Supplementary Information

Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes

Matthieu Le Gallo[1*], Andrea J. O'Hara[1*], Meghan L. Rudd[1], Mary Ellen Urick[1], Nancy F. Hansen[2], Nigel J. O'Neil[3], Jessica C. Price[1], Suiyuan Zhang[2], Bryant M. England[1], Andrew K. Godwin[4], Dennis C. Sgroi[5], NISC Comparative Sequencing Program[6], Philip Hieter[3], James C. Mullikin[2,6], Maria J. Merino[7], and Daphne W. Bell[1]

*Denotes equal contribution

[1] Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2] Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[3] Michael Smith Laboratories, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada
[4] Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66106, USA
[5] Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA
[6] NIH Intramural Sequencing Center (NISC), National Institutes of Health, Bethesda, MD 20892, USA
[7] National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Correspondence to: Dr. Daphne W. Bell, National Human Genome Research Institute, Cancer Genetics Branch, 50 South Drive, MSC-8000, Bethesda, MD 20892; Phone (301) 594-9256; Fax (301) 594-1360; Email belldaph@mail.nih.gov

Supplementary Note

NIH Intramural Sequencing Center (NISC)

Contributions from NISC were made by Betty Benjamin, Robert Blakesley, Gerry Bouffard, Shelise Brooks, Holly Coleman, Mila Dekhtyar, Michael Gregory, Xiaobin Guan, Jyoti Gupta, Joel Han, April Hargrove, Shi-ling Ho, Taccara Johnson, Richelle Legaspi, Sean Lovett, Quino Maduro, Cathy Masiello, Baishali Maskeri, Jenny McDowell, Casandra Montemayor, James Mullikin, Betsy Novotny, Morgan Park, Nancy Riebow, Karen Schandler, Brian Schmidt, Christina Sison, Mal Stantripop, James Thomas, Meg Vemulapalli, and Alice Young.

Clinical data

A determination was made by the NIH Office for Human Subjects Research Protections (OHSRP) that this research was not "human subjects research" per the Common Rule (45 CFR 46).

Genotype and variant calling

The MPG ("Most Probable Genotype") genotype caller takes as its input a set of sequence reads aligned to an orthologous reference sequence. These sequence reads are assumed to be (1) correctly aligned to the reference, and (2) derived from one or two copies of a modified reference sequence, as would be the case in re-sequencing reads from non-repetitive regions of a diploid individual. In addition, if base quality estimates are provided, these are assumed to be accurate reflections of the probability that the read sequence actually represents the sample. The MPG algorithm is implemented in a perl

script called “bam2mpg”, which reads BAM formatted files, and outputs predicted genotypes and scores at all regions of the genome covered by at least one read.

Any variation between a set of aligned sequence reads and the reference sequence can be specified by its flanking reference coordinates and a list of all of the observed sequences between them (which we will refer to as “alleles”). For example, a single nucleotide variant (SNV) can be specified by giving the coordinates to the left and right of the changed base, and all possible bases at that location (e.g., hg18 chr1:16,901,843-16,901,845, T/C). A deletion-insertion variant (DIV) can also be specified by giving the coordinates of the outer boundaries of the feature, with a list of all possible sequences seen between those two boundaries (e.g., hg18 chr1:145,232,693-145,232,694, -/AT represents an insertion of the sequence “AT” between the two coordinates listed). For variants predicted by MPG, the reference sequence is always included in the list of possible alleles.

As a simplification, we can assume that the read bases seen in a resequencing experiment are independently sequenced with equal probability from each of two chromosomes (in the case of autosomes, or one in cases where only one chromosome is present), and are then chosen from a random distribution in which sequencing errors occur with a total probability e_k. Given a group of aligned sequence reads at a variant site, we can calculate the posterior probability of all possible genotypes at that position using Bayes theorem. For a simple variant with only two possible alleles (the reference and a variant sequence), the genotypes considered are homozygous reference, heterozygous, and homozygous

non-reference. If we denote the posterior probability of one of M possible genotypes G_i, given N observed read sequences $A_1,\ldots, A_N$ as $P(G_i \mid A_1,\ldots, A_N)$, the likelihood of observing $A_1,\ldots,A_N$ as $P(A_1,\ldots,A_N \mid G_i)$, and the prior probability of a site having the genotype G_i as $P(G_i)$, then the desired posterior probabilities of the different genotypes can be calculated using the equation

$$P\left(G_i|A_1,\ldots,A_N\right) = \frac{P\left(A_1,\ldots,A_N|G_i\right)P(G_i)}{\sum_{j=1}^{M} P(A_1,\ldots,A_N|G_j)P(G_j)}$$

where

$$P\left(A_1,\ldots,A_N|G_j\right) = \prod_{k=1}^{N} P\left(A_k|G_j\right)$$

and

$$P(A_k|G_j) = (1-e_k)\left(\frac{1}{2}\left(\delta_{G_{j,1},A_k} + \delta_{G_{j,2},A_k}\right)\right) + e_k\left(\frac{1}{2}\left((1-\delta_{G_{j,1},A_k}) + (1-\delta_{G_{j,2},A_k})\right)\right).$$

The calculation is simplified by noting that for any two possible genotypes,

$$\log\left(\frac{P(G_i|A_1,\ldots,A_N)}{P(G_j|A_1,\ldots,A_N)}\right) = \sum_{k=1}^{N}\left\{\log\left(P\left(A_k|G_i\right)\right) - \log\left(P\left(A_k|G_j\right)\right)\right\} + \log\left(P\left(\frac{G_i}{G_j}\right)\right).$$

The last term in this last equation is the log of the ratio of the prior probabilities, and is assumed by MPG to be the same for all non-reference genotypes compared to the reference (and 0 for all other combinations of genotypes). This value can be passed as a parameter to bam2mpg, and its default value is 3, corresponding to a reference genotype that is expected to be 1000 times more probable than any non-reference genotype.

Once the most probable genotype is determined, it is reported by bam2mpg, along with its "MPG score", which is the value of $\log(P(G_i|A)/P(G_j|A))$ when G_i is the most probable genotype and G_j is the second most probable genotype. We have found empirically that when calculated from well-aligned Illumina reads, genotypes with MPG scores of 10 or

greater agree with Infineum genotypes about 99.8% of the time.

The “MPV”, or “Most Probable Variant”, score is an alternative scoring method which scores a site’s genotype call as the log ratio of posterior probabilities of the most probable genotype compared to the homozygous reference genotype (as opposed to the second most probable genotype, as above). This assesses the probability that any variant is present, and is more suitable when variant detection, rather than exact genotype prediction, is the goal.

Supplementary Figure 1
Number of somatic variants called by exome sequencing of 13 serous endometrial tumors in a mutation discovery screen

Supplementary Figure 2
Distribution of 1,544 exonic and splice junction somatic mutations identified by exome sequencing of 13 primary serous endometrial tumors

Supplementary Figure 3
Sequence chromatograms showing somatic mutations in *CHD4*, identified in primary endometrial cancers

Supplementary Figure 4
Somatic mutations in *MAP3K4*, *ABCC9*, and *CYP4X1* in serous endometrial cancers

Supplementary Figure 5
Distribution of somatic mutations in *CHD4*, *FBXW7*, and *SPOP* relative to the microsatellite instability and *MSH6* mutation status

Supplementary Figure 6
CHD4 expression in endometrial cancer cell lines

Supplementary Figure 7
Cross-species conservation of CHD4

Supplementary Figure 8
Alignment of CHD4 and closely related family members

Supplementary Figure 9
Partial, structure-based, sequence alignment of the ATPase/helicase and helicase domains of CHD4 to the Snf2 domains of SMARCA2, SMARCAL1, and SMARCA4

Supplementary Figure 10
Distribution of somatic variants called by whole exome sequencing according to score and coverage

Supplementary Table 2
Depth of whole exome sequencing coverage

Supplementary Table 3
Exonic somatic mutation rate of the 13 serous endometrial tumors included in the exome discovery screen

Supplementary Table 6
Distribution of somatic variants called by whole exome sequencing among 12 tumors in the discovery screen

Supplementary Table 7
Nine genes with validated mutations in more than one tumor in the discovery screen

Supplementary Table 8
Frequency of *TP53*, PIK3CA, and PPP2R1A among serous endometrial tumors

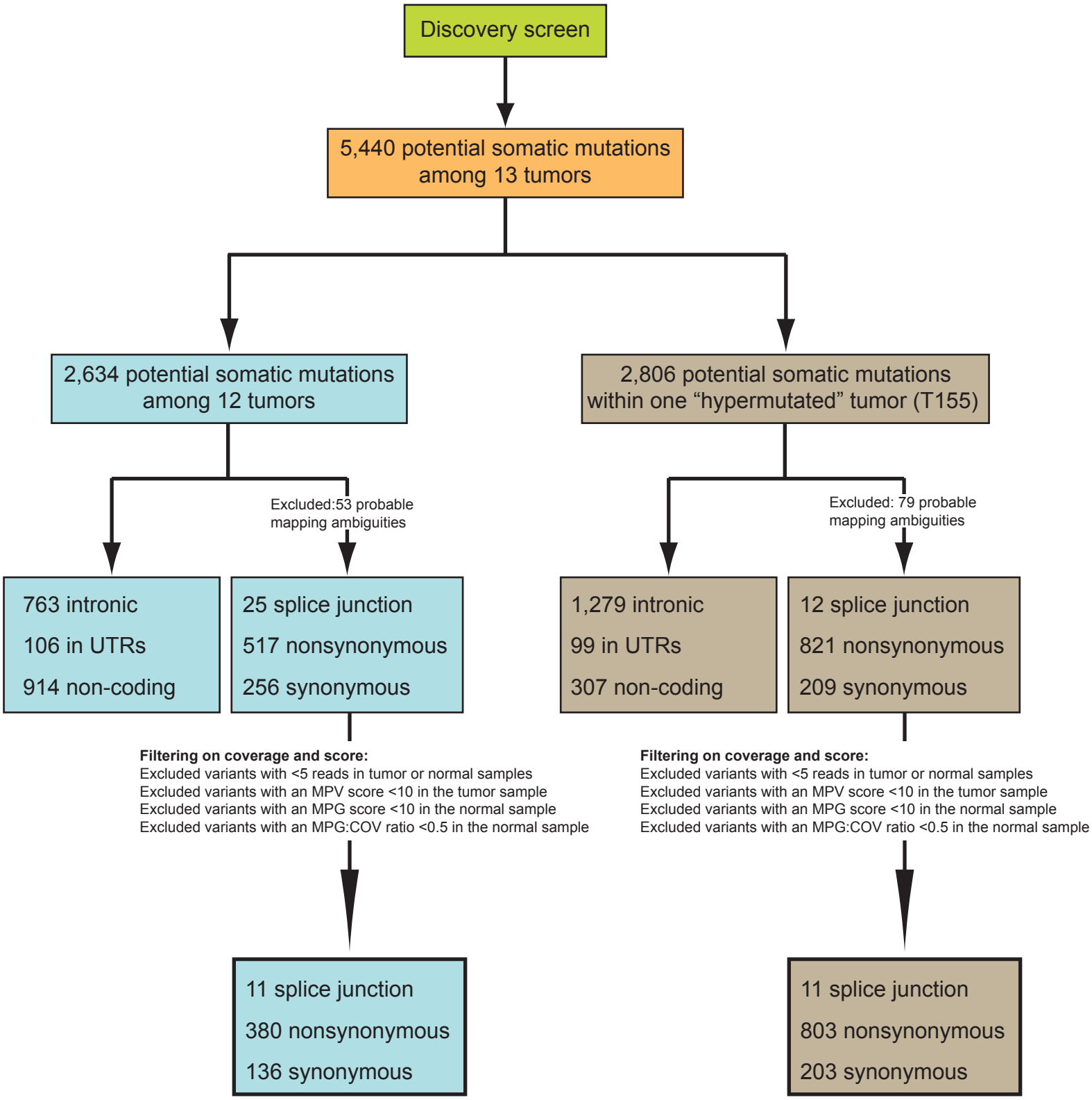
Supplementary Table 9
Overall frequency of somatic mutations in combined discovery and prevalence screens of *CHD4, SPOP, FBXW7, ABCC9, CYP4X1, MAP3K4, TP53, PIK3CA* and *PPP2R1A*

Supplementary Table 10
Mutation rate of individual genes in 52 serous tumors, compared to the background mutation rate for serous endometrial cancers

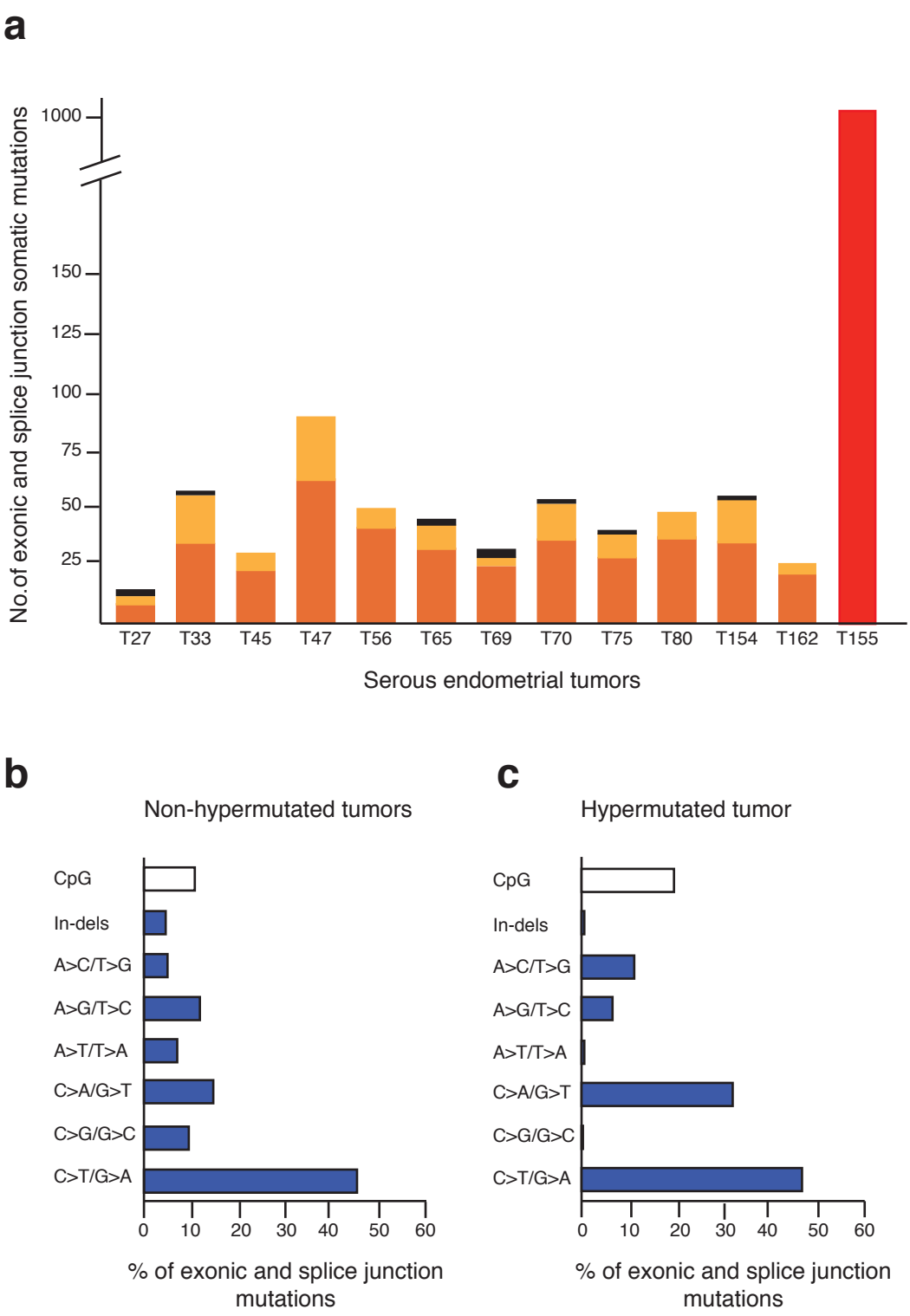
Supplementary Table 12
MSI status and *MSH6* status of tumors with *CHD4*, *FBXW7*, and *SPOP* mutations

Supplementary Table 14
Networks identified by Ingenuity Pathway Analysis

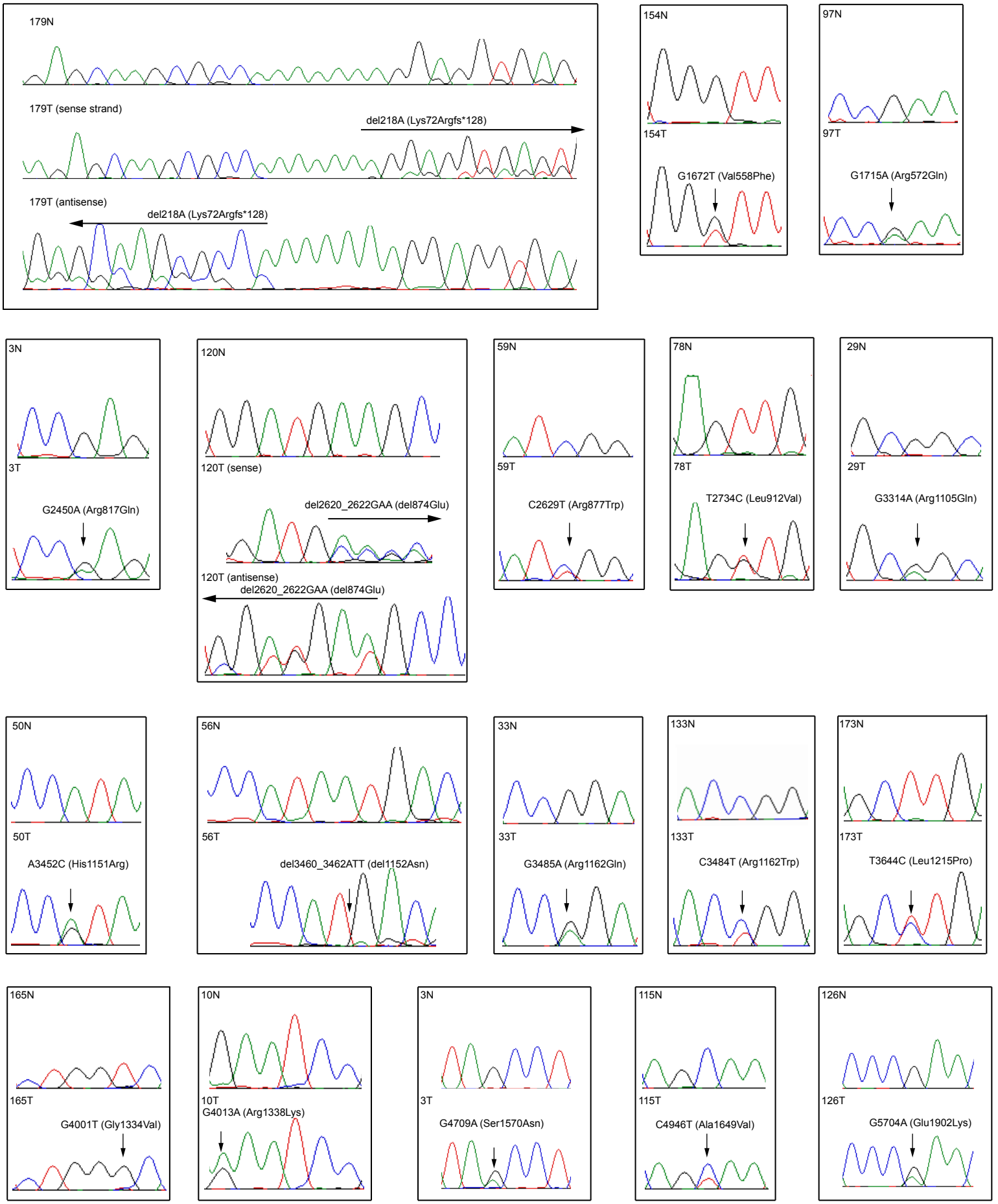
Supplementary Table 15
Estimated power to detect somatically mutated genes in a discovery screen of 12 tumors



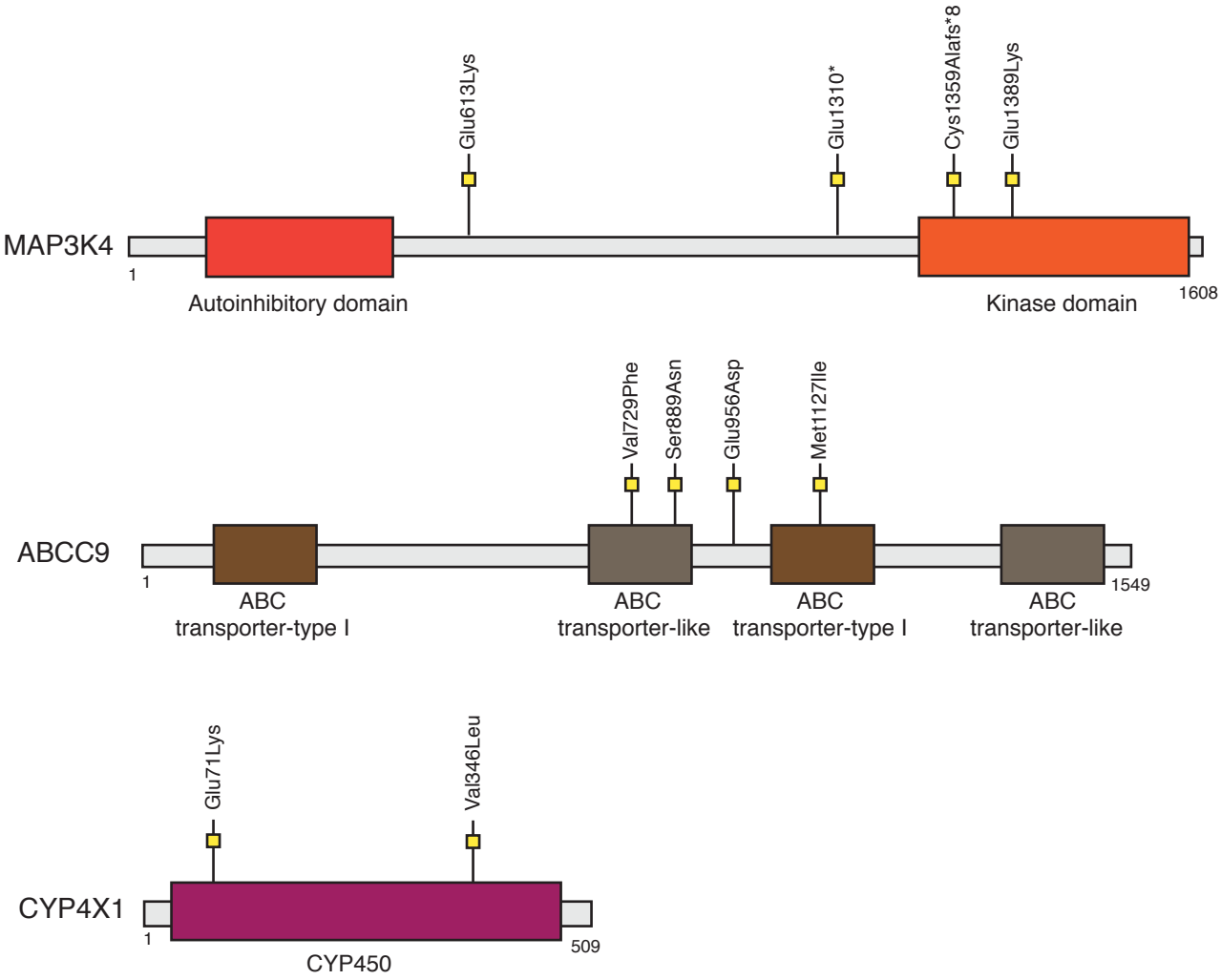
Supplementary Figure 1. Number of somatic variants called by exome sequencing of 13 serous endometrial tumors in a mutation discovery screen. After filtering on coverage and score, there were 1,522 exonic (nonsynonymous and synonymous) somatic variants and 22 somatic splice junction variants among 13 tumors. One tumor (T155) appeared to be hypermutated.



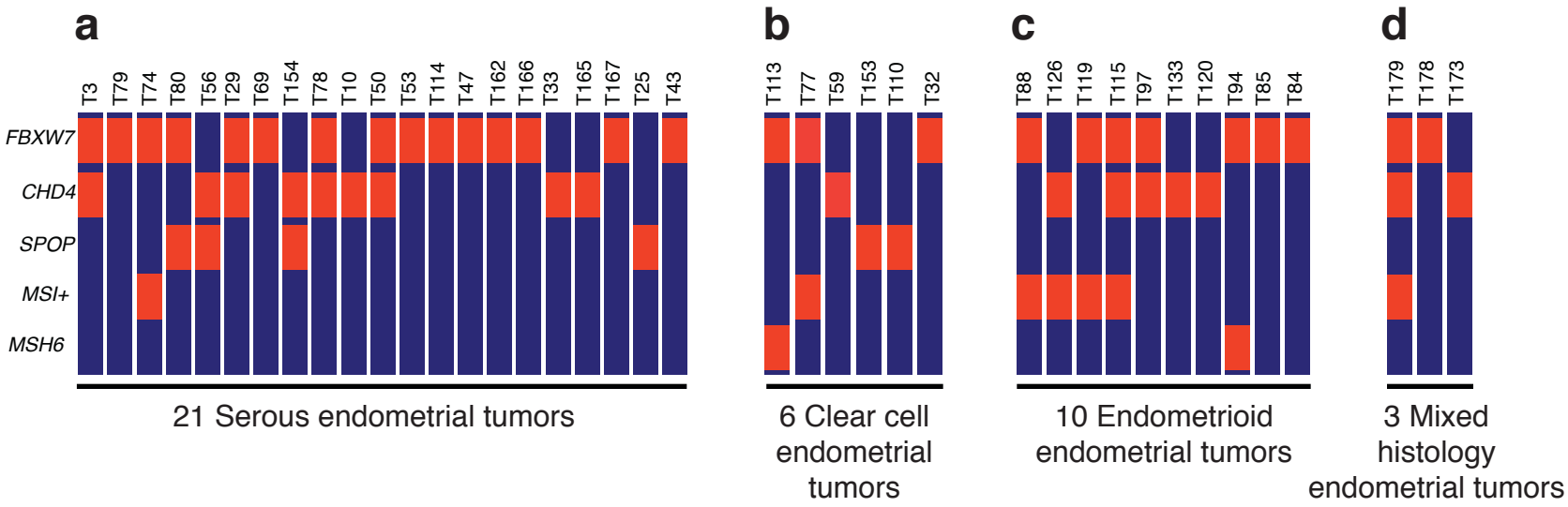
Supplementary Figure 2. Distribution of 1,544 exonic and splice junction somatic mutations identified by exome sequencing of 13 primary serous endometrial tumors. (**a**) Number of nonsynonymous mutations (dark orange), synonymous mutations (light orange), and splice junction mutations (black) in each tumor. T155 (red) is hypermutated, relative to the other 12 tumors. (**b**) Mutation signatures of the 12 non-hypermutated tumors, and (**c**) Mutation signature of T155, a hypermutated tumor.The percentage of mutations that were insertions-deletions (in-dels), transitions,or transversions is displayed. The percentage of nucleotide substitutions that occurred within the context of a CpG dinucleotide, relative to all exonic/splice junction mutations, is indicated.



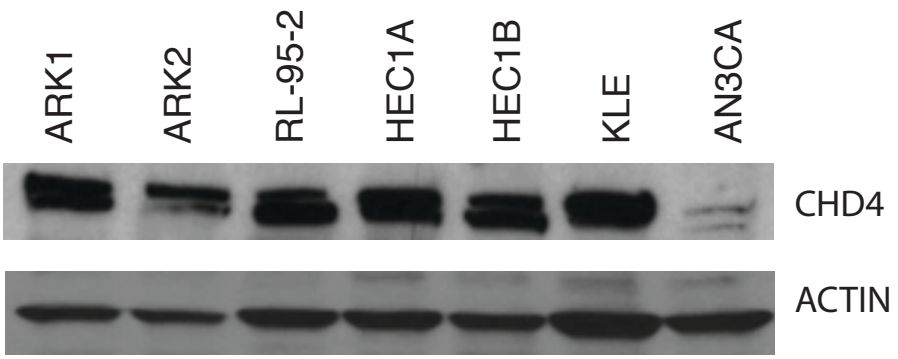
Supplementary Figure 3. Sequence chromatograms showing somatic mutations in *CHD4*, identified in primary endometrial cancers. Mutated nucleotide positions in the tumor (T) DNA are indicated by arrows. Nucleotide sequences of the matched normal (N) DNAs are shown for reference.



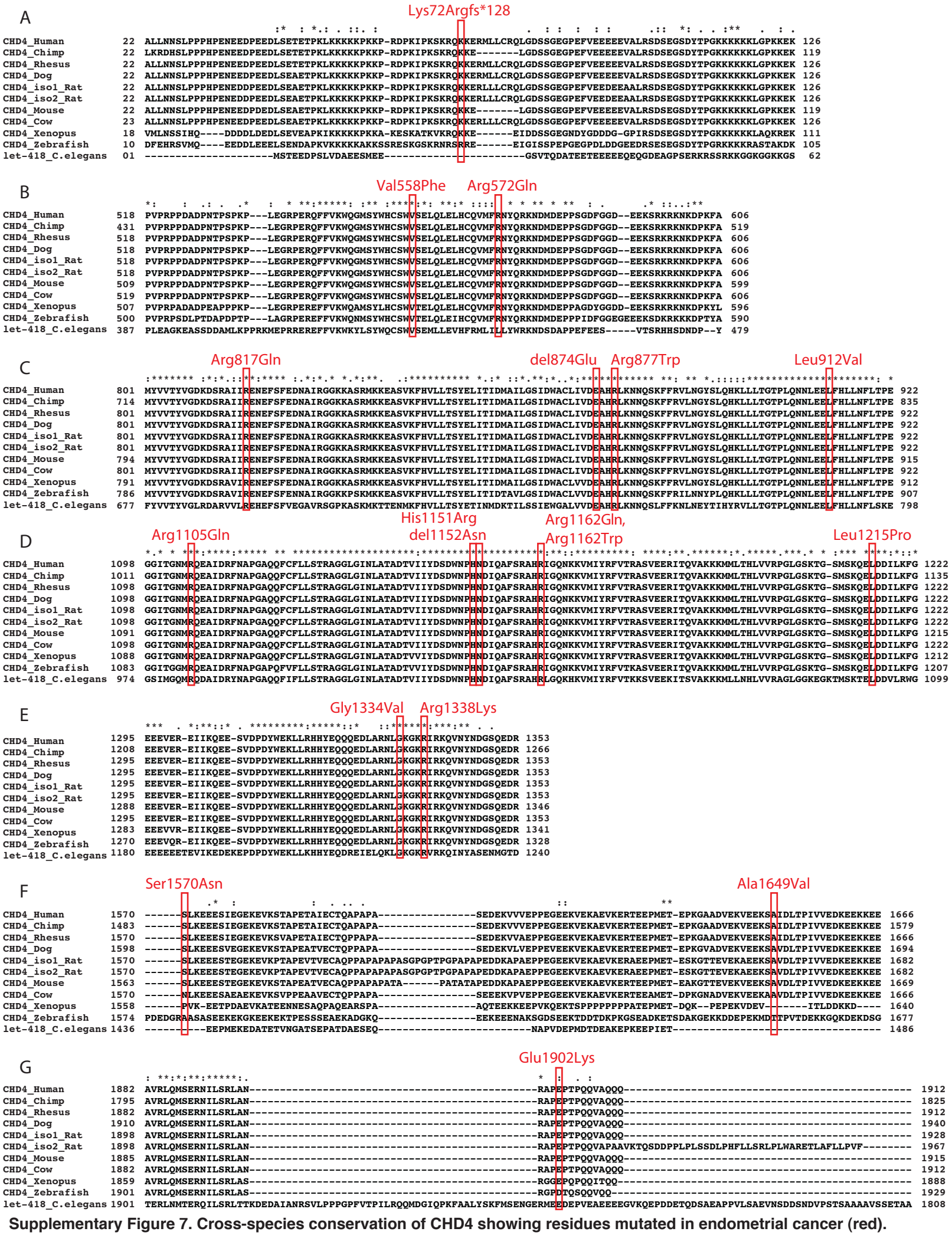
Supplementary Figure 4. Somatic mutations in MAP3K4, ABCC9, and CYP4X1 in serous endometrial cancers. Schematic representation of the MAP3K4, ABCC9, and CYP4X1 proteins showing the relative positions of individual somatic mutations identified among primary serous endometrial tumors.



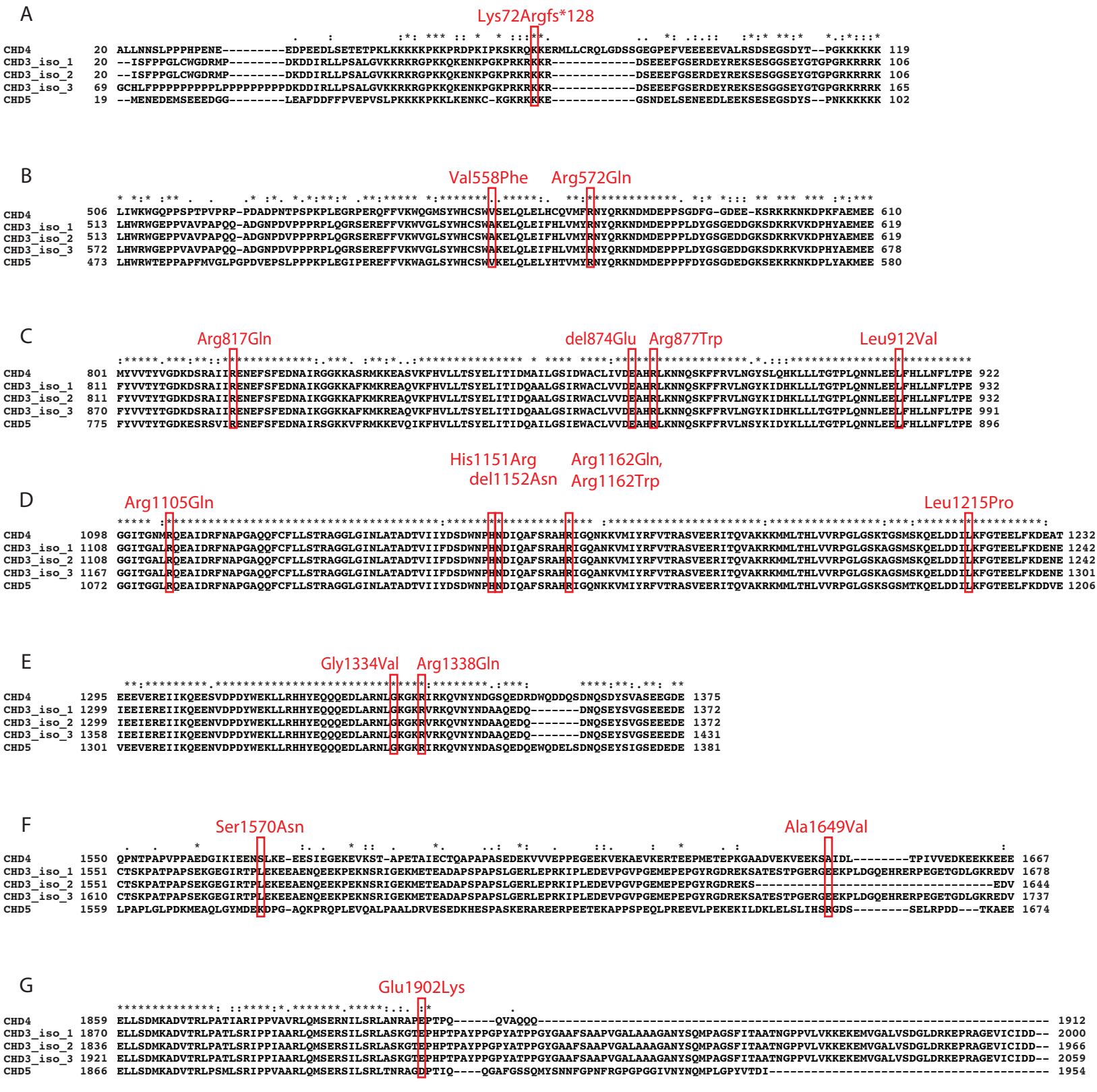
Supplementary Figure 5. Oncoprints displaying the distribution of somatic mutations in *CHD4*, *FBXW7*, and *SPOP* relative to the microsatellite instability (MSI) status, and the mutational status of *MSH6*, among (a) serous, (b) clear cell, (c) endometrioid, and (d) mixed histology endometrial tumors. Each blue bar represents an individual tumor (T). Nonsynonymous somatic mutations and MSI are indicated by the red bars. Only tumors that had somatically mutated *CHD4*, *FBXW7*, or *SPOP* are displayed.



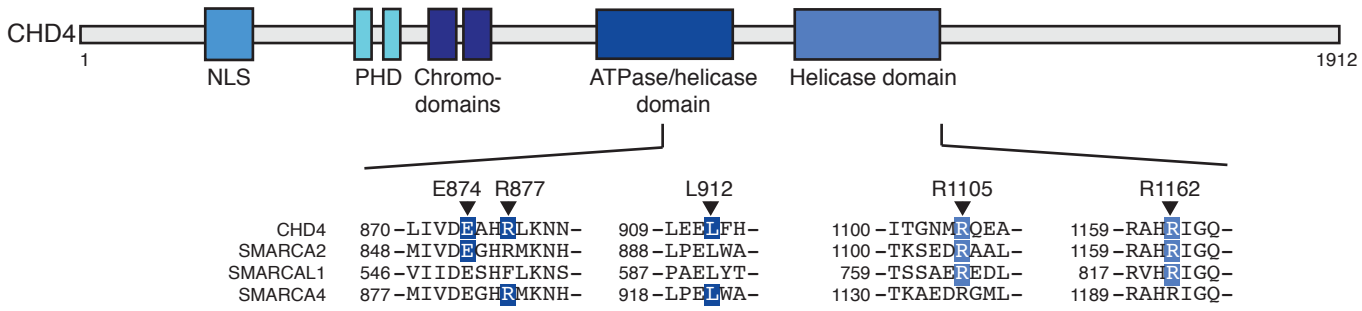
Supplementary Figure 6. CHD4 is endogenously expressed in endometrial cancer cell lines. Western blots showing CHD4 levels (top) and actin levels (bottom).



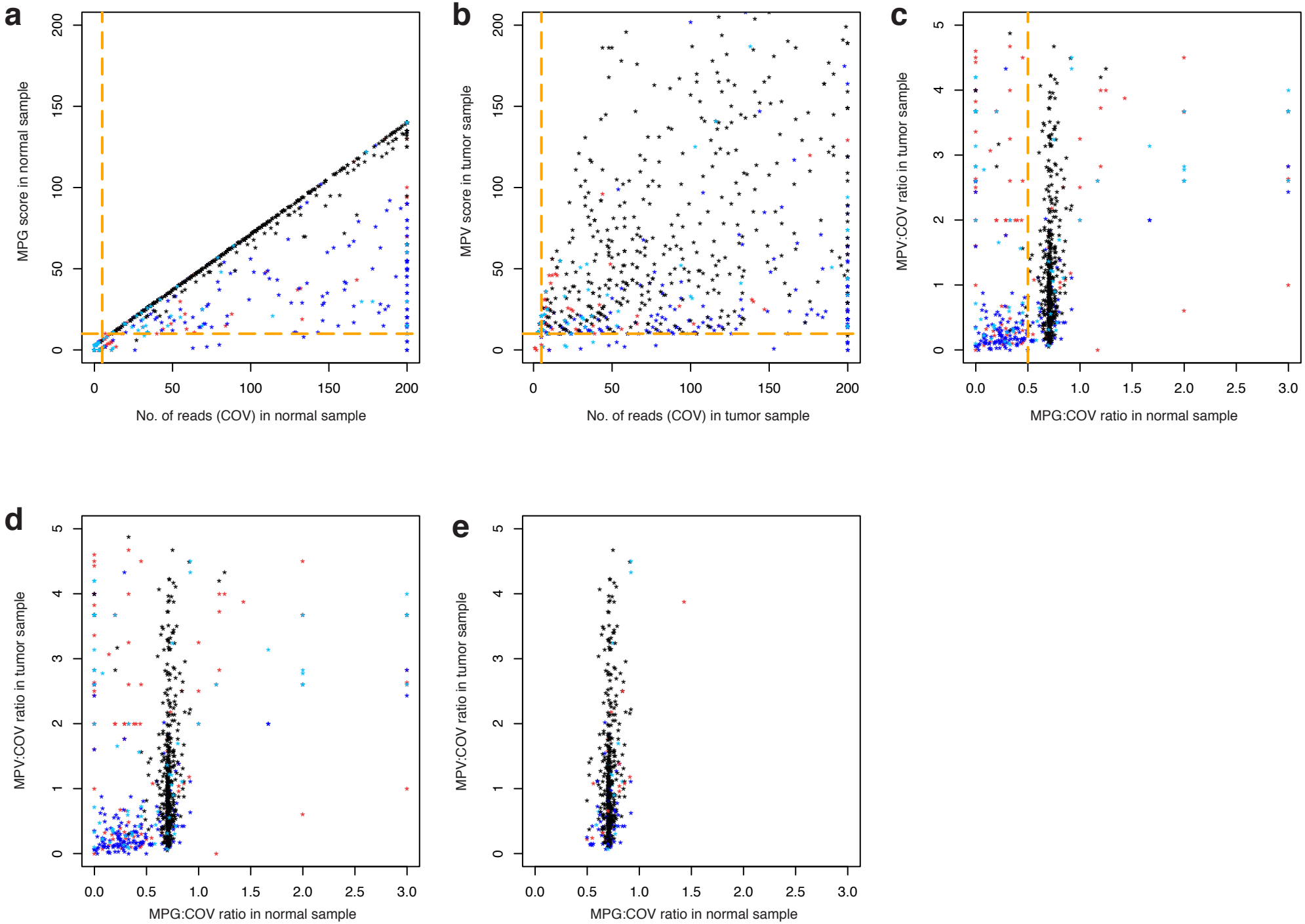
Supplementary Figure 7. Cross-species conservation of CHD4 showing residues mutated in endometrial cancer (red).



Supplementary Figure 8. Alignment of CHD4 and closely related family members. CHD4 residues mutated in endometrial cancer are indicated in red.



Supplementary Figure 9. Partial, structure-based, sequence alignment of the ATPase/helicase and helicase domains of CHD4 to the Snf2 domains of SMARCA2, SMARCAL1, and SMARCA4. Five resides (blue boxes) that are somatically mutated in CHD4 in endometrial cancer, are conserved in SMARCA2, SMARCAL1, and SMARCA4 and undergo pathogenic germline mutations in patients with Nicolaides-Baraister syndrome (SMARCA2), Schimke immuno-osseous dysplasia (SMARCAL1), and Coffin-Siris syndrome (SMARCA4) (references 20-22).



Supplementary Figure 10. **Distribution of somatic variants called by whole exome sequencing (WES) according to score and coverage.** (**a-c**) Dot plots showing thresholds (orange lines) used to filter the calls to achieve a balance between specificity and sensitivity of somatic mutation detection. Dot plots display variants called as somatic by WES, according to their validation status by Sanger sequencing before (**d**) and after (**e**) filtering on score and coverage. Each point represents an individual variant; the validation status of variants is somatic (black), germline (red), not evaluated (light blue), or not detectable in the tumor (dark blue).

Supplementary Table 2: Depth of whole exome sequencing coverage

Sample	No. of bases targeted	No. of bases covered	% of targeted bases covered	Average depth of coverage for aligned reads	No. of targeted bases with sufficient coverage and quality for variant calling	% of targeted bases with sufficient coverage and quality for variant calling
Tumor samples						
T27	37,640,396	37,041,467	98.41%	100.8	35,861,300	91.69%
T33	51,499,639	49,957,360	97.01%	150.4	48,506,495	91.33%
T45	37,640,396	36,951,637	98.17%	105.1	35,684,195	90.73%
T47	51,499,639	49,876,467	96.85%	94.8	48,142,085	89.19%
T56	51,499,639	49,879,322	96.85%	120.2	48,411,478	89.69%
T65	51,499,639	49,730,719	96.57%	75.4	48,505,619	87.06%
T69	51,499,639	49,892,083	96.88%	119.2	48,444,093	89.71%
T70	51,499,639	49,785,681	96.67%	74.8	47,969,296	87.07%
T75	51,499,639	49,733,051	96.57%	69.8	47,526,230	86.89%
T80	37,640,396	36,722,341	97.56%	70.8	35,832,195	85.89%
T154	51,499,639	49,935,941	96.96%	126.6	48,398,890	90.34%
T155	51,499,639	49,903,701	96.90%	128.5	48,066,638	89.87%
T162	51,499,639	49,951,052	96.99%	101.9	48,173,519	89.11%
Matched normal samples						
N27	37,640,396	36,984,288	98.26%	102.5	36,035,725	91.15%
N33	51,499,639	50,040,094	97.17%	124.4	48,552,347	90.79%
N45	37,640,396	36,952,879	98.17%	81.4	35,757,557	89.81%
N47	51,499,639	49,903,532	96.90%	95.2	47,995,032	89.60%
N56	51,499,639	49,953,739	97.00%	117.5	48,103,183	90.63%
N65	51,499,639	50,067,103	97.22%	72.4	47,414,952	89.91%
N69	51,499,639	50,036,195	97.16%	127.4	48,131,346	90.64%
N70	51,499,639	49,900,106	96.89%	80.6	47,435,488	88.89%
N75	51,499,639	49,746,618	96.60%	69.5	47,379,192	87.47%
N80	37,640,396	36,968,802	98.22%	117	34,836,598	91.21%
N154	51,499,639	49,968,334	97.03%	116.2	48,323,669	90.51%
N155	51,499,639	49,892,810	96.88%	108.9	48,121,062	89.54%
N162	51,499,639	50,010,023	97.11%	105.5	47,950,781	89.82%
Weighted average of all tumors			97.13%	102.6		89.53%

Supplementary Table 3: Exonic somatic mutation rate of the 13 serous endometrial tumors included in the exome discovery screen

Tumor ID	No. of exonic synomymous somatic mutations	No. of exonic non synonymous somatic mutations	Total no of exonic somatic mutations	No. of covered bp within coding exons	Mutation rate	Grubbs' Test for outliers approximate P value
T27	3	8	11	26,679,401	4.123.E-07	9.079
T33	20	34	54	30,495,201	1.771.E-06	10.558
T45	7	22	29	26,319,267	1.102.E-06	9.822
T47	27	62	89	29,912,229	2.975.E-06	11.909
T56	8	41	49	30,175,714	1.624.E-06	10.395
T65	10	32	42	29,351,564	1.431.E-06	10.183
T69	3	24	27	30,162,216	8.952.E-07	9.598
T70	15	36	51	29,345,612	1.738.E-06	10.522
T75	10	28	38	29,122,936	1.305.E-06	10.044
T80	11	37	48	25,471,115	1.884.E-06	10.684
T154	18	35	53	30,266,338	1.751.E-06	10.536
T162	4	21	25	29,980,154	8.339.E-07	9.531
T155	203	803	1006	30,070,801	3.345.E-05	6.59E-13
All Tumors	**339**	**1183**	**1522**	**377,352,548**	**4.033.E-06**	**Not applicable**
Non Hypermutated Tumors	**136**	**380**	**516**	**347,281,747**	**1.486.E-06**	**Not applicable**

Supplementary Table 6: Distribution of somatic variants called by whole exome sequencing among 12 tumors in the discovery screen

Tumor	No. of variants called by exome sequencing (No. of variants confirmed by Sanger sequencing)			
	Exonic variants		Splice Junction variants	Total
	Synonymous	Nonsynonymous		
T27	3 (1)	8 (5)	2 (0)	13 (6)
T33	20 (11)	34 (32)	1 (1)	55 (44)
T45	7 (4)	22 (18)	0 (0)	29 (22)
T47	27 (24)	62 (55)	0 (0)	89 (79)
T56	8 (6)	41 (31)	0 (0)	49 (37)
T65	10 (7)	32 (28)	2 (2)	44 (37)
T69	3 (3)	24 (21)	3 (3)	30 (27)
T70	15 (13)	36 (31)	1 (1)	52 (45)
T75	10 (10)	28 (26)	1 (1)	39 (37)
T80	11 (11)	37 (32)	0 (0)	48 (43)
T154	18 (16)	35 (27)	1 (1)	54 (44)
T162	4 (3)	21 (15)	0 (0)	25 (18)
Total	136 (109)	380 (321)	11 (9)	527 (439)
Average no. of variants/tumor	11.3 (9.1)	31.7 (26.8)	0.9 (0.8)	43.9 (36.6)
Average no. of nonsynonymous and splice junction variants/tumor		32.6 (27.5)		

Supplementary Table 7: Nine genes with validated mutations in more than one tumor in the discovery screen

Gene Name	Transcript Accession Id	Tumor	Nucleotide Change	Amino Acid Change	Mutation Assessor Prediction
TP53	uc002gij.2	T47	c.1009C>T	p.Arg337Cys	Medium
TP53	uc002gij.2	T33	c.1025G>C	p.Arg342Pro	Medium
TP53	uc002gij.2	T69	c.614A>G	p.Tyr205Cys	High
TP53	uc002gij.2	T56	c.309C>A	p.Tyr103*	-
TP53	uc002gij.2	T154	c.1045G>T	p.Glu349*	-
TP53	uc002gij.2	T162	c.814G>C	p.Val272Leu	Medium
TP53	uc002gij.2	T65	c.476C>T	p.Ala159Val	Medium
TP53	uc002gij.2	T65	c.475G>T	p.Ala159Ser	High
TP53	uc002gij.2	T70	c.742C>T	p.Arg248Trp	High
TP53	uc002gij.2	T80	c.422G>A	p.Cys141Tyr	High
PIK3CA	uc003fjk.1	T80	c.263G>A	p.Arg88Gln	Medium
PIK3CA	uc003fjk.1	T75	c.3172A>T	p.Ile1058Phe	Low
PIK3CA	uc003fjk.1	T75	c.3207A>G	p.*1069insWKDN	-
PIK3CA	uc003fjk.1	T69	c.3132T>A	p.Asn1044Lys	Low
FBXW7	uc003ims.1	T80	c.1385C>T	p.Ser462Phe	Medium
FBXW7	uc003ims.1	T80	c.1322G>C	p.Arg441Pro	Medium
FBXW7	uc003ims.1	T47	c.1436G>A	p.Arg479Gln	Medium
FBXW7	uc003ims.1	T162	c.1436G>A	p.Arg479Gln	Medium
FBXW7	uc003ims.1	T69	c.2065C>T	p.Arg689Trp	Low
CHD4	uc001qpo.1	T33	c.3485G>A	p.Arg1162Gln	High
CHD4	uc001qpo.1	T154	c.1672G>T	p.Val558Phe	Medium
CHD4	uc001qpo.1	T56	c.3460_3462delATT	p.Asn1152del	-
SPOP	uc002ipg.1	T56	c.240C>G	p.Ser80Arg	Medium
SPOP	uc002ipg.1	T80	c.280C>G	p.Pro94Ala	Low
SPOP	uc002ipg.1	T154	c.362G>A	p.Arg121Gln	Medium
PPP2R1A	uc002pyp.1	T70	c.771G>T	p.Trp257Cys	Medium
PPP2R1A	uc002pyp.1	T65	c.1757T>C	p.Leu586Pro	Medium
PPP2R1A	uc002pyp.1	T154	c.767C>T	p.Ser256Phe	Medium
PPP2R1A	uc002pyp.1	T69	c.536C>G	p.Pro179Arg	Medium
MAP3K4	uc003qtq.1	T80	c.4165G>A	p.Glu1389Lys	High
MAP3K4	uc003qtq.1	T65	c.4077delC	p.Cys1359Alafs*8	-
ABCC9	uc001rfh.1	T70	c.2185G>T	p.Val729Phe	High
ABCC9	uc001rfh.1	T75	c.3381G>A	p.Met1127Ile	Low
CYP4X1	uc001cqr.1	T75	c.1036G>C	p.Val346Leu	Low
CYP4X1	uc001cqr.1	T56	c.211G>A	p.Glu71Lys	Low

(-) Not evaluated by Mutation Assessor

Supplementary Table 8: Frequency of TP53, PIK3CA, and PPP2R1A among serous endometrial tumors

Gene Name (RefSeq ID; UCSC Transcript Accession ID^)	Mutation frequency in serous tumors	Tumor	Histology	Nucleotide Change	Amino Acid Change	Mutation type
TP53 (NM_001126114; uc002gij.2)	71% (37/52)	T164	Serous	c.251delC	p.Pro84Leufs*38	Deletion
		T56	Serous	c.309C>A	p.Tyr103*	Nonsense
		T52	Serous	c.377A>G	p.Tyr126Cys	Missense
		T45	Serous	c.395A>G	p.Lys132Arg	Missense
		T50	Serous	c.399_400insTT	p.Met133_Phe134insFfs*35	Insertion
		T10 (OM-2009-C1)	Serous	c.406C>G	p.Gln136Glu	Missense
		T80	Serous	c.422G>A	p.Cys141Tyr	Missense
		T41	Serous	c.434T>C	p.Leu145Pro	Missense
		T65	Serous	c.475G>T	p.Ala159Phe	Missense
		T65	Serous	c.476C>T	p.Ala159Phe	Missense
		T29	Serous	c.536A>G	p.His179Arg	Missense
		T165	Serous	c.538G>T	p.Glu180*	Nonsense
		T69	Serous	c.614A>G	p.Tyr205Cys	Missense
		T55	Serous	c.637C>T	p.Arg213*	Nonsense
		T163	Serous	c.641A>G	p.His214Arg	Missense
		T83	Serous	c.722C>G	p.S241Cys	Missense
		T185	Serous	c.722C>T	p.S241Phe	Missense
		T75	Serous	c.731G>T	p.G244Val	Missense
		T71	Serous	c.733G>A	p.G245Ser	Missense
		T78	Serous	c.733G>A	p.G245Ser	Missense
		T76	Serous	c.734G>A	p.G245Asp	Missense
		T51	Serous	c.739delA	p.Asn247Thrfs*98	Deletion
		T23	Serous	c.742C>G	p.Arg248Trp	Missense
		T66	Serous	c.742C>T	p.Arg248Trp	Missense
		T70	Serous	c.742C>T	p.Arg248Trp	Missense
		T8 (OM-2511-C1)	Serous	c.743G>A	p.Arg248Gln	Missense
		T68	Serous	c.763A>T	p.Ile255Phe	Missense
		T112	Serous	c.764T>C	p.Ile255Thr	Missense
		T162	Serous	c.814G>C	p.Val272Leu	Missense
		T114	Serous	c.817C>T	p.Arg273Cys	Missense
		T167	Serous	c.818G>A	p.Arg273His	Missense
		T185	Serous	c.818G>A	p.Arg273His	Missense
		T30	Serous	c.824G>T	p.Cys275Phe	Missense
		T166	Serous	c.830G>T	p.Cys277Phe	Missense
		T53	Serous	c.841G>A	p.Asp281Asn	Missense
		T107	Serous	c.841G>C	p.Asp281His	Missense
		T75	Serous	c.844C>T	p.Arg282Trp	Missense
		T47	Serous	c.1009C>T	p.Arg337Cys	Missense
		T33	Serous	c.1025G>C	p.Arg342Pro	Missense
		T154	Serous	c.1045G>T	p.Glu349*	Nonsense
PIK3CA (NM_006218.1; uc003fjk.1) ‡	31% (16/52)	T75	Serous	c.3207A>G	p.*1069_*1069insTrpLysAspAsn*	Insertion
		T3 (OM-1323-C1)	Serous	c.241G>A	p.Glu81Lys	Missense
		T10 (OM-2009-C1)	Serous	c.1634A>C	p.Glu545Ala	Missense
		T41	Serous	c.42_64delinsTCCAA	p.Leu15_Val22insProIle	Insertion
		T29	Serous	c.1624G>A	p.Glu542Lys	Missense
		T49	Serous	c.1624G>C	p.Glu542Gln	Missense
		T53	Serous	c.1637A>C	p.Gln546Lys	Missense
		T68	Serous	c.263G>A	p.Arg88Gln	Missense
		T68	Serous	c.323G>A	p.Arg108His	Missense
		T69	Serous	c.3132T>A	p.Asn1044Lys	Missense
		T71	Serous	c.1093G>A	p.Glu365Lys	Missense
		T74	Serous	c.3139C>T	p.H1047Tyr	Missense
		T74	Serous	c.278G>A	p.Arg93Gln	Missense
		T74	Serous	c.333G>T	p.Lys111Asn	Missense
		T75	Serous	c.3172A>T	p.Ile1058Phe	Missense
		T76	Serous	c.1357G>A	p.Glu453Lys	Missense
		T78	Serous	c.331A>G	p.Lys111Glu	Missense
		T79	Serous	c.3073A>G	p.Thr1025Ala	Missense
		T80	Serous	c.263G>A	p.Arg88Gln	Missense
		T81	Serous	c.3140A>G	p.H1047Arg	Missense
PPP2R1A (NM_014225; uc002pyp.1)	25% (13/52)	T51	Serous	c.536C>G	p.Pro179Arg	Missense
		T69	Serous	c.536C>G	p.Pro179Arg	Missense
		T83	Serous	c.536C>G	p.Pro179Arg	Missense
		T107	Serous	c.536C>G	p.Pro179Arg	Missense
		T164	Serous	c.536C>G	p.Pro179Arg	Missense
		T50	Serous	c.767C>G	p.Ser256Phe	Missense
		T108	Serous	c.767C>G	p.Ser256Phe	Missense
		T154	Serous	c.767C>G	p.Ser256Phe	Missense
		T76	Serous	c.770C>G	p.Trp257Ser	Missense
		T23	Serous	c.771C>G	p.Trp257Cys	Missense
		T55	Serous	c.771C>G	p.Trp257Cys	Missense
		T70	Serous	c.771C>G	p.Trp257Cys	Missense
		T65	Serous	c.1757T>C	p.Leu586Pro	Missense

^UCSC transcript IDs are based on the hg18 assembly of the human genome sequence

‡ Previously reported in Rudd *et. al.* (ref 6)

Supplementary Table 9: Overall frequency of somatic mutations in combined discovery and prevalence screens of *CHD4, SPOP, FBXW7, ABCC9, CYP4X1, MAP3K4, TP53, PIK3CA* and *PPP2R1A*

Gene	Histological subtype						
	Serous	Clear cell	Endometrioid All grades	Endometrioid Grade 1 or Grade 2	Endometrioid Grade 2/3	Endometrioid Grade 3	Mixed Histology
CHD4	17% (9/52)	4% (1/23)	7% (5/67)	11% (5/46)	0% (0/4)	0% (0/17)	11% (2/18)
FBXW7	29% (15/52)	13% (3/23)	10% (7/67)	15% (7/46)	0% (0/4)	0% (0/17)	11% (2/18)
SPOP	8% (4/52)	9% (2/23)	0% (0/67)	0% (0/46)	0% (0/4)	0% (0/17)	0% (0/18)
CHD4/FBXW7/SPOP	40% (21/52)	26% (6/23)	15% (10/67)	22% (10/46)	0% (0/4)	0% (0/17)	17% (3/18)
ABCC9	6% (3/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
CYP4X1	4% (2/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
MAP3K4	6% (3/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
TP53	71% (37/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
PIK3CA	31% (16/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
PPP2R1A	25% (13/52)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

Supplementary Table 10: Mutation rate of individual genes in 52 serous tumors, compared to thecalculated background mutation rate for serous endometrial cancers

				Average background mutation rate in 12 discovery screen tumors (excluding T155)						Average background mutation rate all 13 discovery screen tumors					
Gene Name	**No. of covered bp within coding sequence**	**No. of somatic mutations**	**Number of genes assayed in discovery screen**	**Background mutation rate calculated from discovery screen data**	**Expected number of mutations based on background mutation rate**	**Poisson P-value (one-tailed) for observed number of mutations**	**Rank of P-value**	**q-value (False Discovery Rate)**	**CaMP-like Score (-log_10(q))**	**Background mutation rate calculated from discovery screen data**	**Expected number of mutations based on background mutation rate**	**Poisson P-value (one-tailed) for observed number of mutations**	**Rank of P-value**	**q-value (False Discovery Rate)**	**CaMP-like Score (-log_10(q))**
FBXW7	2121	20	21441	1.486E-06	0.1638746	5.3515E-37	1	1.15E-32	31.940	4.033E-06	0.4448477	5.2452E-28	1	1.12E-23	22.949
CHD4	5244	10	21441	1.486E-06	0.4051667	8.3511E-13	2	8.95E-09	8.048	4.033E-06	1.0998498	2.6138E-08	2	2.80E-04	3.553
SPOP	1122	4	21441	1.486E-06	0.0866890	3.7957E-08	3	2.71E-04	3.567	4.033E-06	0.2353226	4.9455E-06	3	3.53E-02	1.452
TSPYL2	1725	3	21441	1.486E-06	0.1332785	1.1820E-05	4	6.34E-02	1.198	4.033E-06	0.3617927	5.3543E-04	4	2.87	-0.458
ABCC9	4647	4	21441	1.486E-06	0.3590408	3.6912E-05	5	1.58E-01	0.801	4.033E-06	0.9746381	3.2857E-03	5	14.09	-1.149
MAP3K4	4672	4	21441	1.486E-06	0.3609723	3.7855E-05	6	1.35E-01	0.869	4.033E-06	0.9798815	3.3607E-03	6	12.01	-1.080
KDM4B/JMJD2B	2973	3	21441	1.486E-06	0.2297026	9.6595E-05	7	2.96E-01	0.529	4.033E-06	0.6235419	3.8454E-03	7	11.78	-1.071
CYP4X1	1527	2	21441	1.486E-06	0.1179805	2.5059E-04	9	5.97E-01	0.224	4.033E-06	0.3202652	4.3143E-03	8	11.56	-1.063
TRIM16	1692	2	21441	1.486E-06	0.1307289	3.3769E-04	11	6.58E-01	0.182	4.033E-06	0.3548715	5.7216E-03	9	13.63	-1.135
ARID1A	5718	4	21441	1.486E-06	0.4417893	9.7245E-05	8	2.61E-01	0.584	4.033E-06	1.1992642	7.7267E-03	10	16.57	-1.219
YEATS4	681	1	21441	1.486E-06	0.0526160	1.3366E-03	12	2.39	-0.378	4.033E-06	0.1428295	9.2790E-03	11	18.09	-1.257
EP300	7148	4	21441	1.486E-06	0.5522753	2.7107E-04	10	5.81E-01	0.236	4.033E-06	1.4991851	1.8538E-02	12	33.12	-1.520
BAZIB	4452	2	21441	1.486E-06	0.3439745	5.2525E-03	13	8.66	-0.938	4.033E-06	0.9337398	6.8526E-02	13	113.02	-2.053
CTCF	2181	1	21441	1.486E-06	0.1685104	1.2699E-02	14	19.45	-1.289	4.033E-06	0.4574318	7.7581E-02	14	118.82	-2.075
HDAC7	2491	1	21441	1.486E-06	0.1924619	1.6307E-02	15	23.31	-1.368	4.033E-06	0.5224496	9.7087E-02	15	138.78	-2.142

The background mutation rates are calculated based on the tumors that were in the discovery screen; the mutation rate of individual genes is based on 52 tumors (12 in the discovery screen tumors and 40 prevalence screen tumors).

Supplementary Table 12: MSI status and *MSH6* status of endometrial tumors with *CHD4*, *FBXW7*, and *SPOP* mutations

Pair-wise comparison	Frequency in Serous Tumors *	Frequency in Clear Cell Tumors	Frequency in Endometrioid Tumors **	Frequency in Mixed Histology Tumors
CHD4 mutant/MSI STABLE-*MSH6* WILDTYPE	9 of 48 (18.8%)	1 of 18 (5.5%)	3 of 41 (7.3%)	1 of 15 (6.7%)
CHD4 mutant/MSI UNSTABLE-*MSH6* MUTANT	0 of 3 (0%)	0 of 5 (0%)	2 of 25 (8.0%)	1 of 3 (33.3%)
P-value	1	1	1	0.3137
FBXW7 mutant/MSI STABLE-*MSH6* WILDTYPE	14 of 48 (29.2%)	1 of 18 (5.5%)	3 of 41 (7.3%)	1 of 15 (6.7%)
FBXW7 mutant/MSI UNSTABLE-*MSH6* MUTANT	1 of 3 (33.3%)	2 of 5 (40%)	4 of 25 (16%)	1 of 3 (33.3%)
P-value	1	0.1073	0.4122	0.3137
SPOP mutant/MSI STABLE-*MSH6* WILDTYPE	4 of 48 (8.3%)	2 of 18 (11.1%)	0 of 41 (0%)	0 of 15 (0%)
SPOP mutant/MSI UNSTABLE-*MSH6* MUTANT	0 of 3 (0%)	0 of 5 (0%)	0 of 25 (0%)	0 of 3 (0%)
P-value	1	1	1	1

P-value (Fisher's Exact test of significance, 2-tailed)

* The MSI status for 1 of 52 serous tumors could not be determined

** The MSI status for 1 of 67 endometrioid tumors could not be determined

Supplementary Table 14: Networks identified by Ingenuity Pathway Analysis

Network ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	Actin,ARHGEF2,ATM,CHD4,CHKA,Creb,CTCF (includes EG:10664),Cyclin E,CYP11B2,DMD,EP300,FBXW7,FLNA,hCG,Histone h3,Histone h4,LHCGR,MAGEC2/MAGEC3,mediator,NUMA1,PEX2 (includes EG:19302),PLAA,POLR1A,RFX5,Rnr,SIN3A,SLC26A3,SMC3,SMC4,TIP60,TP53 (includes EG:22059),TRRAP,TSPYL2,YEATS4,ZBTB33	38	26	DNA Replication, Recombination, and Repair, Gene Expression, Cancer
2	AMPH,ANO3,ANXA5,CD3,CDCA7,CDH11,DSG3,ERBB2,ERK,ERK1/2,Focal adhesion kinase,GPER,GRM1,HGF,IL2RG,IRS4 (includes EG:16370),ITCH,ITGA3,MAP2K1/2,MAP4K4,Mapk,Mek,NFkB (complex),P38 MAPK,p85 (pik3r),PI3K (complex),PIK3CA,PPP2R1A,PPP2R2A,PUS7,RNF128,SPDEF,SPTB,TCR,TLR7	32	23	Cellular Movement, Cell Morphology, Tissue Morphology
3	ABCA3,AGTR1,ATAD2,C3,CCND1,CDCA7,DNMT3B,DPY30,EMR3,ESR1,EZH2,FHOD3,GPR77,HAMP,HDAC7,HFE2,HK3,HNRNPH2,KANSL1, KCNA1,KCNQ5,LIMCH1,MGA,MLL,NEO1,NRF1,PNRC1,RBMS3,RECK,SLC9B2,SRPK2,TGFBR2,TGM2,TMPO,WDFY3	18	16	Developmental Disorder, Hematological Disease, Hereditary Disorder
4	AKR1C3,CALD1,CCDC82,CEP63,COL7A1,CRMP1,CTDSP2,DISC1,DST,EGFR,ERBB2,estrogen receptor,GRM5,HDC,HFE,HM13,HSPA5,JUN,KHSRP,KIF1A,KRT7,MAPK1,MPHOSPH9,PCDH7,PCDHA9,PGR (includes EG:18667),PTPRK,SBDS,SMAD1,SORCS1,SPAG16,SPOCK1,SYNE1,TNXB,WNT5B	17	15	Endocrine System Development and Function, Small Molecule Biochemistry, Cell Cycle
5	ABLIM,ACSL4,ACTG1,ACVR2A,ADCY7,AKAP9,ANKS1B,ARHGAP35,ATP9A,BCR (complex),CXCR4,DIAPH3,DUSP3,DYRK3,EPSTI1,ERG,FSH,GNLY,Lh,LOC81691,LPXN,MT3,P4HA2,PDXK,PLCL1,PPFIA4,PRIC285,PTGS2,RAB33B,RGS5,SATB1,SPTBN5,TPM2,TRIOBP,VAV1	15	14	Antigen Presentation, Cellular Movement, Hematological System Development and Function
6	ACTBL2,BGN,CACNA1D,CCNA2,CD40LG,CD79A,CDKN1A,COL3A1,CUX1,EIF3M,ESRRG,GAL3ST1,GCH1,HIF1A,HIF1AN,HSPB2,IFIT5,IFNG (includes EG:15978),Igm,KDM4B,MAPK9,MARCKSL1,MRC1 (includes EG:100286774),NF1 (includes EG:18015),NOV,PLEK,POLI,PPP2CA,PPP2R3A,SGOL1,SMTN,SPAG4,TGFB1 (includes EG:21803),TYMP,ZFP36L1	14	13	Embryonic Development, Organ Development, Organismal Development
7	ACTL6A,ADAM9,ADNP,ARID1A,ATP6AP1,BAF110,CCNC,CD46,CDK5,CSPG4,CTSB,EXT2,FAN1,FHL2,ITGA10,ITGB1,LIG3,MSH6,PDP1,PHLDB1,PMS2,POU5F1,PPAP2B,PRPF6,PRPF4B,SALL4,SERPINE2,SLC3A2,Smad2/3,SMARCA4,SMARCC2,STAT3,SYVN1,TRPM3,WDR45	12	12	Cellular Compromise, Nervous System Development and Function, Tissue Development
8	ANLN,BAZ1B,CHEK1,DBF4 (includes EG:10926),DDX11/DDX12P,E2F4,ERCC8,EXOSC8,FANCE,KIAA0101,KIF1B,KISS1,MARK1,MCM5,MED14,PANX3,PCNA,PKN2,PTPRN2,RAD50,RAD51AP1,RBBP8,RINT1,RNA polymerase II,RRM1,RRM2,RUNX2,SHPRH,SND1,SPAG7,SVIL,TLN1,YWHAG,ZNF83,ZNF451	12	12	DNA Replication, Recombination, and Repair, Cancer, Reproductive System Disease
9	ACHE,ADORA2B,AKAP2/PALM2-AKAP2,BIRC6,CASP1,CASP9 (includes EG:100140945),CCL18,CFB,CNTN1,DIABLO,DSC3 (includes EG:13507),FNDC3A,GCLM,IGHM,IGL@,LAMA3,LAMB3,LAMP3,MAFF,MOAP1,MS4A1,NFKB2,OASL,RPS13,RXRA,SEC16A,SERPINB1,SERPINB9,SFMBT2,SLAMF7,STAB1,TNF,TNIP3,TREM1,ZFP91	11	11	Cell Death, Antigen Presentation, Cell Morphology
10	ADAMTS9,Akt,ASGR1,ATP6V0D1,BAG4,CDC37L1,DNAJB1,FGFR4,GDNF,HDAC8,Hsp70,Hsp90,IgG,IRS4 (includes EG:16370),Jnk,KCNA5,KIF26B,KRT31,MDK,mir-29,MTOR,NDFIP1,NQO1,PACRG,PEA15,PIK3CB,PRR5,Rar,SIAH1,SMG5,SNCAIP,STIP1,STK3,STK4,TRIM16	9	10	Cardiovascular System Development and Function, Embryonic Development, Organismal Development
11	ACTN4,AJAP1,AXIN1,CDH1,CDK17,CDK5R1,CPB2,CTNNA3,CTNNB1,CUL3 (includes EG:26554),CYP24A1,DAXX,DBH,DIXDC1,FAM123B,GUCY2F,Hdac,HNF1A,HSD17B2,KLK6,LDL-cholesterol,LRP2,LRP1 (includes EG:16971),MAP3K4,MSX2,NDRG1,PAFAH2,PI3K p85,PKD1,PTPRF,SLC12A6,SLIT2,SMAD4,SPOP,THBD	9	10	Cell-To-Cell Signaling and Interaction, Tissue Development, Renal and Urological System Development and Function
12	PDX1 (includes EG:18609),SEZ6L	1	1	Cell Morphology, Developmental Disorder, Digestive System Development and Function
13	CAPRIN2,mir-17	1	1	Embryonic Development, Hair and Skin Development and Function, Organ Development
14	TUB,WDR35	1	1	Cell Death, Nutritional Disease, Tissue Morphology
15	CCR9,NADPH oxidase	1	1	Cellular Function and Maintenance, Cellular Movement, Hematological System Development and Function
16	NGLY1,RAD23B	1	1	Digestive System Development and Function, Protein Degradation, Protein Synthesis
17	Nuclear factor 1,SLC25A5	1	1	Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry
18	ABCC9,KCNJ11	1	1	Molecular Transport, Cardiovascular Disease, Endocrine System Development and Function
19	NRK,ZNF217	1	1	Cellular Development, Reproductive System Disease, Cardiovascular System Development and Function
20	FOXL2,TCEB3B	1	1	Cellular Development, Dermatological Diseases and Conditions, Developmental Disorder
21	GNPAT,PEX7 (includes EG:18634)	1	1	Lipid Metabolism, Small Molecule Biochemistry, Connective Tissue Disorders
22	NDUFS3,NUBPL	1	1	Hereditary Disorder, Metabolic Disease, Cellular Assembly and Organization
23	RHAG,RHCE/RHD,SLC4A1	1	1	Cellular Assembly and Organization, Hematological Disease, Cell-To-Cell Signaling and Interaction
24	VPS11,VPS18,VPS16 (includes EG:296159)	1	1	Cell Cycle, Cellular Assembly and Organization, Cellular Function and Maintenance

Supplementary Table 15: Estimated power to detect somatically mutated genes in a discovery screen of 12 tumors

Actual mutation frequency of gene "G" in serous endometrial cancer	Power to detect one tumor with a mutation in gene "G" in a discovery screen of 12 tumors	Power to detect two tumors with a mutation in gene "G" in a discovery screen of 12 tumors
8%	63.20%	24.90%
10%	71.80%	34.10%
15%	85.80%	55.70%
20%	93.10%	72.50%