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

Matthieu Le Gallo1,10, Andrea J O’Hara1,10, Meghan L Rudd1, Mary Ellen Urick1, Nancy F Hansen2, Nigel J O’Neil3, Jessica C Price1, Suiyuan Zhang2, Bryant M England1, Andrew K Godwin4, Dennis C Sgroi5,6, NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program7,8, Philip Hieter3, James C Mullikin2,8, Maria J Merino9 & Daphne W Bell1

Endometrial cancer is the sixth most commonly diagnosed cancer in women worldwide, causing ~74,000 deaths annually1. Serous endometrial cancers are a clinically aggressive subtype with a poorly defined genetic etiology2–4. We used whole-exome sequencing to comprehensively search for somatic mutations within ~22,000 protein-encoding genes in 13 primary serous endometrial tumors. We subsequently resequenced 50 genes, which had been mutated in more than 1 tumor and/or were components of an enriched functional grouping, from 40 additional serous tumors. We identified high frequencies of somatic mutations in CHD4 (17%), EP300 (8%), ARID1A (6%), TSPYL2 (6%), FBXW7 (29%), SPOP (8%), MAP3K4 (6%) and ABC29 (6%). Overall, 36.5% of serous tumors had a mutated chromatin-remodeling gene, and 33% had a mutated ubiquitin ligase complex gene, implicating frequent mutational disruption of these processes in the molecular pathogenesis of one of the deadliest forms of endometrial cancer.

We performed targeted exon capture and next-generation sequencing on DNA from 13 primary serous endometrial tumors with high neoplastic cellularity (Supplementary Table 1) and matched normal tissues. The mean depth of coverage for aligned reads was 102.6×, and, on average, 89.5% of targeted bases had sufficient coverage and quality for variant calling (Supplementary Table 2). Using stringent filtering criteria, which included an empirically determined threshold that accounted for read quality and depth and provided an optimal balance between positive predictive value (86.1%) and sensitivity (97.3%) (Online Methods), we identified 1,522 exonic somatic mutations (1,183 nonsynonymous and 339 synonymous) and 22 splice-junction mutations within the protein-encoding genes of the 13 tumors (Supplementary Fig. 1). One tumor had an apparent hypermutable phenotype, with a greater number of mutations and a different mutation signature than the other 12 tumors (Supplementary Fig. 2 and Supplementary Tables 3 and 4), and was excluded from subsequent analyses.

Among the remaining 12 tumors, we identified 516 exonic mutations (380 nonsynonymous and 136 synonymous) and 11 splice-junction mutations (Supplementary Table 5). We could orthogonally assess 510 of the 527 exonic and splice-junction mutations by Sanger sequencing: 86.1% (439 of 510) of mutations were confirmed as somatic (Supplementary Tables 5 and 6). The validated somatic mutations included 321 nonsynonymous mutations (279 missense, 86.9%; 19 nonsense, 5.9%; 17 frameshift, 5.3%; 6 in-frame insertions or deletions, 1.9%) and 9 splice-junction mutations in 304 protein-encoding genes and 109 synonymous mutations (Supplementary Table 5). There was a mean of 27.5 validated nonsynonymous and splice-junction mutations per tumor (range of 5–55) (Supplementary Table 6). The predicted functional impact of the validated missense mutations was assessed in silico: 34.4% of the 241 missense mutations that could be orthogonally assessed by both the SIFT and Mutation Assessor algorithms were predicted functional impact of the validated missense mutations was assessed in silico: 34.4% of the 241 missense mutations that could be orthogonally assessed by both the SIFT and Mutation Assessor algorithms were predicted to affect protein function (Supplementary Table 5).

To prioritize our search for previously unknown driver mutations in serous endometrial cancer, we focused on the nine genes that had validated nonsynonymous somatic mutations in more than one tumor (Supplementary Table 7). We resequenced these genes in a prevalence screen of 40 additional serous endometrial tumors. Three of the nine genes (TP53, PIK3CA and PPP2R1A) have established roles in the pathogenesis of serous endometrial cancer5–8. TP53, PIK3CA and PPP2R1A were mutated, respectively, in 71%, 31% and 25% of the 52 serous tumors in our study, as reported here and in a previous study from our group6 (Supplementary Tables 8 and 9). The other six genes (CHD4, SPOP, FBXW7, ABCC9, CYP4X1 and...
Figure 1 Somatic mutations in CHD4, FBXW7 and SPOP cluster within important functional domains of the encoded proteins. Schematics of the CHD4, FBXW7 and SPOP proteins show the positions of individual somatic alterations identified in primary endometrial tumors relative to important functional domains. Alterations in serous (yellow boxes), clear-cell (brown boxes), endometriod (red boxes) and mixed-histology (white boxes) endometrial tumors are indicated. (a) Fifty percent of all CHD4 alterations localize to the ATPase/helicase and helicase domains. (b) The majority of FBXW7 alterations in endometrial cancer cluster within the WD repeats. The FBXW7 p.Glu65* and p.Lys70fs*27 alterations (data not shown) affect an alternate isoform. (c) All SPOP alterations in endometrial cancer were localized to the MATH domain. BTB domain, broad complex, tramtrack and bric-a-brac domain; D domain, dimerization domain; MATH domain, meprin and TRAF homology domain; NLS, nuclear localization signal; PHD, plant homeodomain-type zinc finger; WD repeat, tryptophan–aspartic acid repeat.

MAP3K4 have no previously reported role in serous endometrial cancer. The combined discovery and prevalence screens identified high-frequency somatic mutations in CHD4 (17%), FBXW7 (29%), SPOP (8%), MAP3K4 (6%), ABCC9 (6%) and CYP4X1 (4%) (Fig. 1, Table 1, Supplementary Figs. 3 and 4 and Supplementary Table 9). The mutation rates for CHD4, FBXW7 and SPOP were significantly higher than the background mutation rate (q ≤ 0.0353) (Supplementary Table 10).

In addition to serous tumors, the major histological subtypes of endometrial cancer are clear cell and endometrioid, with overall 17% of mixed-histology endometrial cancers (Table 1).

| Gene name (RefSeq ID; UCSC transcript accession ID) | Tumor | Histology | Nucleotide change | Amino-acid change | Mutation type | Mutation Assessor predicted effect on function |
|---------------------------------------------------|-------|-----------|-------------------|------------------|--------------|-----------------------------------------------|
| CHD4 (NM_001273; uc001qpo.2)                      | T3 (OM1323) | Serous | c.2450G>A | p.Arg817Gln | Missense | Medium |
| CHD4 (NM_001273; uc001qpo.2)                      | T3 (OM1323) | Serous | c.4709G>A | p.Ser1570Asn | Missense | Low |
| CHD4 (NM_001273; uc001qpo.2)                      | T10 (OM2009) | Serous | c.4013G>A | p.Arg1338Lys | Missense | Medium |
| CHD4 (NM_001273; uc001qpo.2)                      | T29 | Serous | c.3314G>A | p.Arg1105Gln | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T50 | Serous | c.3452A>C | p.His1151Arg | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T78 | Serous | c.2734T>C | p.Leu912Val | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T165 | Serous | c.4001G>T | p.Gly1334Val | Missense | Medium |
| CHD4 (NM_001273; uc001qpo.2)                      | T33 | Serous | c.3485G>A | p.Arg1162Gln | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T154 | Serous | c.1672G>T | p.Val558Phe | Missense | Medium |
| CHD4 (NM_001273; uc001qpo.2)                      | T56 | Serous | c.3460_3462delATT | p.Asn1152del | In-frame deletion | – |
| CHD4 (NM_001273; uc001qpo.2)                      | T59 | Serous | c.2629C>T | p.Arg877Trp | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T97 | Endometriod | c.1715G>A | p.Arg572Gln | Missense | Medium |
| CHD4 (NM_001273; uc001qpo.2)                      | T115 | Endometriod | c.4946C>T | p.Ala1649Val | Missense | Low |
| CHD4 (NM_001273; uc001qpo.2)                      | T120 | Endometriod | c.2620_2622delGAA | p.Glu874del | In-frame deletion | – |
| CHD4 (NM_001273; uc001qpo.2)                      | T126 | Endometriod | c.5704G>A | p.Glu1902Lys | Missense | Low |
| CHD4 (NM_001273; uc001qpo.2)                      | T133 | Endometriod | c.3484C>T | p.Arg1162Trp | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T173 | Mixed | c.3644T>C | p.Leu1215Pro | Missense | High |
| CHD4 (NM_001273; uc001qpo.2)                      | T179 | Mixed | c.2186delA | p.Lys72Argfs*128 | Frameshift | – |
| FBXW7 (NM_0018315; uc003imr.2)                     | T3 (OM1323) | Serous | c.193G>T | p.Glu65* | Nonsense | – |
| FBXW7 (NM_0018315; uc003imr.2)                     | T113 | Clear cell | c.207_208insA | p.Asn69Lysfs*26 | Frameshift | – |
| FBXW7 (NM_0033632; uc003ims.2)                     | T3 (OM1323) | Serous | c.288A>C | p.Glu96Asp | Missense | Neutral |

CHD4 (chromodomain helicase DNA-binding protein 4) is a catalytic subunit of the NuRD (nucleosome remodeling and deacetylase) complex, which regulates transcriptional repression, chromatin assembly and the DNA damage response10–19. We confirmed endogenous CHD4 expression in endometrial cancer cells (Supplementary Fig. 6). CHD4 was highly mutated in serous tumors (17%) and was also mutated in clear-cell (4%), endometrioid (7%) and mixed-histology (11%) tumors (Supplementary Table 9). We found that 80% of CHD4 missense mutations, including those affecting an Arg1162 hotspot, were predicted to have an impact on protein function (Fig. 1 and Table 1). Most CHD4 mutations were missense; 83% of all CHD4 mutations affected residues that are highly conserved throughout evolution or across closely related family members (Supplementary Figs. 7 and 8).

Half of all CHD4 mutations affected the ATPase/helicase and helicase domains (Fig. 1a); two-thirds (6 of 9) of these mutations affected conserved functional domains. Alterations in serous (yellow boxes), clear-cell (brown boxes), endometriod (red boxes) and mixed-histology (white boxes) endometrial tumors are indicated. (a) Fifty percent of all CHD4 alterations localize to the ATPase/helicase and helicase domains. (b) The majority of FBXW7 alterations in endometrial cancer cluster within the WD repeats. The FBXW7 p.Glu65* and p.Lys70fs*27 alterations (data not shown) affect an alternate isoform. (c) All SPOP alterations in endometrial cancer were localized to the MATH domain. BTB domain, broad complex, tramtrack and bric-a-brac domain; D domain, dimerization domain; MATH domain, meprin and TRAF homology domain; NLS, nuclear localization signal; PHD, plant homeodomain-type zinc finger; WD repeat, tryptophan–aspartic acid repeat.

Table 1 Somatic mutations identified in the discovery and prevalence screens of CHD4, FBXW7, SPOP, MAP3K4, ABCC9 and CYP4X1

(continued)
residues that undergo germline or de novo pathogenic mutations in SMARCAL1, SMARCA4 or SMARCA2, causing Schimke immunosseous dysplasia, Coffin-Siris syndrome and Nicolaides-Baraitser syndrome, respectively (Supplementary Fig. 9)\(^{20–22}\). This observation leads us to speculate that somatic mutations affecting the ATPase/helicase domain of CHD4 may be driver mutations in endometrial cancer.

Other frequently mutated genes in our study were FBXW7 and SPOP. The FBXW7 (F-box and WD repeat domain containing 7) tumor suppressor is a component of the FBXW7-SKP1-CUL1 ubiquitin ligase complex, which mediates ubiquitination and proteasomal degradation of phosphoprotein substrates, including cyclin E, Notch, JUN and MYC\(^{23}\). Previous reports of FBXW7 mutations in endometrial cancer

| Gene name (RefSeq ID; UCSC transcript accession ID\(^a\)) | Tumor | Histology | Nucleotide change | Amino-acid change | Mutation type | Mutation Assessor predicted effect on function |
|---|---|---|---|---|---|---|
| T43 | Serous | c.1590_1608delTGAAA | p.Glu531Cysfs*19 | Frameshift | – |
| T50 | Serous | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T53 | Serous | c.1393C>T | p.Arg465Cys | Missense | Medium\(^b\) |
| T74 | Serous | c.1394G>A | p.Arg465His | Missense | High\(^b\) |
| T78 | Serous | c.1394G>A | p.Arg465His | Missense | High\(^b\) |
| T79 | Serous | c.504G>T | p.Met168Ile | Missense | Low |
| T79 | Serous | c.41G>A | p.Arg14Gln | Missense | Neutral |
| T79 | Serous | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T114 | Serous | c.1513C>T | p.Arg505Cys | Missense | Medium\(^b\) |
| T162 | Serous | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T166 | Serous | c.1634A>T | p.Tyr545Cys | Missense | Medium |
| T167 | Serous | c.2065C>T | p.Arg689Trp | Missense | Medium |
| T29 | Serous | c.1387A>G | p.Thr463Ala | Missense | Medium |
| T80 | Serous | c.1385C>T | p.Ser462Phe | Missense | Medium |
| T80 | Serous | c.1322G>C | p.Arg441Pro | Missense | Medium |
| T47 | Serous | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T69 | Serous | c.2065C>T | p.Arg689Trp | Missense | Medium |
| T32 | Clear cell | c.1199_1201delACA | p.Asn401del | In-frame deletion | – |
| T77 | Clear cell | c.1513C>T | p.Arg505Cys | Missense | Medium\(^b\) |
| T84 | Endometrioid | c.1268G>T | p.Gly423Val | Missense | Medium |
| T85 | Endometrioid | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T88 | Endometrioid | c.994G>T | p.Glu332* | Nonsense | – |
| T88 | Endometrioid | c.1452G>A | p.Arg484Ser | Missense | High |
| T94 | Endometrioid | c.1393C>T | p.Arg465Cys | Missense | Medium\(^b\) |
| T97 | Endometrioid | c.1436G>A | p.Arg479Gln | Missense | Medium\(^b\) |
| T115 | Endometrioid | c.1508G>T | p.Ala503Val | Missense | Medium |
| T119 | Endometrioid | c.1513C>T | p.Arg505Cys | Missense | Medium\(^b\) |
| T178 | Mixed | c.1508G>T | p.Ala503Val | Missense | Medium |
| T179 | Mixed | c.791A>G | p.Gln264Arg | Missense | Medium |
| T179 | Mixed | c.1972C>T | p.Arg658* | Nonsense | – |
| T25 | Serous | c.240C>G | p.Ser80Arg | Missense | Medium |
| T56 | Serous | c.240C>G | p.Ser80Arg | Missense | Medium |
| T80 | Serous | c.280C>G | p.Pro94Ala | Missense | Low |
| T154 | Serous | c.362G>A | p.Arg121Gln | Missense | Low |
| T110 | Clear cell | c.351G>A | p.Met117Ile | Missense | Low |
| T153 | Clear cell | c.139G>A | p.Glu47lys | Missense | Medium |
| T79 | Serous | c.1837G>A | p.Glu613Lys | Missense | Low |
| T79 | Serous | c.3928G>T | p.Glu1310* | Nonsense | – |
| T80 | Serous | c.4165G>A | p.Glu1389Lys | Missense | High |
| T65 | Serous | c.4077delC | p.Cys1359Alafs*8 | Frameshift | – |
| T3 (OM1323) | Clear cell | c.2666G>A | p.Ser889Asn | Missense | Low |
| T3 (OM1323) | Serous | c.2686G>T | p.Glu956Asp | Missense | Low |
| T70 | Serous | c.2185G>T | p.Val729Phe | Missense | Medium |
| T75 | Serous | c.3381G>A | p.Met1127Ile | Missense | Low |
| T75 | Serous | c.1036G>C | p.Val346Leu | Missense | Neutral |
| T56 | Serous | c.211G>A | p.Glu71Lys | Missense | Neutral |

\(^{a}\)UCSC transcript IDs are based on the hg18 assembly of the human genome sequence. \(^{b}\)Known loss-of-function mutations in FBXW7.
either did not include serous and clear-cell tumors or did not report the histology of mutated cases24–27. We identified FBXW7 mutations in 29% of serous, 13% of clear-cell, 10% of endometrioid and 11% of mixed-histology endometrial cancers (Supplementary Table 9). The mutation frequency was significantly higher in serous tumors than high-grade endometrioid tumors (29% versus 0%; \( P = 0.0146 \) by Fisher’s two-tailed exact test of significance). Most FBXW7 mutations affected the substrate-binding WD repeats (Fig. 1b), consistent with the mutation spectrum in other cancers28. Forty-six percent of the FBXW7 missense mutations in our study are known loss-of-function mutations; another 39% of missense mutations were predicted to affect function29 (Table 1). Our findings may be clinically relevant because loss-of-function of FBXW7 function correlates with resistance to antitubulin chemotherapeutics30 and sensitivity to an HDAC inhibitor31.

Table 2

| Gene name (RefSeq ID; UCSC transcript accession ID) | Mutation frequency in serous tumors | Mutation frequency in clear-cell tumors | Nucelotide change | Amino-acid change | Mutation type | Mutation Assessor predicted effect on function |
|-----------------------------------------------|-----------------------------------|---------------------------------------|------------------|-----------------|-------------|---------------------------------------------|
| EP300 (NM_001429; uc003azi.3)                  | 8% (4/52)                         | 4% (1/23)                             | c.1959G>A        | p.Arg1732His    | Missense    | Medium                                      |
| T3 (OM1323)                                    | Serous                            | Serous                               | c.4879C>T        | p.Ala1727Thr    | Missense    | Medium                                      |
| T74                                           | Serous                            | Serous                               | c.4880G>A        | p.Ala1727Gln    | Missense    | Medium                                      |
| T79                                           | Serous                            | Serous                               | c.4404A>C        | p.Lys1468Asn    | Missense    | Medium                                      |
| T113                                          | Clear cell                        | c.2863C>T                            | p.Pro955Ser      | Missense        | Neutral    |                                            |
| ARID1A (NM_006015; uc001trmv.1)                | 6% (3/52)                         | 13% (3/23)                           | c.5161C>T        | p.Arg1721*      | Nonsense    |                                            |
| T74                                           | Serous                            | c.3826C>T                            | p.Ala1267*       | Nonsense        |            |                                            |
| T79                                           | Serous                            | c.5299G>T                            | p.Glu1767*       | Nonsense        |            |                                            |
| T79                                           | Serous                            | c.5965C>T                            |                     |                 |             |                                            |
| T76                                           | Clear cell                        | c.1645delC                          | p.Tyr551Hfrs*68  | Frameshift      |            |                                            |
| T78                                           | Clear cell                        | c.3478_3479insT                     |                   |                 |             |                                            |
| T113                                          | Clear cell                        | c.5142G>T                           | p.Glu1714Asp     | Missense        | Low         |                                            |
| TSPYL2 (NM_022117; uc004dpx.1)                  | 6% (3/52)                         | 0% (0/23)                            | c.752A>C         | p.His251Pro     | Missense    |                                            |
| TSPYL2 (NM_022117; uc004dwr.2)                 |                                  |                                      |                  |                 |             |                                            |
| T70                                           | Serous                            | c.1636G>T                            | p.Glu546Cys      | Missense        | Neutral    |                                            |
| T3 (OM1323)                                    | Serous                            | c.1952G>C                            | p.Glu651Asa      | Missense        | Neutral    |                                            |
| KDM4B (NM_015015; uc002mfbq.3)                  | 4% (2/52)                         | 0% (0/23)                            | c.2138C>T        | p.Pro713Leu     | Missense    | Neutral                                     |
| T115                                          | Serous                            | c.964G>A                             | p.Ala322Met      | Low             |            |                                            |
| T3 (OM1323)                                    | Serous                            | c.1855C>T                            | p.Pro619Ser      | Low             |            |                                            |
| TRIM16 (NM_006470; uc002gxo.2)                 | 2% (1/52)                         | 0% (0/23)                            | c.1319delC       | p.Y440Metfs*3   | Frameshift |                                            |
| CTCF (NM_006565; uc002eti.2)                   | 2% (1/52)                         | 0% (0/23)                            | c.1094A>C        | p.Lys365Thr     | Missense    | Neutral                                     |
| HDAC7 (NM_015401; uc001rqi.2)                  | 2% (1/52)                         | 0% (0/23)                            | c.1024G>A        | p.Glu342Lys     | Low         |                                            |
| TRRAP (NM_003496; uc003.app.1)                  |                                  |                                      | c.5035G>A        | p.Ala1679Thr    | Low         |                                            |
| YEATS4 (NM_006530; uc001sux.1)                 | 2% (1/52)                         | 0% (0/23)                            | c.617delA        | p.Thr206Leufs*1 | Frameshift |                                            |
| BAZ1B (NM_032408; uc003ydc.1)                  | 4% (2/52)                         | 4% (1/23)                            | c.3448G>A        | p.Glu1150Lys    | Missense    | Neutral                                     |
| T77                                           | Clear cell                        | c.3167G>A                            | p.Arg1056His     | Missense        | Low         |                                            |

Table 2 Somatic mutations identified in the discovery and prevalence screens of ten chromatin-remodeling genes

Figure 2 Distribution of nonsynonymous somatic mutations in endometrial cancers. (a,b) OncoPrint showing the distribution of nonsynonymous somatic mutations in ubiquitin ligase complex and chromatin-remodeling genes and TP53, PPP2R1A and PIK3CA in 26 serous tumors (a) and mutations in ubiquitin ligase complex and chromatin-remodeling genes in 8 clear-cell endometrial tumors (b). Individual tumors are indicated by blue bars. Nonsynonymous somatic mutations are indicated by yellow bars. Only tumors that had somatically mutated ubiquitin ligase complex and/or chromatin-remodeling genes are shown. Collectively, the two ubiquitin ligase complex genes that regulate ubiquitin-mediated proteolysis were mutated in 35% (18 of 52) of serous endometrial tumors and 22% (5 of 23) of clear-cell endometrial tumors; the 11 proteolysis were mutated in 35% (18 of 52) of serous endometrial tumors and 22% (5 of 23) of clear-cell endometrial tumors.

SPOP (encoding speckle-type POZ protein) was somatically mutated in serous (8%) and clear-cell (9%) tumors but was not mutated in endometrioid or mixed-histology tumors (Supplementary Table 9). The mutation frequency was significantly higher in serous relative to endometrioid tumors (8% versus 0%; \( P = 0.0341 \) by Fisher’s two-tailed exact test of significance). The SPOP protein forms part of a multisubunit cullin 3 (CUL3)-dependent ubiquitin ligase complex and has recently been shown to be mutated at high frequency in prostate cancer32. All SPOP alterations in endometrial cancer, including a recurrent SPOP p.Ser80Arg alteration, were localized to
highly evolutionarily conserved residues within the MATH domain that acts as the substrate recognition domain to bind proteins targeted for ubiquitin-mediated degradation (Fig. 1c). This localization is very similar to that of loss-of-function alterations affecting the substrate recognition domain of FBXW7 (Fig. 1b). Thus, we predict that SPOP mutations in endometrial cancer are likely to result in loss-of-function mutants with impaired substrate binding. Notably, the SRC-3 oncoprotein (also known as AIB1 and NCOA3), an SPOP substrate, is overexpressed in endometrial cancer independent of gene amplification.

To identify additional candidate driver genes for serous endometrial cancer, we evaluated the functional relationships of the 304 protein-encoding genes that had orthogonally validated mutations (nonsynonymous or splice junction) in the discovery screen (Supplementary Tables 13 and 14). One of the enriched functional groupings was chromatin modification, which was formed by CHD4 and ten other genes (EP300, ARID1A, TSPYL2, KDM4B, TRIM16, HDAC7, CTCF, YEATS4, TRRAP and BAZ1B) (Supplementary Table 13). Although this enriched grouping did not achieve statistical significance after correction for multiple testing, we focused on it because it contained CHD4, one of the most highly mutated genes identified in our study, and because chromatin-remodeling genes are a frequent target of somatic mutations in other types of cancer.

We therefore resequenced the 10 additional chromatin-remodeling genes in 40 additional serous and 23 clear-cell endometrial tumors. In the combined discovery and prevalence screens, the 11 chromatin-remodeling genes were somatically mutated in 36.5% of serous tumors and 22% of clear-cell tumors (Fig. 2 and Tables 1 and 2). Two of the mutated genes, EP300 (encoding E1A-binding protein p300) and ARID1A (encoding AT-rich interactive domain 1A (SWI-like)) are consensus cancer genes. EP300 and ARID1A were mutated, respectively, in 8% and 6% of serous tumors and in 4% and 13% of clear-cell tumors (Table 2). Most EP300 mutations were localized to the sequence encoding the histone acetyltransferase domain of p300, a global transcriptional coactivator (Fig. 3a). p300 Arg1627 constituted a mutational hotspot in endometrial cancer; the p300 p.Arg1627Trp alteration has also been described in lymphoma. ARID1A encodes the BAF250A tumor suppressor, a subunit of the SWI/SNF-A chromatin-remodeling complex. Most ARID1A mutations we identified were predicted to truncate BAF250A (Fig. 3b), consistent with the mutation spectrum in other tumors. Our finding of ARID1A mutations in 6% of serous endometrial cancers is consistent with a recent study that documented ARID1A mutations in 11% of serous endometrial tumors. To our knowledge, this is the first report of ARID1A mutations in clear-cell endometrial cancer, substantiating previous reports of loss of BAF250A expression in this histotype.

Our study provides new insights into the somatic mutations present in serous endometrial cancer exomes. However, it is important to acknowledge that our discovery screen is underpowered to detect all somatically mutated genes that drive serous tumors. For example, PIK3R1, which we previously found to be somatically mutated in 8% of serous endometrial tumors, was not somatically mutated in the tumors that formed our discovery screen. We estimate that, for genes that are mutated in 8% of all serous endometrial cancers, a discovery screen of 12 tumors has 25% power to detect 2 mutated tumors and 63% power to detect 1 mutated tumor (Supplementary Table 15); for genes that are mutated in 20% of all serous endometrial cancers, our discovery screen had an estimated 72.5% power to detect 2 mutated tumors and 93% power to detect 1 mutated tumor. Massively parallel sequencing of additional cases will undoubtedly yield deeper insights into the mutational landscape of serous endometrial cancer.

Herein, we report one of the first exome sequencing analyses of serous endometrial cancers, which are clinically aggressive tumors that have been poorly characterized genomically. Our findings implicate the disruption of chromatin-remodeling and ubiquitin ligase complex genes in 50% of serous endometrial tumors and 35% of clear-cell endometrial tumors (Fig. 2). The high frequency and specific distributions of mutations in CHD4, FBXW7 and SPOP strongly suggest that these are likely to be driver events in serous endometrial cancer.

Note added in proof: While this manuscript was in the final stages of publication, Kuhn et al. reported frequent FBXW7 mutations in serous endometrial cancer exomes.

URLs. Cross_match, http://www.phrap.org/phredphrapconsed.html; bam2mpg, http://research.nghri.nih.gov/software/bam2mpg/; SIFT, http://sift.jcvi.org/; Mutation Assessor, http://mutationassessor.org/; Ingenuity Systems Pathway Analysis, www.ingenuity.com/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank our colleagues for critical reading of the manuscript: R.T. Moreland and N. Trivedi of the National Human Genome Research Institute Bioinformatics and Scientific Programming Core, respectively, for performing in silico PCR and giving advice on statistics; and J. Teer for sharing expertise on VarSifter. A. Santin (Yale School of Medicine) kindly provided the ARK1 and ARK2 cell lines. The study was funded in part by the Intramural Program of the National Human Genome Research Institute, US NIH (D.W.B., J.C.M. and M.J.M.); NIH grant R01CA112021 (D.C.S.); the Avon Foundation (D.C.S.); NIH grant R01CA140323 (A.K.G.); and the Ovarian Cancer Fund (A.K.G.). P.H. is supported by grants from the NIH (CA016519) and by the Canadian Institutes for Health Research (MOP-39067).

AUTHOR CONTRIBUTIONS

D.W.B. designed and directed the study and wrote the manuscript. A.K.G. contributed clinical specimens. M.J.M. and D.C.S. conducted pathological review of clinical specimens. M.L.R. prepared DNA samples and performed identity...
testing and microsatellite instability analysis. NISC performed library construction and whole-exome sequencing. NISC and N.F.H. performed variant calling. M.L.G. and A.J.O. created and orthogonally validated exome sequencing data. M.L.G., A.J.O. and D.W.B. interpreted the exome data and established filtering criteria. M.L.G., A.J.O., M.L.R., J.C.P., B.M.E., S.Z. and D.W.B. designed, performed, analyzed and interpreted the mutation prevalence screens. A.J.O. and M.L.G. analyzed MSH6. M.E.U. and M.L.G. generated sequence conservation alignments. M.E.U. performed cell culture and immunoblotting. N.F.H., M.L.G. and J.C.M. performed statistical analyses. N.F.H. performed the power calculation. D.W.B., M.E.U., M.L.G., M.L.R., A.J.O., N.F.H., J.C.M., A.K.G., P.H. and N.J.O. edited and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/dofinder/10.1038/ng.2465. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
Online Methods

Clinical Material. Anonymized, snap-frozen primary tumor tissues, corresponding hematoyxin- and eosin-stained tumor sections, matched normal tissues (uninvolved reproductive tissue or whole blood) and clinicopathological information were obtained from the Cooperative Human Tissue Network, which is funded by the National Cancer Institute. The NIH Office of Human Subjects Research determined that this research was not ‘human subjects research’ per the Common Rule (45 CFR 46) and that no institutional review board (IRB) review was required for whole-exome sequencing of the samples. A small number of samples in the prevalence screen were obtained from the Biosample Repository at the Fox Chase Cancer Center or Oncomatrix. Tumor specimens were collected at surgical resection before treatment. Histological classifications were based on the entire specimen at the time of diagnosis. Tumors consisted of 53 serous cases, 23 clear-cell cases, 67 endometrioid cases and 18 cases of mixed histology. A pathologist reviewed hematoyxin- and eosin-stained sections of banked tumor tissues to verify that they were representative of the original histological classification and to delineate regions of high tumor cell content (>70%) for macrodissection.

DNA Extraction and Identity Testing. Genomic DNA was isolated from macrodissected tumor tissues and normal tissues using the Puregene kit (Qiagen) and purified by phenol–chloroform extraction. Tumor-normal pairs were typed using the Coriell Identity Mapping kit. Genotyping fragments were resolved on an ABI 3730xl DNA analyzer (Applied Biosystems) and scored using GeneMapper software (Applied Biosystems).

Exome Capture, Library Construction and Next-generation Sequencing. The exomes of tumor-normal pairs were captured using the Agilent SureSelect Human All Exon kit (three pairs) or the Agilent SureSelect Human All Exon 50Mb Target Enrichment kit (ten pairs) according to the manufacturer’s instructions. DNA captured using the SureSelect Human All Exon kit was run on the Illumina Genome Analyzer IIX platform with version 4 chemistry and version 2 flow cells; DNA captured using the SureSelect Human All Exon 50Mb Target Enrichment kit was run on the Illumina Genome Analyzer IIX platform with version 5 chemistry and version 4 flow cells, according to the manufacturer’s instructions, to generate 75- or 100-base-paired-end reads.

Read Mapping and Genotype Calling. Next-generation sequence reads were initially mapped to the human reference sequence NCBI Build 36 (hg18) using the Illumina ELAND alignment algorithm. When at least one read in a pair mapped to a unique location in the genome, that read and its pair were subjected to more accurate gapped alignment to the 100-kb region surrounding the location via Cross_match (see URLs). Alignments were stored in BAM format and fed as input to bam2mpq (see URLs) to call genotypes at all covered positions using a probabilistic Bayesian algorithm referred to as the most probable genotype (MPG) algorithm.60 As noted elsewhere60, highly reliable genotypes have an MPG score of ≥2.1 × 10^59. For tumor samples, bam2mpq was run with the score_variant option to calculate a most probable variant (MPV) score, which assesses the posterior probability of the existence of any variant at a position and is therefore more sensitive than the MPG score at positions for which there is uncertainty about whether a variant is heterozygous or homozygous non-reference. Additional information on the MPG and MPV scores is provided in the Supplementary Note.

Filtering of Variant Calls. We used a number of steps to filter nucleotide variants identified in the whole-exome screen. Germline variants called in paired tumor-normal samples were excluded from further analysis. Variants that were present within dbSNP build 132 that were not annotated as pathogenic or probable pathogenic variants in dbSNP were also excluded. We compared the remaining variants in each tumor exome to the variants in all 13 normal exomes sequenced in this study; variants called in both a tumor exome and a normal exome were excluded. Variants representing probable mapping ambiguities were also excluded. All remaining variants were considered to be potential somatic mutations and were annotated using the VarSifter software package61 into bins representing mutations in exons, introns, splice junctions, UTRs and non-coding RNAs.

Establishment of Filtering Criteria Based on Read Score and Read Depth. After filtering the exome data to exclude germline variants and probable mapping ambiguities, 798 somatic variants were called in the exons and splice junctions of 12 tumors (Supplementary Fig. 1 and Supplementary Table 16). We were able to orthogonally assess 730 of the 798 variants by Sanger sequencing; PCR products could not be generated for the remaining 68 variants. Of the 730 variants tested by Sanger sequencing, 451 were orthogonally validated as somatic (present in the tumor and absent in the matched normal DNA), yielding a positive predictive value of 61.8% (451 of 730). The remaining variants were either not detected within the tumor or were germline (present in DNA from both the tumor and matched normal tissue).

Because a positive predictive value of 61.8% was unacceptably low, we sought to empirically establish filtering criteria on the basis of sequence quality (MPV and MPG scores) and read depth (coverage (COV)) to achieve an optimal balance between accuracy and sensitivity in mutation detection. We observed (Supplementary Fig. 10) that the majority of mutations that did not orthogonally validate as somatic by Sanger sequencing had one or more of the following characteristics: (i) < 5 reads in the tumor and/or normal sample; (ii) an MPG score of <10 in the normal sample and/or an MPV score of <10 in the tumor sample; and (iii) an MPV/COV ratio of <0.5 in the normal sample. Therefore, we therefore retrospectively imposed filtering criteria to the 798 somatic variants called by exome sequencing requiring that there be (i) at least 5 reads in the tumor and normal samples and (ii) an MPV score of ≥10 in the normal sample and an MPV score of ≥10 in the tumor sample and; and (iii) an MPV/COV ratio of ≥0.5 in the normal sample; 527 variants were retained after filtering (Supplementary Table 5). Filtering on score and read depth gave a positive predictive value of 86.1% (510 of 527 retained variants could be assessed by Sanger sequencing; 439 of 510 (86.1%) assessed variants were orthogonally validated as somatic mutations) and a sensitivity of 97.3% (439 of 451 orthogonally validated mutations observed before applying the coverage score filter were retained after filtering).

These filtering criteria were also applied to the 1,042 somatic variants called in T155 by exome sequencing (Supplementary Table 17); 1,017 variants were retained after filtering (Supplementary Table 4).

Predicting the Functional Significance of Missense Mutations. Non-synonymous missense mutations called by whole-exome sequencing were evaluated in silico using the SIFT and Mutation Assessor algorithms (see URLs) to predict their impact on protein function. A SIFT prediction of deleterious and Mutation Assessor predictions of medium impact or high impact were considered to predict an effect on protein function.

Power of Study Design. Assuming N tumor samples are sequenced in a discovery screen and a fraction x of all tumor samples have gene G mutated, the probability that the N samples sequenced contain 0 samples with the gene mutation is (1 – x)^N. Therefore, the probability of observing gene G mutated at least once in the discovery screen was 1 – (1 – x)^N, which is 93% for 12 discovery samples, assuming x = 0.20 or 20%. Likewise, the probability of observing 0 or 1 sample with a mutation in gene G is (1 – x)^N + N(1 – x)^N(x)1x, giving the probability of seeing the mutation twice or more as 1 – (1 – x)^N – N(1 – x)(N – 1)x, which is 73% for 12 tumors, assuming x = 0.20.

PCR Amplification and Sanger Sequencing. Genomic DNA (5 ng) was amplified using M13-tailed primers (Supplementary Table 18) in a 10-µl PCR volume containing 1× AmpliTaq Gold PCR buffer (Applied Biosystems), 1.5 mM MgCl2, 75 mM dNTP, 400 nM sense primer, 400 nM antisense primer and 0.5 U of AmpliTaq Gold DNA polymerase. PCR amplification was performed on a GeneAmp PCR System 9700 (Applied Biosystems). PCR amplicons were purified using exonuclease I (Epicenter Biotechnologies) and shrimp alkaline phosphatase (USB Corporation) and were bidirectionally sequenced using the BigDye Terminator v.3.1 kit (Applied Biosystems) and M13-tailed primers. Cycled sequencing products were run on an ABI 3730xl DNA analyzer. Tumor and reference sequences were aligned and compared using in-house software to determine the genotype at variant nucleotide positions. Non-pathogenic variants present in dbSNP build 132 were excluded from further analysis. Somatic mutations were confirmed by reamplification.
and sequencing of DNA from matched tumor-normal pairs and analyzed using Sequencher software (Gene Codes Corporation).

**Determination of mutation rates and statistical analyses.** The somatic mutation rate was determined by dividing the total number of exonic mutations present within a tumor (after filtering on quality score and read depth (coverage, COV)) by the number of the exonic bases that had adequate quality and coverage in both the tumor sample (MPV score ≥ 10 and at least 5 reads) and paired normal sample (MPG score ≥ 10, MPG/COV ratio ≥ 0.5 and at least 5 reads). A Grubb’s test was used to calculate an approximate P value for each tumor to identify outliers. A uniform background mutation rate equal to the rate observed in the discovery phase was assumed, and a Poisson distribution function was used to calculate P values for the observed number of mutations in each gene. False discovery rates were calculated using the Benjamini-Hochberg method, correcting for 21,441 genes tested. This method is a simplified version of the CaMP (cancer mutation prevalence) scoring method, including subsequently suggested corrections.

**Functional enrichment analyses in silico.** The 304 somatically mutated protein-encoding genes identified and orthogonally validated in the discovery screen were analyzed for enriched functional groupings using the Database for Annotation Visualization and Integrated Discovery (DAVID) and Ingenuity Systems Pathway Analysis (IPA) in silico tools (see URLs). The Bonferroni, Benjamini and false discovery rate values, computed within DAVID, were assessed for significance.

**Determination of defective mismatch repair.** Tumor-normal DNA pairs were screened for the presence of microsatellite instability (MSI) using the Promega Microsatellite Instability Analysis System v1.2 according to the manufacturer’s instructions. All coding exons of MSH6 were PCR amplified and sequenced by Sanger sequencing.

**Cell lines.** Endometrial cancer cell lines (RL-95-2, HEC1A, HEC1B, KLE and ANC3A) were obtained from the American Type Culture Collection or the National Cancer Institute Developmental Therapeutics Program cell line repository. ARK1 and ARK2 serous endometrial cancer cell lines were kindly provided by A. Santin.

**Immunoblotting.** Cells were lysed in RIPA buffer (Thermo-Scientific) containing 1 mM sodium orthovanadate, 10 mM sodium fluoride and 1× protease inhibitor cocktail (Roche). Lysates were centrifuged and denatured at 95 °C in 2× SDS sample buffer (Sigma-Aldrich) before SDS-PAGE and transfer to PVDF membranes (Bio-Rad). The antibodies used included primary antibodies to CHD4 (Cell Signaling Technology, 4245) and β-actin (Sigma-Aldrich, A2228) and horseradish peroxidase (HRP)-conjugated secondary antibodies to mouse (Cell Signaling Technology, 7076) and horseradish peroxidase (HRP)-conjugated secondary antibodies to rabbit (Cell Signaling Technology, 7074). Immunoreactive proteins were visualized with enhanced chemiluminescence (Pierce).

---

60. Teer, J.K. et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res. 20, 1420–1431 (2010).
61. Teer, J.K., Green, E.D., Mullikin, J.C. & Biesecker, L.G. VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. Bioinformatics 28, 599–600 (2012).
62. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J.R. Stat. Soc. 57, 289–300 (1995).
63. Sjöblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. Science 314, 268–274 (2006).
64. Rubin, A.F. & Green, P. Comment on “The consensus coding sequences of human breast and colorectal cancers”. Science 317, 1500 (2007).