores lower and higher than this threshold, representing tumors likely to be unamplified and amplified, respectively, at the *PIK3CA* locus. P-values correspond to a test of the null hypothesis that the Pearson product-moment correlation was equal to 0. High 3q26–27 functional amplification scores are also associated with (*D*) high grade and (*E*) non-endometrioid histology, as are (*F*, *G*) high 29-gene summary scores, representing membership in expression Cluster 2. These comparisons were made in the validation dataset. P-values were calculated using a 1-sided Wilcoxon test.

1. Carter SL, et al. (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38:1043–1048.

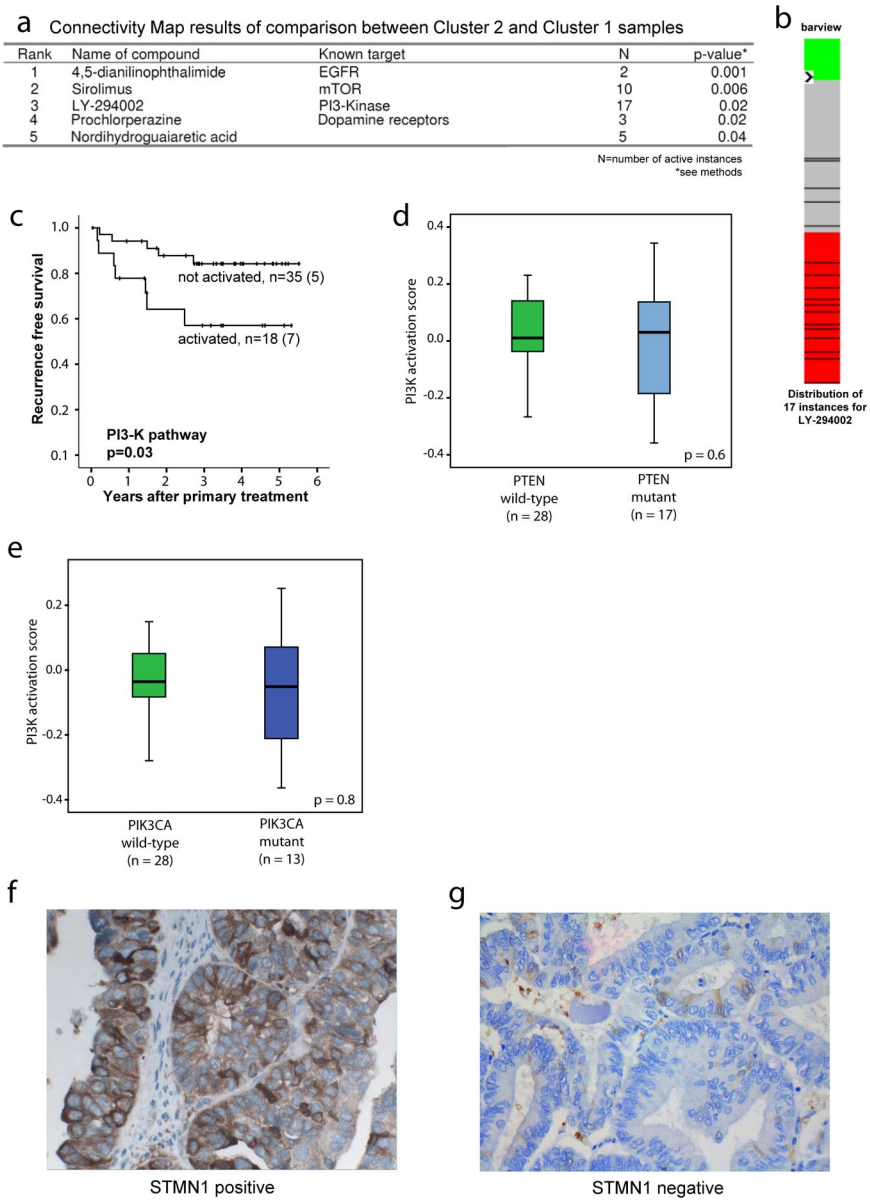

Fig. S4. Association between measures of PI3K activation and genomic and proteomic features of endometrial cancer. (*A*) The Connectivity Map (1) showed that the anticorrelation between the Cluster 2 expression signature and the PI3K inhibitor LY-294002 was the third strongest among the 164 small molecules tested. (*B*) The rankings of the Cluster 2 signature scores for the individual instances of LY-294002 in the Connectivity Map are displayed as in Fig. S2C. (*C*) PI3K scores, an alternative measure of PI3K pathway activation, are significantly associated with decreased recurrence-free survival. (*D*) PI3K scores are not significantly associated with *PTEN* mutations or with (*E*) *PIK3CA* mutations. P-values were calculated using a 2-sided Wilcoxon test. (*F*) High expression of the PI3K pathway member STMN1, seen by immunohistochemistry in a tumor with amplification of *PIK3CA* and elevated expression of *PIK3CA* and the PI3K activation signature. (*G*) Low expression of STMN1, seen in a tumor without amplification of *PIK3CA* or elevated expression of *PIK3CA* or the PI3K activation signature.

1. Lamb J, et al. (2006) The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935.

Table S1. Patient characteristics and histopathologic variables for the endometrial carcinoma series studied compared with a population-based patient series from the same region

		Expression Arrays (n = 57)		SNP Arrays (n = 74)	
Characteristic	General Population (n = 285)	Raw value	P-value*	Raw value	P-value*
Median age	65 years (range, 33–92 years)	63 years (range, 39–91 years)	0.8	64 years (range, 39–91)	0.8
FIGO stage III or IV	55 (19%)	9 (16%)	0.7	12 (16%)	0.7
Non-endometrioid histology	29 (10%)	6 (11%)	1	8 (11%)	0.8
Histologic grade 3	59 (21%)	13 (23%)	0.7	15 (21%)	1

*Compared to general population, using a Fisher's exact test, except for age comparisons in which a Kruskal-Wallis test was used

Table S2. Differences in clinical and histopathologic characteristics between expression clusters I and II and summarized gene mRNA predictor in microarray and qPCR datasets

Characteristic	Cluster 1 ($n = 29$)	Cluster 2 ($n = 28$)	P-value*	Microarray P-value*	qPCR P-value*
Median age	62 years	67 years	0.6		
FIGO stage III or IV	1 (3%)	8 (29%)	0.01	0.02	0.01
Non-endometrioid histology	0 (0%)	6 (21%)	0.01	0.03	0.003
Histologic grade 3	1 (3%)	12 (43%)	< 0.001	0.001	0.001
Depth of myometrial infiltration	6†	8.5†	0.06	n.s.	0.05
Mitotic figures	4‡	10‡	0.002	0.002	0.01
Loss of estrogen receptor	4 (14%)	12 (43%)	0.02	n.s.	0.01
Loss of progesterone receptor	2 (7%)	11 (39%)	0.005	n.s.	0.002
Presence of necrosis	13 (45%)	22 (76%)	0.01	0.03	0.002
Vascular invasion	6 (27%)	16 (73%)	0.007	0.02	0.002
Type II tumor§	1 (3%)	16 (57%)	< 0.001	< 0.001	< 0.001
Aggressive cluster 2				< 0.001	< 0.001

*Using a Fisher's exact test except for comparisons between continuous variables, in which a Mann-Whitney U test was used.
†Median value in millimeters.
‡Median number of mitotic figures per 10 fields at original magnification × 40.
§Defined as non-endometrioid, high-grade endometrioid, or lacking both estrogen receptor and progesterone receptor.
N.s., not significant.

Table S3. Univariate (UV) and multivariate survival analysis of recurrence-free survival among 53 endometrial cancer patients cured by surgery and adjusted hazard ratios (HRs) for recurrence among endometrial carcinomas for cluster 2 and histopathologic variables

Histopathologic Variable	Univariate Survival P-value*	Hazard Ratio Cluster 2†	P-value‡	Hazard Ratio Histopathologic Variable§	P-value‡
FIGO stage III or IV	n.s.	3.0 (0.9–10.4)	0.08	1.3 (0.3–6.3)	0.7
Non-endometrioid histology	0.17	2.9 (0.8–10.2)	0.1	1.6 (0.3–7.9)	0.6
Histologic grade 3	0.009	2.0 (0.5–8.2)	0.3	2.9 (0.8–10.8)	0.1
Loss of estrogen receptor	n.s.	3.2 (0.9–11.2)	0.07	1.0 (0.3–3.3)	0.9
Loss of progesterone receptor	0.14	2.7 (0.7–10.2)	0.1	1.5 (0.4–5.6)	0.6

*Log rank test.
†Hazard ratio for cluster 2, adjusted for listed histopathologic variable in column 1; 95% confidence intervals in parentheses.
‡Using a log-ratio test.
§Hazard ratio for histologic variable in column 1, adjusted for cluster segregation; 95% confidence intervals in parentheses.
N.s., not significant.

Table S4. List of peak regions of amplification, deletion, and non-overlapping LOH

Amplification	Cytoband	Boundaries of Peak*	Q-value	Frequency† (%)	Oncogene or Tumor Suppressor Gene in Region
1	1p34.3	30.1–41.9	2e-3	40	
2	1q42.13	224.8–237.7	3e-18	47	*LMYC1*
3	3q26.32	173.3–184.9	1e-14	15	*PIK3CA*
4	6p21.2	37.5–38.3	0.11	15	
5	7p11.2	54.5–65.2	0.13	12	*EGFR*
6	8q24.21	122.6–129.7	4e-8	26	*MYC*
7	10q22.2	75.0–79.7	0.07	16	
8	12p12.1	25.5–27.8	0.02	12	*KRAS*
9	17q12	36.3–49.1	0.08	26	*ERBB2*
10	19q12	34.0–35.6	4e-7	45	*CCNE1*
11	20q13.2	45.3–57.7	2e-3	23	*AURKA*
Deletion					
1	1p36.32	0–19.3	0.24	10	*CHD5/PAX7*
2	3p26.2	0–3.5	0.13	4	*FBXW7*
3	4q34.3	92.1–183.3	8e-4	8	
4	6q16.3	10.48–10.49	0.04	9	
5	7p22.2	0–4.2	0.12	10	
6	7q33	109.2–137.1	0.11	4	
7	8p21.2	25.1–29.8	0.02	12	
8	11q23.3	116.5–134.5	0.18	8	
9	13q12.3	0–27.8	0.22	6	
10	15q26.3	97.9–100.3	0.07	10	
11	16q21	58.3–65.1	1e-3	18	
12	17p12	11.5–12.2	0.02	8	
13	22q13.2	41.7–44.3	0.18	10	

*Mb coordinates using hg16 build.

†Frequency of amplification or deletion to any level. High-level amplifications were seen for LMYC, PIK3CA, EGFR, 6p21.2, EGFR, CCNE1 (1 case each), and MYC (2 cases).

Table S5. Correlations between amplifications and recurrence-free survival

Cytoband	Oncogene in Region	Number of Amplified Samples (and Recurrences)*	Mean Survival Among Amplified Samples†	Mean Survival Among Unamplified Samples†	Bonferroni-Corrected P-Value‡
1p34.3		25 (5)	4.2	4.7	1.0
1q42.13	*LMYC*	33 (8)	4.2	4.9	0.9
3q26.32	*PIK3CA*	9 (4)	2.4	4.8	0.03
6p21.2		8 (3)	2.6	4.7	0.4
7p11.2	*EGFR*	8 (2)	3.7	4.6	1.0
8q24.21	*MYC*	16 (4)	4.0	4.7	1.0
10q22.2		10 (4)	3.0	4.8	0.6
12p12.1	*KRAS*	7 (4)	2.4	4.9	0.01
17q12	*ERBB2*	16 (4)	4.2	4.7	1.0
19q12	*CCNE1*	29 (6)	4.3	4.7	1.0
20q13.2	*AURKA*	14 (5)	3.3	4.8	0.4

*Number of samples amplified to any degree among the 68 completely resected tumors with follow-up clinical data, with number of recurrences among amplified samples (of 13 recurrences overall) in parentheses.
†In years after primary surgery.
‡By log-rank test, after Bonferroni correction for 11 hypotheses.

Table S6. Immunohistochemical Stathmin (STMN1) expression in a population-based series of endometrial carcinomas: correlation with clinical phenotype

Characteristic	Stathmin Negative	Stathmin Positive	P-value[†]
Original tumor series*	$n = 53$	$n = 19$	
FIGO stage III or IV	13%	11%	n.s.
Non-endometrioid histology	9%	21%	0.19
Histologic grade 3	17%	47%	0.01
Depth of myometrial infiltration[‡]	6.5	7.5	n.s.
Mitotic rate[‡]	7	18	0.01
Loss of ER	25%	32%	n.s.
Loss of PR	19%	53%	0.005
Presence of necrosis	50%	77%	0.06
Vascular invasion	33%	65%	0.02
Type II tumor[§]	26%	58%	0.01
Aggressive cluster[¶]	44%	75%	0.03
Validation series*	$n = 175$	$n = 66$	
FIGO stage III or IV	17%	21%	n.s.
Non-endometrioid histology	9%	15%	0.18
Histologic grade 3	14%	32%	0.001
Depth of myometrial infiltration[‡]	4	7	0.003
Mitotic rate[‡]	8	14	< 0.001
Loss of estrogen receptor	18%	45%	< 0.001
Loss of progesterone receptor	23%	48%	< 0.001
Presence of necrosis	52%	77%	< 0.001
Vascular invasion	34%	62%	< 0.001
Type II tumor[§]	35%	64%	< 0.001

*In the original tumor series, data are missing in 3 cases for histologic grade, depth of myometrial infiltration, mitotic rate, necrosis and vascular invasion, and for cluster annotation in 19 cases. In the validation series, data are missing in 1 case for FIGO stage, in 6 cases for estrogen receptor/progesterone receptor status, in 38 cases for depth of myometrial infiltration, and in 7 cases for type I/II classification.

[†]Using a Pearson chi-square test when otherwise not specified.

[‡]Median depth of myometrial infiltration in millimeters, number of mitotic figures per 10 fields at magnification 40 $\times$, Mann-Whitney U test.

[§]Defined as either non-endometrioid, high-grade endometrioid, or lacking both estrogen receptor and progesterone receptor.

[¶]$N = 53$ cases, 1-sided test.

Ns = not significant.

Table S7. Immunohistochemical Stathmin (STMN1) expression in a population-based series of endometrial carcinomas: multivariate survival analysis of clinicopathologic variables

Characteristic	Hazard Ratio*	P-value†
Age‡	1.048 (1.024–1.072)	< 0.001
FIGO stage		
I	1	< 0.001
II	3.04 (1.36–6.77)	
III	10.65 (5.87–19.31)	
IV	48.74 (20.74–114.57)	
Histologic subtype		
Endometrioid	1	0.2
Non-endometrioid	1.59 (0.79–3.23)	
Histologic grade		
1	1	0.38
2	1.73 (0.75–3.96)	
3	1.36 (0.50–3.76)	
STMN1 staining		
Negative	1	0.004
Positive	2.14 (1.28–3.59)	

*Hazard ratio with 95% confidence intervals in parentheses, based on the Cox proportional hazards model.
†Using a log-ratio test.
‡Continuous variable with hazard ratio given per year.