Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.0806514106">http://dx.doi.org/10.1073/pnas.0806514106</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/50250">http://hdl.handle.net/1721.1/50250</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation


aDepartment of Obstetrics and Gynecology, Haukeland University Hospital, 5021 Bergen, Norway; bDepartment of Clinical Medicine, University of Bergen, 5020 Bergen, Norway; cThe Harvard and MIT Division of Health Sciences and Technology, Cambridge, MA 02142; dThe Broad Institute of Harvard and MIT, Cambridge, MA 02142; eThe Gade Institute, Section for Pathology, University of Bergen, 5020 Bergen, Norway; fDepartment of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115; gDepartment of Pathology, Haukeland University Hospital, 5020 Bergen, Norway; hMax Planck-Institute for Neurological Research, and Klaus-Joachim-Zülch Laboratories of the Max Planck Society and the Medical Faculty of the University of Cologne, 509315 Cologne, Germany; iDepartment of Medicine, Brigham and Women’s Hospital, Boston, MA 02115; jDepartment of Medicine, Harvard Medical School, Boston, MA 02115; kDepartment for Informatics and Computational Biology Unit, University of Bergen, 5020 Bergen, Norway; lChemical Genomics Center of the Max Planck Society, 44227 Dortmund, Germany; mDepartment of Internal Medicine and Center of Integrated Oncology, University of Cologne, 50931 Cologne, Germany; nCenter for Cancer Genome Discovery, Dana–Farber Cancer Institute, Boston, MA 02115; oNovartis Institutes for Biomedical Research, Cambridge, MA 02139; and pDepartment of Pathology, Harvard Medical School, Boston, MA 02115

Edited by Joe W. Gray, Lawrence Berkeley National Laboratory, Berkeley, CA, and accepted by the Editorial Board January 26, 2009 (received for review July 4, 2008)

Although 75% of endometrial cancers are treated at an early stage, 15% to 20% of these recur. We performed an integrated analysis of genome-wide expression and copy-number data for primary endometrial carcinomas with extensive clinical and histopathological data to detect features predictive of recurrent disease. Unsupervised analysis of the expression data distinguished 2 major clusters with strikingly different phenotypes, including significant differences in disease-free survival. To identify possible mechanisms for these differences, we performed a global genomic survey of amplifications, deletions, and loss of heterozygosity, which identified 11 significantly amplified and 13 significantly deleted regions. Amplifications of 3q26.32 harboring the oncogene PIK3CA were associated with poor prognosis and segregated with the aggressive transcriptional cluster. Moreover, samples with PIK3CA amplification carried signatures associated with in vitro activation of PI3 kinase (PI3K), a signature that was shared by aggressive tumors without PIK3CA amplification. Tumors with loss of PTEN expression or PIK3CA overexpression that did not have PIK3CA amplification also shared the PI3K activation signature, high protein expression of the PI3K pathway member STMN1, and an aggressive phenotype in test and validation datasets. However, mutations of PTEN or PIK3CA were not associated with the same expression profile or aggressive phenotype. STMN1 expression had independent prognostic value. The results affirm the utility of systematic characterization of the cancer genome in clinically annotated specimens and suggest the particular importance of the PI3K pathway in patients who have aggressive endometrial cancer.

The needs for effective systemic therapies and for reliable prognostic markers have been addressed only partly. The most common basis for determining risk of recurrent disease has been the categorization of endometrial cancer into 2 subtypes. The majority are type I, associated with good prognosis, low stage and grade, and endometrioid histology. In contrast, type II cancers are characterized by high stage and grade, nonendometrioid histology, and poor prognosis. The prognostic value of this distinction is limited, however, because up to 20% of type I cancers recur, and half of type II cancers do not (2).

The molecular basis of the distinction between type I and II cancer is understood only partially. Type I cancer is associated with hyperestrogenic risk factors, is more often estrogen and progesterone receptor positive, diploid, microsatellite unstable, and KRAS or PTEN mutant. Type II cancer is more often aneuploid and harbors alterations in CDKN2A, TP53, and ERBB2 (3). Such molecular alterations are of prognostic value (4) but have not provided a basis for improved therapy (3). Hormone receptor status influences the choice of treatment in metastatic disease, but most aggressive tumors are receptor negative (1).

We hypothesized that tumors with an aggressive phenotype are likely to be distinguished by underlying genetic alterations reflected in distinct transcriptional signatures, and we investigated whether tumors that recur share transcriptional signatures that suggest shared underlying genetic alterations.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.W.G. is a guest editor invited by the Editorial Board.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.gov/geo (accession no. GSE14860).

1To whom correspondence may be addressed. E-mail: helga.salvesen@uib.no, lars.akslen@gades.uib.no, or rameen_beroukhim@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0806514106/DCSupplemental.


Results and Discussion

Unsupervised Analysis of Expression Profiles Distinguishes Aggressive Tumors. We collected genome-wide expression and clinical and histopathologic data from a random sampling of 57 endometrial carcinomas in a population-based tissue bank of gynecologic cancer in Hordaland County, Norway. The characteristics of these patients were not significantly different from those of all patients diagnosed with endometrial carcinoma in a 10-year period from the same region (4) [supporting information (SI) Table S1].

An unsupervised analysis of these data distinguished 2 major groups of tumors (clusters 1 and 2, Fig. L4). Significant analysis of microarrays (SAM) (5) between these clusters identified a 2-fold or higher change for 138 significant genes, of which 64 were up-regulated and 74 downregulated in cluster 2 (data not shown). A set of 29 genes, validated by quantitative RT-PCR, predicted the clusters with 100% accuracy (Fig. 1A and B).

The 2 clusters had strikingly different clinical and histopathologic characteristics. Cluster 2 contained more aggressive tumors, with higher International Federation of Gynecology and Obstetrics (FIGO) stage, histologic grade, number of mitoses, presence of non-endometrioid histologic subtype, tumor necrosis, and vascular invasion, \( P < 0.001 \) for the presence of any of these characteristics; Fig. 1A, Table S2). Cluster 2 contained almost all the type II tumors \( (P < 0.001) \) (Fig. 1A, Table S2) but it also contains almost 1/3 of the type I tumors, and these tumors have more vascular invasion, necrosis, and frequent mitoses than the type I tumors in cluster 1 \( (P = 0.01) \). The 29-gene summary set also was significantly correlated with aggressive cancer (Table S2).

Most prominently, patients who had tumors in cluster 2 had significantly poorer recurrence-free survival \( (P = 0.05, \text{Fig. 1C}) \). Segregation into cluster 2 predicted recurrence better than FIGO stage, histologic subtype, or receptor status and slightly less well than grade (Table S3) but did not exhibit independent prognostic impact, probably because of the limited number of cases and events.

Regions of Significant Amplification, Deletion, and Loss of Heterozygosity. To identify the underlying somatic changes distinguishing aggressive tumors with the cluster 2 signature, we performed a genome-wide survey of copy-number changes and loss of heterozygosity (LOH) among 84 tumors. The majority of tumors exhibit a small number of amplifications (median of 4 in each tumor) and even fewer deletions (median of 1). Nevertheless, virtually every region of the genome is amplified or deleted in at least 1 tumor (Fig. 2A).

To distinguish copy-number changes associated with endometrial cancer from potentially random events, we applied the statistical method Genomic Identification of Significant Targets in Cancer (GISTIC) (6). GISTIC assigns each region of the genome 2 G-scores (Fig. 2B), each representing the combined frequency and amplitude of either local amplifications or deletions. It then compares these scores with similar scores generated from random permutations of the data to determine false discovery rate q-values (7), representing the likelihood of obtaining the observed G-scores from chance events alone. The G-scores tend to be larger for amplifications than deletions (Fig. 2B) because of the greater prevalence of amplifications (Fig. 2A). Conversely, deletions attain statistical significance (using a q-value threshold of 0.25, green lines in Fig. 2B) at lower prevalence because of their overall infrequency.

We found 11 significantly amplified and 13 significantly deleted regions of the genome (Table S4), including several that had been noted previously (8, 9). For each we selected the peak region (ie, the highest frequency and amplitude of events) as the region most likely to contain a cancer gene target. Known oncogenes are located within these peaks for 8 amplified regions,
and known tumor suppressors are located within deletion peaks on chromosomes 1 and 3 (Table S4), but functional data tying any of these genes to endometrial carcinogenesis are lacking. Also, 14 regions contain no known cancer genes. These peaks usually represent infrequent events (<17% of tumors), with the exception of 1q amplification, in which the gene target is unclear because of the large size of the amplicon. The consistent breadth of this amplicon, in fact, may suggest more than 1 target (6). LOH generally reflects deletions, with the exception of prevalent copy-neutral LOH on 10q (Fig. S1) containing the known endometrial tumor suppressor PTEN (10).

**Amplifications of KRAS and PIK3CA Associate with Poor Prognosis.** Among the 11 significant amplifications, only 2 (3q26.32 and 12p12.1) are associated with recurrence-free survival (in both cases poor survival) after correction for multiple hypotheses (Table S5, Fig. 3A). We considered only amplifications because of the low prevalence of deletions. Amplification of 3q26.32 (3qamp) also is associated with non-endometrioid (44% vs. 11% prevalence; \( P = 0.02 \)) and high-grade (\( P < 0.001 \)) histology. The association between 12p12.1 amplification and poor survival is surprising, because mutations of KRAS, which is within the peak region, are known to associate with better survival. However, we sequenced KRAS in 64 tumors and found mutant KRAS in none of the 12p12.1-amplified samples, although mutations were seen in 4 unamplified samples. Amplification of 12p12.1 is also associated with high grade (\( P = 0.02 \)) and FIGO stage (\( P = 0.04 \)). Although 3q26.32 and 12p12.1 tended to be amplified in the same tumors (\( P = 0.03 \)), they usually did not coincide. We directed further analyses at 3qamp because all the samples with this amplification segregated into expression cluster 2 (\( P = 0.01 \)) (Fig. 1A), suggesting that the amplification could be associated with the cluster 2 transcriptional profile (SI Text Note 1).

**Integrated Analyses Associate Markers of PI3 Kinase Activation with Aggressive Cancer.** We wanted to investigate whether 3qamp leads to an aggressive phenotype through activation of PIK3CA. Although PIK3CA has not been shown to be the 3qamp target, we suspected it for 4 reasons: (i) PIK3CA is 1 of 36 genes within the peak region (indeed, PIK3CA amplifications were noted previously; see refs. 11, 12); (ii) tumors with 3qamp overexpress PIK3CA compared with unamplified tumors (\( P = 0.003 \); Fig. 3B); (iii) similar amplifications in ovarian cancer act through PIK3CA (13); and (iv) the PI3 kinase (PI3K) pathway frequently is aberrant in endometrial cancer (11), including point mutations in PIK3CA (14).

We therefore looked for wider effects of PIK3CA activation in the transcriptome of tumors with 3qamp. We used published data (15) from cell lines transfected with mutationally activated PIK3CA to define a PI3K activation score (PI3K score), representing the expression levels of genes that correlate with activated PIK3CA (see Methods). Tumors with 3qamp scored higher than unamplified samples (\( P = 0.05 \); Fig. 3C). However, the impact of this finding is limited by its borderline statistical significance and by the possibility that the PI3K score may reflect PI3K activation only in the model systems in which it was measured but not generally.

We therefore sought to corroborate this finding by seeing whether samples with 3qamp have an expression profile opposite that induced by PI3K pathway inhibition. To that end, we queried the 50 most overexpressed and underexpressed genes in samples with 3qamp relative to unamplified samples (Fig. S2A) using the
Connectivity Map (16). Among 164 small molecules represented
in the Connectivity Map, the PI3K inhibitor LY-294002 (17) had
an expression signature most significantly anticorrelated with
the 3qamp signature (Fig. S2 B and C, P/H11005 0.003). LY-294002 is
known to bind to additional kinases (18), raising the possibility
that this anticorrelation results from nonspecific effects. The
anticorrelation between the 3qamp signature and inhibitors of
adenylate cyclase and Hsp90 (Fig. S2 B) also suggests potentially
complex effects of the amplicon. Nevertheless, the findings that
the 3qamp signature correlates with a PI3K activation signature
and anticorrelates with the signature of a PI3K inhibitor support
the hypothesis that an effect of 3qamp may be to increase PI3K
activity.

We also validated the correlation between PIK3CA amplification
and the PI3K score in an independent expression dataset. We first inferred amplification of 3q26–27 from local gene
expression levels, as reflected in a “functional amplification”
(FA) score (19). As expected, samples determined to have 3qamp
by SNP array analysis also had high 3q26–27 FA scores (P/H11005
0.001 and 0.005, respectively; Fig. S3 D and E) and with high values of the 29-gene summary
predictor for membership in cluster 2 (P = 3 × 10^{-4} and P
= 0.004, respectively; Fig. S3 F and G).

The finding that both PIK3CA amplification and the cluster 2
expression profile indicate aggressive tumors, coupled with the
association between PIK3CA amplification and the in vitro PI3K
activation signature, suggested that the broader set of aggressive
tumors in cluster 2 might share the in vitro PI3K activation
signature. This conjecture seems to be true: tumors in cluster 2
without PIK3CA amplification have PI3K scores significantly
higher than the scores of tumors in cluster 1 (P < 0.001) and
equal to those in tumors with amplification of PIK3CA (Fig. 3D).
Moreover, the cluster 2 signature is highly anticorrelated with
the signature of treatment with LY-294002 (P = 0.02; Fig. S4 A
and B). Furthermore, tumors with high PI3K scores are associated
with poor survival (P = 0.03, Fig. S4 C) and other markers
of aggressive phenotype in both the test and validation datasets (P = 0.01 and P = 0.001, respectively).

A possible cause of overexpression of the PI3K activation signature among tumors without PIK3CA amplification is decreased expression of the downstream PI3K pathway member PTEN. Decreased PTEN expression was associated with increased PI3K3 scores in both our test and validation datasets (P < 0.001 and P < 0.03, respectively), regardless of PIK3CA amplification status. Decreased PTEN expression also was associated with markers of aggressive disease (P = 0.02), in agreement with prior studies (21).

Conversely, among the 45 tumors with expression data that we sequenced for PTEN, mutations did not associate with high PI3K scores (P = 0.6; Fig. S4D). On the contrary, we observed more mutations in the nonaggressive cluster 1 than in cluster 2 (P = 0.04; Fig. 1A), in line with prior studies associating PTEN mutations with indolent disease (22, 23).

Overexpression and mutation of PIK3CA also seem to have different implications. We found significantly higher PIK3CA expression in tumors with aggressive features, including those without PIK3CA amplification (P = 0.05 and P = 0.0009 among test and validation data, respectively). However, among the 41 tumors with expression data that we sequenced for PIK3CA, mutations did not associate with high PI3K3 scores (P = 0.8; Fig. S4E) or features of aggressive disease (P = 0.5). We also cannot confirm the finding that exon 20 mutations correspond to aggressive tumors (24) (data not shown). These results were surprising in light of evidence that overexpression of mutated, but not wild-type, PIK3CA leads to transformation (25, 26) and suggest either that PIK3CA suffers from prevalent cryptic mutations or that transformation assays did not capture the effects of wild-type PIK3CA overexpression in human tumors. Although PIK3CA mutations previously have been noted primarily in endometrioid cancers (11, 12), we also find no correlation with histologic subtype (P = 1).

Expression of the PI3 Kinase Pathway Member STMN1 Is an Independent Prognostic Indicator. The suggestion that PI3K activation associates with poor prognosis suggested that measuring PI3K activity might improve prognostication of localized endometrial cancer. Expression of STMN1 previously has been shown to correlate with PI3K activity in breast cancer (27) and can be measured by immunohistochemistry in paraffin-embedded tissue. We therefore measured STMN1 expression by immunohistochemistry (SI Methods and Fig. S4 F and G) in 72 tumors, including 66 with SNP array and 53 with expression data (Fig. 1A). Although STMN1 is not a member of our PI3K activation signature, STMN1 expression correlated with PI3K scores (P = 0.05). High STMN1 expression also correlated with PIK3CA amplification (P = 0.04) and overexpression (P = 0.04) and with segregation in cluster 2 (P = 0.03), supporting our prior associations between these features and PI3K pathway activation.

High STMN1 expression also was associated with poor recurrence-free survival in our original tumor set (P = 0.006) and with poor recurrence-free (P = 0.01) and overall (P = 0.01) survival in a validation set of 241 tumors from a population-based series of all endometrial carcinomas in Hordaland County from 1981 to 1990 (4, 21). In both tumor sets, STMN1 expression correlated with grade, mitotic rate, presence of necrosis or vascular invasion, and type II status (Table S6). Nevertheless, across all 313 cases (except for 5 cases with missing clinical data), high STMN1 expression was a prognostic indicator independent of FIGO stage, histologic subtype, grade, and age (P = 0.004; Fig. 3G, Table S7). In particular, high STMN1 expression was associated with poor prognosis in the otherwise low-risk endometroid subgroup (P = 0.007, data not shown).

Ultimately, the goals of integrated genomic analyses of localized tumors are to enable development of clinical assays to distinguish aggressive tumors requiring therapy beyond resection and of effective therapeutics for such tumors. We have shown that both transcriptional and copy-number profiles of endometrial tumors contain prognostic information that is partly reflected in expression levels of PIK3CA, in vitro PI3K activation, PTEN, and STMN1. We conclude that PTEN and PIK3CA mutations seem to have different transcriptional and phenotypic correlates than changes in expression of these genes. These results suggest that further investigation of the specific consequences of mutation and altered expression is warranted. They also emphasize the potential utility of clinical assays for PI3K pathway activation to identify patients who have aggressive disease and the particular relevance of therapeutics that inhibit this pathway.

Aside from the PI3K pathway, our general survey of chromosomal changes in endometrial carcinoma also identified approximately 20 other regions of significant copy-number change. Most of these copy-number changes involve tens to hundreds of genes, so even when known oncogenes or tumor suppressors are within the regions most affected by these copy-number changes, the genomic data are ambiguous as to the actual target. In many cases, including amplification of 3q26.32, the size of these events suggests the possibility of multiple targets. Moreover, functional data tying even known oncogenes and tumor suppressors to carcinogenesis in model systems of endometrial cancer are lacking, for the most part. The limited number of significant regions of copy-number change suggests that comprehensive, systematic experiments to identify these oncogenes and tumor suppressors in endometrial cancer are feasible. Such experiments are likely to point to therapeutic targets for women with all stages of endometrial carcinoma.

Materials and Methods

Patient Series. For the primary investigation series, primary endometrial carcinomas were frozen immediately during hysterectomies conducted from 2001 to 2003. All samples were reviewed by a pathologist according to published criteria (28). Treatment included bilateral salpingo-oophorectomy and pelvic lymphadenectomy. Adjuvant therapy was recommended for patients who had disease of FIGO surgical stage IIB or higher or nonendometrioid histology. Patients were followed from primary surgery until June, 2007 or death, with a median follow-up for survivors of 3.6 years (range, 0.8–5.5). Deaths not attributable to endometrial cancer were censored. No patient was lost to follow-up.

RNA Analysis. RNA was extracted from biopsies with at least 50% (usually >80%) tumor content using the RNeasy kit (Qiagen). Quality and yield were assessed by agarose electrophoresis, the Agilent Bioanalyser 2100, and spectrophotometry. RNA was prepared in 2 batches and hybridized to Agilent 21K and 22K arrays, respectively, according to manufacturer’s instructions (www.agilent.com). Arrays were scanned using the Agilent Microarray Scanner Bundle.

Signal intensities were determined using J-Express (www.molmine.com) and filtered to remove genes with signal intensities below 2 SD over background in either channel (G5, Gy3) in more than 30% of samples. Batch adjustment was performed as described previously (29). Genes were mean-centered across the tumor set.

Hierarchical clustering was performed using the 5000 genes with highest variance using weighted average linkage and Pearson correlation as similarity measures. Clustering with more or fewer genes gave stable results (data not shown). A SAM analysis using these clusters as class labels identified 138 significantly changed genes in primary tumors, which were selected for their combined discriminatory power as described in SI Methods. Messenger RNA levels for these 29 genes and PTEN were validated by quantitative PCR using the TaqMan Low Density Array (Applied Biosystems) according to manufacturer’s instructions (29).

For the external dataset (Affymetrix U133 + 2 arrays), individual probes were sequence-matched against Aceview (NCBI35) (19) to construct transcript-level probe sets. Summary expression levels then were derived by batch normalization across samples via RMA (30).

The PI3K score was obtained by comparing previously published expression data of 9 replicate transfections of activated PIK3CA with 5 GFP controls (15).
and includes the 495 genes surpassing a Bonferroni-corrected 2-sided t test p-value of 0.05. To evaluate this signature, expression data for each gene were normalized to a common mean and scaled to the same SD. For each sample, the activation score is the sum of genes significantly up-regulated in the cells with activated PIK3CA (relative to the cells with GFP control) minus genes significantly down-regulated in those cells.

**DNA Analysis.** Genomic DNA was extracted from surgically dissected, fresh-frozen primary tumors and from 9 cell lines: Ishiakawa, HeC1A, KLE, AN3-CA, EFE184, MFE-280, MFE-296, MFE-319, and RL-95–2. Tumors were needle dissected to ensure 80% purity. PIK3CA, KRAS, and PTEN were sequenced as previously described (31). Genomic DNA was analyzed by SNP arrays interrogating 116,204 SNP loci (Affymetrix) and the GISTIC algorithm, as previously described (6). SNP, gene, and cytogenetic band locations are based on the hg16 (July 2003) genome build (genome.ucsc.edu).

**Statistics.** For relations of molecular data to clinical phenotype, Pearson’s χ² (χ²), Fisher’s exact-, Mann-Whitney-, or Kruskal-Wallis tests were used as appropriate. P-values represent 2-sided tests except when testing the 1-sided hypothesis that 3qAMP correlates with measures of PI3K activation. Univariate survival analyses were performed by the Kaplan-Meier method. The log-rank (Mantel-Cox) test with Bonferroni correction was used to compare survival curves for different categories. Variables with significant impact on survival (P ≤ 0.05) were examined further by log-minus-log plot before incorporation in the Cox proportional hazards regression model.

**ACKNOWLEDGMENTS.** The authors thank Gerd Lilian Halleth, Bendik Nordanger, Britt Edvardsen, Randi Nygaard, Beth Johannessen, and Hua My Hoang for technical assistance. This study was supported by Helse Vest Grants 911005, 911351, and 911069, the University of Bergen, the Norwegian Cancer Society Grants D94070/04, D0019/01, H501–2006-0446 (Harald Andersen legat), the Research Council of Norway Grants 154924/Y0 and 163920/V50, Department of Defense Grants PC040638 and PC061642, National Institutes of Health Grant KO8CA122833, and the Dana-Farber Harvard Cancer Center Prostate Specialized Program in Research Excellence. This research was approved by the Norwegian Data Inspectorate (961478–2), Norwegian Social Sciences Data Services (15501), and local ethics committee (REKII nr. 052.01).