Integrated genomic characterization of endometrial carcinoma

The Cancer Genome Atlas Research Network

We performed an integrated genomic, transcriptomic and proteomic characterization of 373 endometrial carcinomas using array- and sequencing-based technologies. Uterine serous tumours and ~25% of high-grade endometrioid tumours had extensive copy number alterations, few DNA methylation changes, low oestrogen receptor/progesterone receptor levels, and frequent TP53 mutations. Most endometrioid tumours had few copy number alterations or TP53 mutations, but frequent mutations in PTEN, CTNNB1, PIK3CA, ARID1A and KRAS and novel mutations in the SWI/SNF chromatin remodelling complex gene ARID5B. A subset of endometrioid tumours that we identified had a markedly increased transversion mutation frequency and newly identified hotspot mutations in POLE. Our results classified endometrial cancers into four categories: POLE ultramutated, microsatellite instability hypermutated, copy-number low, and copy-number high. Uterine serous carcinomas share genomic features with ovarian serous and basal-like breast carcinomas. We demonstrated that the genomic features of endometrial carcinomas permit a reclassification that may affect post-surgical adjuvant treatment for women with aggressive tumours.

Somatic copy number alterations

Somatic copy number alterations (SCNAs) were assessed in 363 endometrial carcinomas. Unsupervised hierarchical clustering grouped the tumours into four clusters (Fig. 1a). The first three copy-number clusters were composed almost exclusively (97%) of endometrioid tumours without significant differences in tumour grades. Cluster 1 tumours were nearly devoid of broad SCNAs, averaging less than 0.5% genome alteration, with no significant recurrent events. Cluster 1 tumours also had significantly increased non-synonymous mutation rates compared to all others (median 7.2 × 10⁻⁶ versus 1.7 × 10⁻⁶ mutations per megabase (Mb), P < 0.001). Copy-number clusters 2 and 3 consisted mainly of endometrioid tumours, distinguished by more frequent 1q amplification in cluster 3 than cluster 2 (100% of cluster 3 tumours versus 33% of cluster 2 tumours) and worse progression-free survival (P = 0.003, log-rank versus clusters 1 and 2; Fig. 1b).

Most of the serous (50 out of 53; 94%) and mixed histology (8 out of 13; 62%) tumours clustered with 36 (12%) of the 289 endometrioid tumours, including 24% of grade 3 and 5% of grade 1 or 2, into copy-number cluster 4; a single group characterized by a very high degree of SCNAs (Supplementary Fig. 2.1; focal SCNAs with false discovery rate (FDR) < 0.15, and Supplementary Data 2.1). Cluster 4 tumours were characterized by significantly recurrent previously reported focal amplifications of the oncopgenes MYC (8q24.12), ERBB2 (17q12) and CCNE1 (19q12)13, and by SCNAs previously unreported in endometrial cancers including those containing FGFR3 (4p16.3) and SOX17 (8q11.23). Cluster 4 tumours also had frequent TP53 mutations (90%).
identified hotspot mutations in *POLE* at Pro286Arg and Val411Leu present in 13 (76%) of the 17 ultramutated samples. Significantly mutated genes (SMGs) identified at low FDRs (Q) in this subset included *PTEN* (94%, Q = 0), *PIK3R1* (65%, Q = 8.3 × 10^{-7}), *PIK3CA* (71%, Q = 9.1 × 10^{-4}), *FBXW7* (82%, Q = 1.4 × 10^{-4}), *KRAS* (53%, Q = 9.2 × 10^{-4}) and *POLE* (100%, Q = 4.2 × 10^{-2}). Mutation rates in *POLE* mutant endometrial and previously reported ultramutated colorectal tumours exceeded those found in any other lineage including lung cancer and melanoma. Germline susceptibility variants have been reported in *POLE* (Leu424Val and POLD1 Ser478Asn), but were not found in our endometrial normal exome-seq reads.

The MSI endometrioid tumours had a mutation frequency approximately tenfold greater than MSS endometrioid tumours, few SCNAs, frameshift deletions in *RPL22*, frequent non-synonymous *KRAS* mutations, and few mutations in *FBXW7*, *CTNNB1*, *PPP2R1A* and *TP53*. The MSS, copy-number low, endometrioid tumours had an unusually high frequency of *CTNNB1* mutations (52%); the only gene with a higher mutation frequency than the MSI samples. The copy-number high group contained all of the remaining serous cases and one-quarter of the grade 3 endometrioid cases. Most of these tumours had *TP53* mutations and a high frequency of *FBXW7* (22%, Q = 0) and *PPP2R1A* (22%, Q = 1.7 × 10^{-16}) mutations, previously reported as common in uterine serous but not endometrioid carcinomas. Thus, a subset of high-grade endometrioid tumours had similar SCNAs and mutation spectra as uterine serous carcinomas, suggesting that these patients might benefit from treatment approaches that parallel those for serous tumours.

There were 48 genes with differential mutation frequencies across the four groups (Fig. 2d and Supplementary Data 3.1). *ARID5B*, a member of the same AT-rich interaction domain (ARID) family as *ARID1A*, was more frequently mutated in MSI (23.1%) than in either MSS endometrioid (5.6%) or high SCNA serous tumours (0%), a novel finding for endometrial cancer. Frameshifting *RPL22* indels near a homopolymer at Lys 15 were almost exclusively found in the MSI group (36.9%). The *TP53* mutation frequency (>90%) in serous tumours differentiated them from the endometrioid subtypes (11.4%). However, many (10 out of 20; 50%) endometrioid tumours with a non-silent *TP53* mutation also had non-silent mutations in *PTEN*, compared to only 1 out of 39 (2.6%) serous tumours with non-silent *TP53* mutations. Although *TP53* mutations are not restricted to serous tumours, the co-existing *PTEN* mutations in the endometrioid cases suggest a distinct tumorigenic mechanism.

Comparisons of 66 SMGs between traditional histological subtypes are provided (Supplementary Methods 3), and SMGs across other subcohorts can be found in Supplementary Data 3.2. The spectrum of *PIK3CA* and *PTEN* mutations in endometrial cancer also differed from other solid tumours (Supplementary Methods 3). Integrated analysis may be useful for identifying histologically misclassified cases. For example, a single serous case was identified without a *TP53* mutation or extensive SCNAs and with a *KRAS* mutation and high mutation rate. After re-review of the histological section, the case was deemed consistent with a grade 3 endometrioid tumour, demonstrating how molecular analysis could reclassify tumour histology and potentially affect treatment decisions.

**Multiplatform subtype classifications**

All of the endometrial tumours were examined for messenger RNA expression (*n* = 333), protein expression (*n* = 293), microRNA expression (*n* = 367), and DNA methylation (*n* = 373) (Supplementary Methods 4–7). Unsupervised k-means clustering of mRNA expression from RNA sequencing identified three robust clusters termed ‘mitotic’, ‘hormonal’ and ‘immunoreactive’ (Supplementary Fig. 4.1) that were significantly correlated with the four integrated clusters; *POLE*, MSI, copy-number low and copy-number high (*P* < 0.0001). Supervised analysis identified signature genes of the *POLE* cluster (*n* = 17) most involved in cellular metabolism (Fig. 3a). Among the few signature genes
in the MSI cluster was decreased MLH1 mRNA expression, probably due to its promoter methylation. Increased progesterone receptor (PGR) expression was noted in the copy-number low cluster, suggesting responsiveness to hormonal therapy. The copy-number high cluster, which included most of the serous and serous-like endometrioid tumours, exhibited the greatest transcriptional activity exemplified by increased cell cycle deregulation (for example, CCNE1, PIK3CA, MYC and CDKN2A) and TP53 mutation (Supplementary Figs 4.2 and 4.3). This is consistent with reports that increased CDKN2A can distinguish serous from endometrioid carcinomas. Approximately 85% of cases in the copy-number high cluster shared membership with the ‘mitotic’ mRNA subtype.

Supervised clustering of the reverse phase protein array (RPPA) expression data was consistent with loss of function for many of the mutated genes (Fig. 3b). TP53 was frequently mutated in the copy-number high group ($P = 2.5 \times 10^{-27}$) and its protein expression was also increased, suggesting that these mutations are associated with increased expression. By contrast, PTEN ($P = 2.8 \times 10^{-19}$) and ARID1A ($P = 1.2 \times 10^{-6}$) had high mutation rates in the remaining groups, but their expression was decreased, suggesting inactivating mutations in both genes. The copy-number high group also had decreased levels of phospho-AKT, consistent with downregulation of the AKT pathway. The copy-number low group had raised RAD50 expression, which is associated with DNA repair, explaining some of the differences between the copy-number high and low groups. The POLE group had high expression of ASNS and CCNB1, whereas the MSI tumours had both high phospho-AKT and low PTEN expression.

Unsupervised clustering of DNA methylation data generated from Illumina Infinium DNA methylation arrays revealed four unique subtypes (MCI–4) that support the four integrative clusters. A heavily methylated subtype (MCI) reminiscent of the CpG island methylator phenotype (CIMP) described in colon cancers and glioblastomas was associated with the MSI subtype and attributable to promoter hypermethylation of MLH1. A serous-like cluster (MC3) with minimal DNA methylation changes was composed primarily of serous tumours and some endometrioid tumours (Supplementary Fig. 7.1) and contained most of the copy-number high tumours.

Integrative clustering using the iCluster framework returned two major clusters split primarily on serous and endometrioid histology highlighting TP53 mutations, lack of PTEN mutation and encompassing almost exclusively copy-number high tumours (Supplementary Fig. 8.1). We developed a new clustering algorithm, called SuperCluster, to derive overall subtypes based on sample cluster memberships across all data types (Supplementary Fig. 9.1). SuperCluster identified four clusters that generally confirmed the contributions of individual platforms to the overall integrated clusters. No major batch effects were identified for any platform (Supplementary Methods 10).

**Structural aberrations**
To identify somatic chromosomal aberrations, we performed low-pass, paired-end, whole-genome sequencing on 106 tumours with matched normals. We found recurrent translocations involving genes in several pathways including WNT, EGFR–RAS–MAPK, PI(3)K, protein kinase A, retinoblastoma and apoptosis. The most frequent translocations (5 out of 106) involved a member of the BCL family (BCL2, BCL7A, BCL9 and BCL2L11). Four of these were confirmed by identification of the translocation junction point and two were also confirmed by high-throughput RNA sequencing (RNA-Seq). In all cases the translocations result in in-frame fusions and are predicted to result in activation or increased expression of the BCL family members (Supplementary Fig. 3.2). Translocations involving members of the BCL family leading to reduced apoptosis have been
used MEMo25 to identify gene networks with mutually exclusive recurrently mutated, homozygously deleted, and amplified genes, we proteasomal degradation of in situ escence potential role for human epidermal growth factor receptor 2 (HER2)-expression in 25% of the serous or serous-like tumours, suggesting a ERBB2 mechanisms may be operative here.

Because of the high background mutation rate and small sample size, pathways in the four endometrial cancer integrated subgroups. Multiple platform data were integrated to identify recurrently altered described in other tumour types24 and our results suggest that similar mechanisms may be operative here.

Pathway alterations
Multiple platform data were integrated to identify recurrently altered pathways in the four endometrial cancer integrated subgroups. Because of the high background mutation rate and small sample size, we excluded the POLE subgroup from this analysis. Considering all recurrently mutated, homozygously deleted, and amplified genes, we used MEMo25 to identify gene networks with mutually exclusive alteration patterns in each subgroup. The most significant module was found in the copy-number low group and contained CTNNB1, KRAS and SOX17 (Fig. 4a). The very strong mutual exclusivity between mutations in these three genes suggests that alternative mechanisms activate WNT signalling in endometrioid endometrial cancer. Activating KRAS mutations have been shown to increase the stability of β-catenin via glycogen synthase kinase 3β (GSK-3β), leading to an alternative mechanism of β-catenin activation other than adenomatous polyposis colo degradation26. SOX17, which mediates proteasomal degradation of β-catenin27,28, is mutated exclusively in the copy-number low group (8%) at recurrent positions (Ala96Gly and Ser403Ile) not previously described. Other genes with mutually exclusive alteration patterns in this module were FBXW7, CTNNB1 and CTNNB1 (ref. 29). ERBB2 was focally amplified with protein overexpression in 25% of the serous or serous-like tumours, suggesting a potential role for human epidermal growth factor receptor 2 (HER2)-targeted inhibitors. A small clinical trial of trastuzumab found no activity in endometrial carcinoma, but accrued few HER2 fluorescence in situ hybridization (FISH)-amplified serous carcinomas30.

PIK3CA and PIK3R1 mutations were frequent and showed a strong tendency for mutual exclusivity in all subgroups, but unlike other tumour types, they co-occurred with PTEN mutations in the MSI and copy-number low subgroups as previously reported3,9 (Fig. 4b).

The copy-number high subgroup showed mutual exclusivity between alterations of all three genes. Overall, 93% of endometrioid tumours had mutations that suggested potential for targeted therapy with PI(3)K/AKT pathway inhibitors.

Consensus clustering of copy number, mRNA expression and pathway interaction data for 324 samples yielded five PARADIGM clusters with distinct pathway activation patterns31 (Fig. 4c and Supplementary Methods 11). PARADIGM cluster 1 had the lowest level of MYC pathway activation and highest level of WNT pathway activation, consistent with its composition of copy-number low cases having frequent CTNNB1 mutations. PARADIGM cluster 3 was composed predominantly of the copy-number high cases, with relatively high MYC/ MAX signalling but low oestrogen receptor/FOXA1 signalling and p53 activity. Only TP53 truncation and not nonsense mutations were implicated as loss-of-function mutations, suggesting different classes of p53 mutations may have distinct signalling consequences. PARADIGM cluster 5 was enriched for hormone receptor expression.

Comparison to ovarian and breast cancers
The clinical and pathologic features of uterine serous carcinoma and high-grade serous ovarian carcinoma (HGSOC) are quite similar. HGSOC shares many similar molecular features with basal-like breast carcinoma22. Focal SCNA patterns were similar between these three tumour subtypes and unsupervised clustering identified relatedness (Fig. 5a and Supplementary Fig. 12.1). Supervised analysis of transcriptome data sets showed high correlation between tumour subtypes (Supplementary Fig. 12.2). The MC3 DNA methylation subtype with minimal DNA methylation changes was also similar to basal-like breast and HGSOCs (Supplementary Fig. 12.3). A high frequency of TP53 mutations is shared across these tumour subtypes (uterine serous, 91%; HGSOC, 96%; basal-like breast, 84%)33,34, as is the very low frequency of PTEN mutations (uterine serous, 2%; HGSOC, 1%; basal-like breast, 1%). Differences included a higher frequency of FBXW7, PPP2R1A and PIK3CA mutations in uterine serous compared to basal-like breast and HGSOCs (Fig. 5b). We showed that uterine serous carcinomas share many molecular features with both HGSOCs and basal-like breast carcinomas, despite more frequent mutations, suggesting new opportunities for overlapping treatment paradigms.

Discussion
This integrated genomic and proteomic analysis of 373 endometrial cancers provides insights into disease biology and diagnostic classification that could have immediate therapeutic application. Our analysis identified four new groups of tumours based on integrated genomic data, including a novel POLE subtype in ~10% of endometrioid tumours. Ultrahigh somatic mutation frequency, MSS, and common, newly identified hotspot mutations in the exomeclase domain of POLE characterize this subtype. SCNAs add a layer of resolution, revealing that most endometrioid tumours have few SCNAs, most serous and serous-like tumours exhibit extensive SCNAs, and the extent of SCNA roughly correlates with progression-free survival.

Endometrial cancer has more frequent mutations in the PI(3)K/ AKT pathway than any other tumour type studied by The Cancer Genome Atlas (TCGA) so far. Endometrioid endometrial carcinomas share many characteristics with colorectal carcinoma including a high frequency of MSI (40% and 11%, respectively), POLE mutations (7% and 3%, respectively) leading to ultrahigh mutation rates, and frequent activation of WNT/CTNNB1 signallng; yet endometrial carcinomas have novel exclusivity of KRAS and CTNNB1 mutations and a distinct mechanism of pathway activation. Uterine serous carcinomas share many similar characteristics with basal-like breast and HGSOCs; three tumour types with high-frequency non-silent TP53 mutations and extensive SCNA. However, the high frequency of PIK3CA, FBXW7, PPP2R1A and ARID1A mutations in uterine serous carcinomas are not found in basal-like breast and HGSOCs. The frequency of mutations in PIK3CA, FBXW7 and PPP2R1A was ~30% higher than in a recently

©2013 Macmillan Publishers Limited. All rights reserved
reported study of 76 uterine serous carcinomas, but similar to another study. Uterine serous carcinomas have ERBB2 amplification in 27% of tumours and PIK3CA mutations in 42%, which provide translational opportunities for targeted therapeutics.

Early stage type I endometrioid tumours are often treated with adjuvant radiotherapy, whereas similarly staged type II serous tumours are treated with chemotherapy. High-grade serous and endometrioid endometrial carcinomas are difficult to subtype correctly, and intra-observer concordance among specialty pathologists is low. Our molecular characterization data demonstrate that ~25% of tumours classified as high-grade endometrioid by pathologists have a molecular phenotype similar to uterine serous carcinomas, including frequent TP53 mutations and extensive SCNA. The compelling similarities between this subset of endometrioid tumours and uterine serous carcinomas suggest that genomic-based classification may lead to improved management of these patients. Clinicians should carefully consider treating copy-number-altered endometrioid patients with chemotherapy rather than adjuvant radiotherapy and formally test such hypotheses in prospective clinical trials. Furthermore, the marked molecular differences between endometrioid and serous-like tumours suggest that these tumours warrant separate clinical trials to develop the independent treatment paradigms that have improved outcomes in other tumour types, such as breast cancer.

**METHODS SUMMARY**

Biospecimens were obtained from 373 patients after Institutional Review Board-approved consents. DNA and RNA were co-isolated using a modified AllPrep kit (Qiagen). We used Affymetrix SNP 6.0 microarrays to detect SCNA in 363 samples and GISTIC analysis to identify recurrent events. The exomes of 248 samples and GISTIC analysis to identify recurrent events. The exomes of 248 tumors were sequenced to a read-depth of at least 20. We performed low-pass whole-genome sequencing on 107 tumours to a mean depth of $40 \times$. Consensus clustering was used to analyse miRNA, mRNA, RPPA and methylation data with methods previously described. Integrated cross-platform analyses were performed using MEs, iCluster and PARADIGM.

Received 10 December 2012; accepted 21 March 2013.


