

## **Supplementary Information**

Supplementary Figure S1. Test of the methylation status of candidate genes used pooled DNA from endometrial tissues and cervical scrapings.

Supplementary Figure S2. Verification of 14 candidate genes in the pilot set of cervical scrapings.

Supplementary Table S1: Demographics relating to the two data sets of genome-wide DNA methylation.

Supplementary Table S2. Demographics related to the clinical samples.

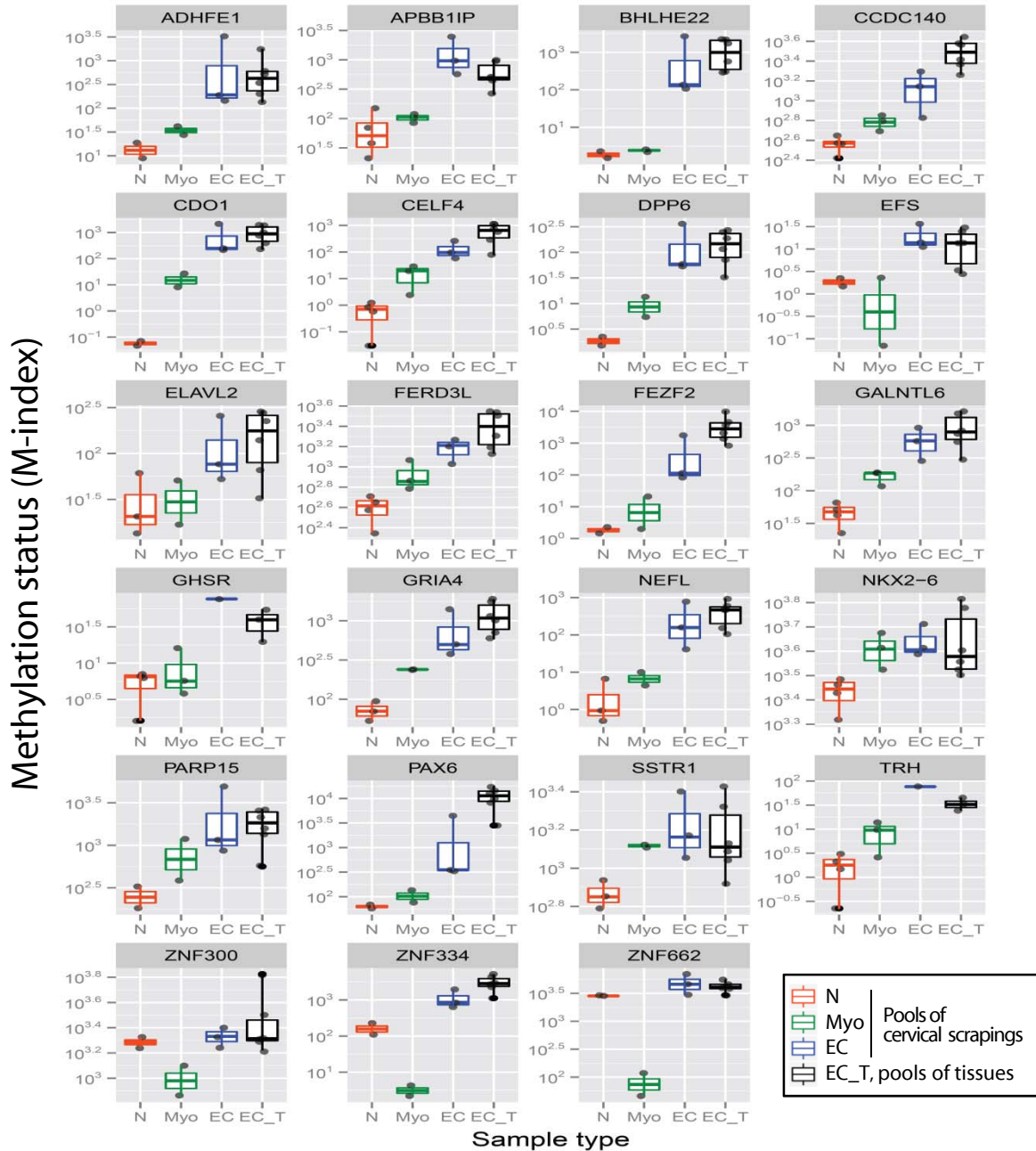
Supplementary Table S3: Summary the area under the receiver-operating characteristic curve for distinction between endometrial cancer and noncarcinoma samples in pilot testing.

Supplementary Table S4: Summary of cophenetic correlation results for analysis of consensus clustering of DNA methylation patterns.

Supplementary Table S5: List of 180 hypermethylated genes in four clusters.

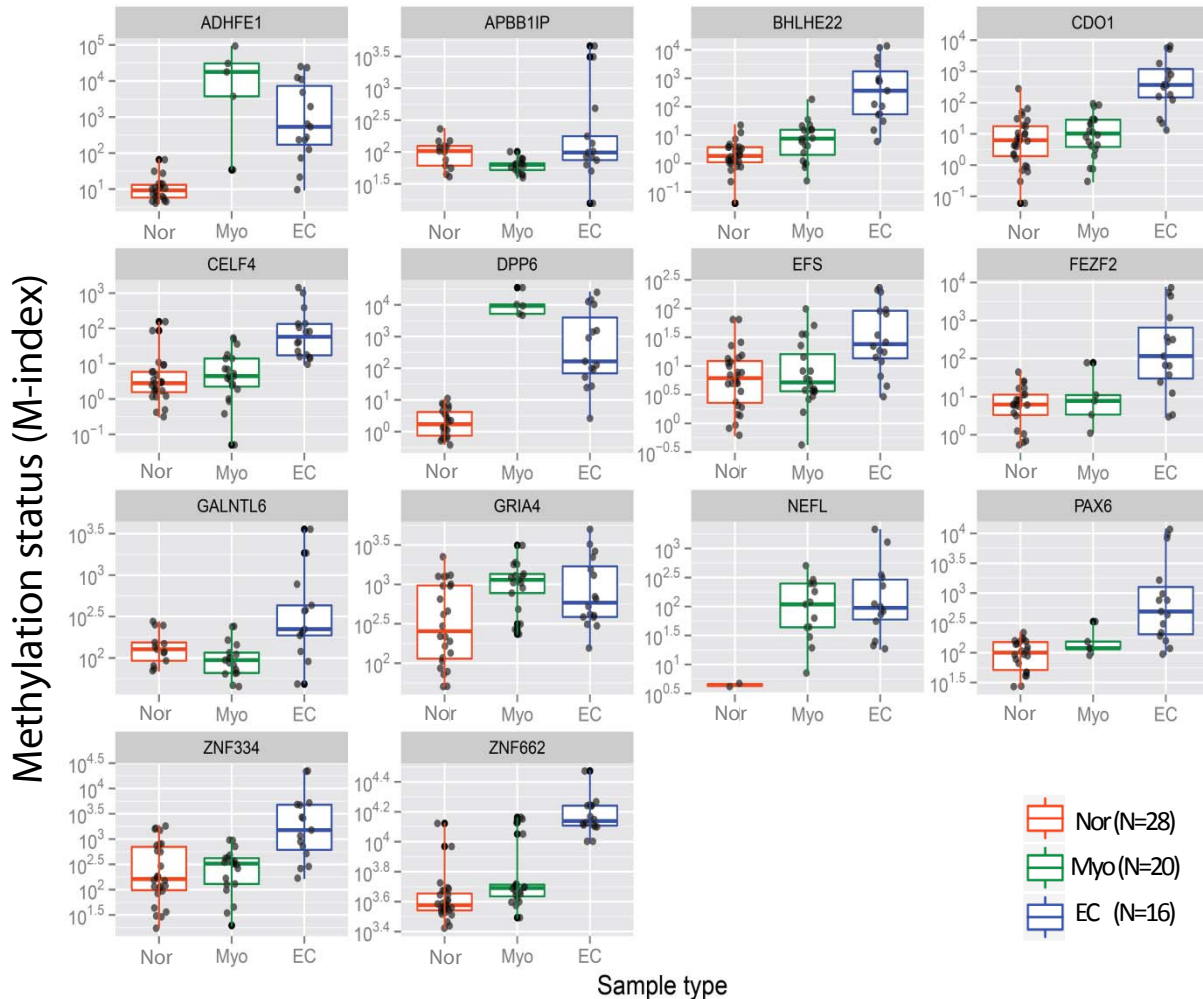
Supplementary Table S6. Summary of the methylation levels of four candidate genes in cervical scrapings from endometrial cancer and noncarcinoma samples.

## DNA pools



**Supplemental Figure S1. Test of the methylation status of candidate genes used pooled DNA from endometrial tissues and cervical scrapings.** Twenty-three candidate genes show hypermethylation in tissue and cervical scrapings from EC. Each spot is a pooled sample from 5 equal amounts DNA of tissue samples. We selected fourteen genes, *ADHFE1*, *APBB1IP*, *BHLHE22*, *CDO1*, *CELF4*, *DPP6*, *EFSFEZF2*, *GALNTL6*, *GRIA4*, *NEFL*, *PAX6*, *ZNF334*, and *ZNF662*, that demonstrated more specific higher DNA methylation in pooled cervical scrapings from endometrial cancer for further validation. N, normal; Myo, myoma, EC, endometrial cancer.

## Pilot set of individual cervical scrapings



**Supplemental Figure S2. Verification of fourteen candidate genes in the pilot set of cervical scrapings.** We examined the DNA methylation in pilot set for reducing candidate genes. Nor, normal; Myo, myoma; EC, endometrial cancer; N, case number.

**Supplementary Table S1:** Demographics relating to the two data sets of genome-wide DNA methylation.

Variable	Set A (TCGA)				Set B (MethylCap-seq)			
	EC	(%)	Adjacent-normal	(%)	EC	(%)	Adjacent-normal	(%)
Case number	270		28		78		11	
Age(yrs)±SD	62.1±11.7		58.6±10.0		64.6±10.2			
30 ~ 50 yrs	34	12.6%	5	17.9%	3	4.5%		
> 50 yrs	236	87.4%	23	82.1%	64	95.5%		
Stage								
I	186	68.9%			50	74.6%		
II	25	9.3%			11	16.4%		
III	50	18.5%			6	9.0%		
IV	9	3.3%			0	0.0%		
Grade								
1	55	20.4%			32	47.8%		
2	76	28.1%			29	43.3%		
3	139	51.5%			6	9.0%		

TCGA, The cancer genome atlas; MethylCap-seq, methyl-DNA binding domain capture coupled with next-generation sequencing; EC, endometrial cancer; SD, standard deviation.

**Supplementary Table S2.** Demographics related to the clinical samples.

Variable	Tissue	Cervical scrapings
<b>Endometrial cancer</b>		
Case number	30	50
Age (yrs)±SD	52.4 ± 10.8	54.0 ± 8.4
Stage		
I	15	38
II	1	3
III	14	8
Unkown	0	1
Grade		
G1	4	20
G2	17	13
G3	7	4
Unkown	2	13
<b>Myoma</b>		
Case number		40
Age (yrs)±SD		44.1 ± 5.9
Histology		
Adenomyosis		7
Leiomyoma		18
Mix		16
<b>Normal</b>		
Case number		56
Age (yrs)±SD		49.0 ± 12.3

SD, standard deviation; Stage, the International Federation of Gynecology and Obstetrics (FIGO) stage

**Supplementary Table S3:** Summary the area under the receiver-operating characteristic curve for distinction between endometrial cancer and noncarcinoma samples in pilot testing.

NO.	Gene Name	AUC (95% CI)	
1	BHLHE22	0.96	(0.91 to 1.00)
2	ZNF662	0.96	(0.91 to 1.00)
3	CDO1	0.95	(0.89 to 1.00)
4	CELF4	0.92	(0.86 to 0.99)
5	PAX6	0.87	(0.74 to 1.00)
6	ZNF334	0.85	(0.73 to 0.96)
7	FEZF2	0.84	(0.68 to 1.00)
8	GALNTL6	0.80	(0.63 to 0.97)
9	EFS	0.80	(0.67 to 0.93)
10	DPP6	0.78	(0.59 to 0.96)
11	ELAVL2	0.77	(0.56 to 0.98)
12	APBB1IP	0.69	(0.50 to 0.88)
13	NEFL	0.61	(0.40 to 0.82)
14	GRIA4	0.40	(0.23 to 0.57)

Abbreviation: AUC, area under an receiver-operating curve. CI, confidence interval. The pilot set included 16, 20, and 28 cervical scrapings from patients with EC, myoma, and normal endometrium.

**Supplementary Table S4:** Summary of cophenetic correlation results for analysis of consensus clustering of DNA methylation patterns.

Maximum iteration	Run multiple rank (number of cluster)			
	2	3	4	5
1000	0.860	0.834	0.564	0.797
1500	0.870	0.840	0.904	0.815
2000	0.846	0.871	0.902	0.569
2500	0.865	0.865	0.902	0.842
3000	0.874	0.902	0.912	0.867
3500	0.877	0.886	0.903	0.839
4000	0.877	0.909	0.922	0.881
4500	0.877	0.919	0.928	0.879
5000	0.877	0.587	0.932	0.893

The red boxes show the cophenetic correlation >0.92. The highest value means the best results for clustering.

**SupplementaryTable S5: List of 180 hypermethylated genes in four clusters.**

NO.	RefSeq	ENTREZ_ID	Gene Name	NMF cluster
1	NM_001111270	115557	ARHGEF25	cluster c
2	NM_001801	1036	CDO1	cluster c
3	NM_001145135	1375	CPT1B	cluster c
4	NM_183374	340665	CYP26C1	cluster c
5	NM_007046	11117	EMILIN1	cluster c
6	NM_001977	2028	ENPEP	cluster c
7	NM_033266	10595	ERN2	cluster c
8	NM_012306	23017	FAIM2	cluster c
9	NM_004122	2693	GHSR	cluster c
10	NM_012205	23498	HAAO	cluster c
11	NM_022377	3386	ICAM4	cluster c
12	NM_033267	153572	IRX2	cluster c
13	NM_005545	3671	ISLR	cluster c
14	NM_001024660	8997	KALRN	cluster c
15	NM_001297433	10296	MAEA	cluster c
16	NM_006211	5179	PENK	cluster c
17	NM_001193341	221421	RSPH9	cluster c
18	NM_001105204	23623	RUSC1	cluster c
19	NM_020949	57709	SLC7A14	cluster c
20	NM_012443	9576	SPAG6	cluster c
21	NM_152380	6913	TBX15	cluster c
22	NM_024600	79652	TMEM204	cluster c
23	NM_007117	7200	TRH	cluster c
24	NM_006634	10791	VAMP5	cluster c
25	NM_001198552	7490	WT1	cluster c
26	NM_144650	137872	ADHFE1	cluster d
27	NM_170732	627	BDNF	cluster d
28	NM_001082534	56934	CA10	cluster d
29	NM_005181	761	CA3	cluster d
30	NM_021827	60494	CCDC81	cluster d
31	NM_033646	1005	CDH7	cluster d
32	NM_001039936	1124	CHN2	cluster d
33	NM_001029865	120237	DBX1	cluster d
34	NM_004432	1993	ELAVL2	cluster d
35	NM_001099439	284656	EPHA10	cluster d
36	NM_152898	222894	FERD3L	cluster d
37	NM_001127643	2554	GABRA1	cluster d
38	NM_001127644	2554	GABRA1	cluster d
39	NM_000829	2893	GRIA4	cluster d
40	NM_001112812	2893	GRIA4	cluster d
41	NM_001278067	2911	GRM1	cluster d
42	NM_016815	2995	GYPC	cluster d
43	NM_138610	9555	H2AFY	cluster d
44	NM_021062	3018	HIST1H2BB	cluster d
45	NM_002500	4760	NEUROD1	cluster d
46	NM_002511	4829	NMBR	cluster d
47	NM_138735	9378	NRXN1	cluster d
48	NM_001270524	5015	OTX2	cluster d
49	NM_001113523	165631	PARP15	cluster d
50	NM_018898	56135	PCDHAC1	cluster d



**SupplementaryTable S5: List of 180 hypermethylated genes in four clusters. (Con't)**

NO.	RefSeq	ENTREZ_ID	Gene Name	NMF cluster
51	NM_032092	56105	PCDHGA11	cluster d
52	NM_003735	26025	PCDHGA12	cluster d
53	NM_018918	56110	PCDHGA5	cluster d
54	NM_018920	56108	PCDHGA7	cluster d
55	NM_018924	56102	PCDHGB3	cluster d
56	NM_018926	56100	PCDHGB6	cluster d
57	NM_018927	56099	PCDHGB7	cluster d
58	NM_032101	56099	PCDHGB7	cluster d
59	NM_001100391	138046	RALYL	cluster d
60	NM_001206671	79608	RIC3	cluster d
61	NM_032291	84251	SGIP1	cluster d
62	NM_052910	114798	SLITRK1	cluster d
63	NM_014926	22865	SLITRK3	cluster d
64	NM_001145102	4088	SMAD3	cluster d
65	NM_018967	54212	SNTG1	cluster d
66	NM_015879	51046	ST8SIA3	cluster d
67	NM_001059	6870	TACR3	cluster d
68	NM_080872	137970	UNC5D	cluster d
69	NM_199424	11060	WWP2	cluster d
70	NM_032153	84107	ZIC4	cluster d
71	NM_003451	7730	ZNF177	cluster d
72	NM_001172832	91975	ZNF300	cluster d
73	NM_018102	55713	ZNF334	cluster d
74	NM_052852	90649	ZNF486	cluster d
75	NM_001134442	91392	ZNF502	cluster d
76	NM_182609	342926	ZNF677	cluster d
77	NM_020063	343472	BARHL2	cluster a
78	NM_170731	627	BDNF	cluster a
79	NM_152414	27319	BHLHE22	cluster a
80	NM_153038	151278	CCDC140	cluster a
81	NM_020180	56853	CELF4	cluster a
82	NM_001010852	134829	CLVS2	cluster a
83	NM_014141	26047	CNTNAP2	cluster a
84	NM_021110	7373	COL14A1	cluster a
85	NM_019609	56265	CPXM1	cluster a
86	NM_004746	9229	DLGAP1	cluster a
87	NM_018008	55079	FEZF2	cluster a
88	NM_021911	2561	GABRB2	cluster a
89	NM_052917	114805	GALNT13	cluster a
90	NM_001034845	442117	GALNTL6	cluster a
91	NM_001480	2587	GALR1	cluster a
92	NM_145793	2674	GFRA1	cluster a
93	NM_000830	2897	GRIK1	cluster a
94	NM_000831	2899	GRIK3	cluster a
95	NM_133445	116443	GRIN3A	cluster a
96	NM_021973	9464	HAND2	cluster a
97	NM_002148	3236	HOXD10	cluster a
98	NM_006043	9956	HS3ST2	cluster a
99	NM_000524	3350	HTR1A	cluster a
100	NM_001277285	22997	IGSF9B	cluster a

**SupplementaryTable S5: List of 180 hypermethylated genes in four clusters. (Con't)**

NO.	RefSeq	ENTREZ_ID	Gene Name	NMF cluster
101	NM_024337	79192	IRX1	cluster a
102	NM_002202	3670	ISL1	cluster a
103	NM_001160132	56479	KCNQ5	cluster a
104	NM_022363	64211	LHX5	cluster a
105	NM_173576	283078	MKX	cluster a
106	NM_004540	4685	NCAM2	cluster a
107	NM_177400	84504	NKX6-2	cluster a
108	NM_000905	4852	NPY	cluster a
109	NM_003269	7101	NR2E1	cluster a
110	NM_007226	11249	NXPH2	cluster a
111	NM_005806	10215	OLIG2	cluster a
112	NM_181459	5077	PAX3	cluster a
113	NM_001127612	5080	PAX6	cluster a
114	NM_002590	5100	PCDH8	cluster a
115	NM_001037341	5142	PDE4B	cluster a
116	NM_002600	5142	PDE4B	cluster a
117	NM_004575	5458	POU4F2	cluster a
118	NM_024504	63978	PRDM14	cluster a
119	NM_001199764	5798	PTPRN	cluster a
120	NM_007050	11122	PTPRT	cluster a
121	NM_012293	7837	PXDN	cluster a
122	NM_013435	30062	RAX	cluster a
123	NM_001282863	340419	RSPO2	cluster a
124	NM_001035	6262	RYR2	cluster a
125	NM_015265	23314	SATB2	cluster a
126	NM_005068	6492	SIM1	cluster a
127	NM_001044	6531	SLC6A3	cluster a
128	NM_003108	6664	SOX11	cluster a
129	NM_001049	6751	SSTR1	cluster a
130	NM_013305	29906	ST8SIA5	cluster a
131	NM_001270484	6862	T	cluster a
132	NM_001166220	57057	TBX20	cluster a
133	NM_001256272	30813	VSX1	cluster a
134	NM_022479	64409	WBSCR17	cluster a
135	NM_003433	7691	ZNF132	cluster a
136	NM_139056	170690	ADAMTS16	cluster b
137	NM_018702	105	ADARB2	cluster b
138	NM_001115	114	ADCY8	cluster b
139	NM_001099733	116	ADCYAP1	cluster b
140	NM_001117	116	ADCYAP1	cluster b
141	NM_000680	148	ADRA1A	cluster b
142	NM_019043	54518	APBB1IP	cluster b
143	NM_173800	206338	AQPEP	cluster b
144	NM_001145194	642597	C18orf42	cluster b
145	NM_001082533	56934	CA10	cluster b
146	NM_003716	8618	CADPS	cluster b
147	NM_080617	140689	CBLN4	cluster b
148	NM_001220490	1012	CDH13	cluster b
149	NM_004361	1005	CDH7	cluster b
150	NM_001796	1006	CDH8	cluster b

**SupplementaryTable S5: List of 180 hypermethylated genes in four clusters. (Con't)**

NO.	RefSeq	ENTREZ_ID	Gene Name	NMF cluster
151	NM_020986	1103	CHAT	cluster b
152	NM_175607	152330	CNTN4	cluster b
153	NM_130773	129684	CNTNAP5	cluster b
154	NM_000104	1545	CYP1B1	cluster b
155	NM_005215	1630	DCC	cluster b
156	NM_001039350	1804	DPP6	cluster b
157	NM_001277174	10278	EFS	cluster b
158	NM_001171195	1993	ELAVL2	cluster b
159	NM_000129	2162	F13A1	cluster b
160	NM_207426	399823	FOXI2	cluster b
161	NM_001204267	2557	GABRA4	cluster b
162	NM_001134366	2572	GAD2	cluster b
163	NM_000844	2917	GRM7	cluster b
164	NM_000863	3351	HTR1B	cluster b
165	NM_000217	3736	KCNA1	cluster b
166	NM_002235	3742	KCNA6	cluster b
167	NM_002239	3760	KCNJ3	cluster b
168	NM_001011705	11061	LECT1	cluster b
169	NM_001102659	200726	LOC200726	cluster b
170	NM_006158	4747	NEFL	cluster b
171	NM_001105541	4741	NEFM	cluster b
172	NM_005382	4741	NEFM	cluster b
173	NM_138966	81832	NETO1	cluster b
174	NM_001136271	137814	NKX2-6	cluster b
175	NM_000439	5122	PCSK1	cluster b
176	NM_001177875	5122	PCSK1	cluster b
177	NM_001135690	5179	PENK	cluster b
178	NM_022068	63895	PIEZO2	cluster b
179	NM_145040	112464	PRKCDBP	cluster b
180	NM_004248	2834	PRLHR	cluster b
181	NM_001142333	54715	RBFOX1	cluster b
182	NM_144682	146857	SLFN13	cluster b
183	NM_001287813	54212	SNTG1	cluster b
184	NM_014978	22986	SORCS3	cluster b
185	NM_030623	80309	SPHKAP	cluster b
186	NM_001130003	132204	SYNPR	cluster b
187	NM_000192	6910	TBX5	cluster b
188	NM_181486	6910	TBX5	cluster b
189	NM_174937	256536	TCERG1L	cluster b
190	NM_001136103	92293	TMEM132C	cluster b
191	NM_133448	121256	TMEM132D	cluster b
192	NM_198570	375567	VWC2	cluster b
193	NM_001168379	84107	ZIC4	cluster b
194	NM_001010879	284307	ZIK1	cluster b
195	NM_001134656	389114	ZNF662	cluster b
196	NM_207404	389114	ZNF662	cluster b

NMF, nonnegative matrix factorization.

**Supplementary Table S6.** Summary of the methylation levels of four candidate genes in cervical scrapings from endometrial cancer and noncarcinoma samples.

	Normal			Myoma			EC			<i>P</i>
	N	Median	Rang	N	Median	Rang	N	Median	Rang	
<i>BHLHE22</i>	55	1.72	0.04–33.31	40	3.64	0.25–182.62	49	121.74	1.77–13707.83	<10 <sup>-4</sup>
<i>CDO1</i>	56	6.27	0.06–280.67	40	9.45	0.30–259.16	50	378.15	6.60–6529.30	<10 <sup>-4</sup>
<i>CELF4</i>	49	3.86	0.31–154.63	40	8.09	0.05–53.73	50	81.87	9.11–3242.10	<10 <sup>-4</sup>
<i>ZNF662</i>	56	3717.80	2199.12–13240.89	40	4649.05	2932.09–14539.73	50	8570.91	3164.39–29690.47	<10 <sup>-4</sup>

EC, endometrial cancer; *P* values were calculated using the Kruskal-Wallis test. N, case number of available testing data using quantitative methylation PCR (QMSP).