2014

Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers

Bo Zhang
Washington University School of Medicine in St. Louis

Xiao Yun Xing
Washington University School of Medicine in St. Louis

Jing Li
Washington University School of Medicine in St. Louis

Rebecca F. Lowdon
Washington University School of Medicine in St. Louis

Yan Zhou
Northeast Normal University

See next page for additional authors

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Authors

Bo Zhang, Xiao Yun Xing, Jing Li, Rebecca F. Lowdon, Yan Zhou, Nan Lin, Baoxue Zhang, Vasavi Sundaram, Katherine B. Chiappinelli, Ian S. Hagemann, David G. Mutch, Paul J. Goodfellow, and Ting Wang

BMC
Genomics

RESEARCH ARTICLE

Open Access

Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers

Bo Zhang[1†], XiaoYun Xing[1†], Jing Li[1,2†], Rebecca F Lowdon[1], Yan Zhou[3], Nan Lin[4], Baoxue Zhang[3], Vasavi Sundaram[1], Katherine B Chiappinelli[5], Ian S Hagemann[6], David G Mutch[7], Paul J Goodfellow[8*] and Ting Wang[1*]

Abstract

Background: Aberrant DNA methylation is a hallmark of many cancers. Classically there are two types of endometrial cancer, endometrioid adenocarcinoma (EAC), or Type I, and uterine papillary serous carcinoma (UPSC), or Type II. However, the whole genome DNA methylation changes in these two classical types of endometrial cancer is still unknown.

Results: Here we described complete genome-wide DNA methylome maps of EAC, UPSC, and normal endometrium by applying a combined strategy of methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylation-sensitive restriction enzyme digestion sequencing (MRE-seq). We discovered distinct genome-wide DNA methylation patterns in EAC and UPSC: 27,009 and 15,676 recurrent differentially methylated regions (DMRs) were identified respectively, compared with normal endometrium. Over 80% of DMRs were in intergenic and intronic regions. The majority of these DMRs were not interrogated on the commonly used Infinium 450K array platform. Large-scale demethylation of chromosome X was detected in UPSC, accompanied by decreased *XIST* expression. Importantly, we discovered that the majority of the DMRs harbored promoter or enhancer functions and are specifically associated with genes related to uterine development and disease. Among these, abnormal methylation of transposable elements (TEs) may provide a novel mechanism to deregulate normal endometrium-specific enhancers derived from specific TEs.

Conclusions: DNA methylation changes are an important signature of endometrial cancer and regulate gene expression by affecting not only proximal promoters but also distal enhancers.

Background

Endometrial cancer is the most common gynecologic malignancy in the United States, with an estimated 47,130 new cases and 8,010 deaths annually [1]. Most cases of endometrial cancer are endometrioid adenocarcinoma (EAC), are of low grade, and are diagnosed at an early stage, with a 5-year survival rate of greater than 85% [1].

Uterine papillary serous carcinoma (UPSC), an aggressive histologic subtype of endometrial cancer, represents less than 10% of all endometrial cancers. However, UPSC accounts for more than 50% of recurrences and deaths attributed to endometrial carcinoma [2,3]. EAC commonly displays near-diploid karyotypes, microsatellite instability, and mutations in the *PTEN*, *KRAS*, and *CTNNB1* (β-catenin) genes. UPSC is characterized by frequent *TP53* mutation, *Her-2/neu* overexpression, and an aneuploid karyotype [2,3].

Like many malignancies, endometrial cancer is a complex disease driven by both genetic, epigenetic and environmental factors. DNA methylation has long been implicated in the development and progression of tumors in various

* Correspondence: Paul.Goodfellow@osumc.edu; twang@genetics.wustl.edu
†Equal contributors
[8]The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA
[1]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA
Full list of author information is available at the end of the article

tissue types [4-6]. An overall reduction in total 5-methylcytosine level and focal hypermethylation in CpG islands near tumor-suppressor gene transcriptional start sites were found in many different types of cancers [7-10]. Among endometrial cancers, promoters of important tumor suppressor genes including *MLH1*, *RASSF1A*, *PTEN*, and *APC*, were found to be hypermethylated in EAC. However, methylation status of these genes was largely unaltered in UPSC [11-16]. The divergent hypermethylation between these tumor types might be caused by the significantly increased expression of DNA methyltransferases *DNMT1* and *DNMT3B* observed in EAC, whereas expressions of these enzymes were unchanged or even decreased in UPSC [14,17].

Recently, The Cancer Genome Atlas Consortium (TCGA) profiled DNA methylation of more than 300 endometrial cancer samples using array-based DNA methylation platforms (HumanMethylation27 BeadChip and HumanMethylation450 BeadChip), which interrogate 27,578 CpG sites and 482,421 CpG sites respectively [18]. Here we took a complementary approach to identify DNA methylation changes unique to the two endometrial cancer subtypes in an unbiased fashion. Our strategy allowed us to systematically measure DNA methylation levels of more than 20 million CpG sites in the cancer genome in an unbiased fashion with respect to intergenic and intronic regions, including repetitive regions derived from transposable elements. These regions are classically not measured by array-based methods, such as the ones employed by TCGA [19,20].

We generated complete DNA methylome maps for endometrioid adenocarcinoma (EAC, three samples), uterine papillary serous carcinoma (UPSC, three samples), and normal endometrium (ten pooled samples) by integrating data from methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylation-sensitive restriction enzyme sequencing (MRE-seq) [19-24]. Comparative analysis of these seven DNA methylomes identified cancer-associated differentially methylated regions (DMRs) and distinct EAC and UPSC genomic DNA methylation patterns. Many methylation changes were found in CpG island shores and these changes were predicted to affect expression of nearby genes. We found demethylation across large domains on the X chromosome in UPSC accompanied by decreased *XIST* expression. We also identified methylation differences at numerous miRNA gene promoters that correlated with expression changes of the associated miRNAs. We discovered that cancer type-specific DMRs were enriched for not only promoters, but also for enhancer elements. Moreover, specific transposable elements (TEs), a rich genomic resource for potential enhancers [23,25-32], were affected by both *de novo* methylation and demethylation in cancer samples. Together, these

results suggest that DNA methylation broadly impacts cancer gene expression via regulation of promoters as well as enhancers and TEs.

Results

### Distinct global and focal DNA methylation signatures in two types of endometrial cancer

DNA methylation changes have been reported in many types of cancers and exhibit strong tissue-specific and tumor type-specific characteristics [8,33,34]. DNA methylation changes in the two subtypes of endometrial cancers, endometrioid adenocarcinoma (EAC) and uterine papillary serous carcinoma (UPSC), were recently reported by The Cancer Genome Atlas Consortium [18]. This seminal study provided the first insight into DNA methylation changes at a genome-wide scale for this important cancer using array-based platforms. Here we took a different strategy, which is complementary to that of TCGA, to deeply profile complete DNA methylomes of a small number of tumor specimens using newly developed sequencing-based epigenomics technology [19,20]. This strategy allowed us to systematically discover DNA methylation changes in cancer genomes without the biases imposed by microarray methods. Findings based on analyzing a small number of deeply profiled DNA methylomes (i.e., a discovery panel) can then be validated using array-based approaches, which can be applied to a much bigger number of samples (i.e., a validation panel). In this study, we generated complete DNA methylomes from six tumor samples (3 EACs and 3 UPSCs) and 1 pooled normal endometrium sample by applying methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylation-sensitive restriction enzyme sequencing (MRE-seq) [22]. We were able to comprehensively measure DNA methylation levels of more than 20 million CpGs for each sample (Additional file 1: Table S1). Sequencing data from the tumor samples were normalized by their genomic copy number variations identified using the Affymetrix SNP6.0 array.

Previous tumor DNA methylome studies have shown global demethylation accompanied by focal hypermethylation in many cancers [7-10]. To understand the global pattern of DNA methylation alterations in endometrial cancer, we directly compared the overall distribution of signal density from MeDIP-seq data at 5 kb resolution across the seven methylomes. As expected, cancer samples showed both hypomethylated and hypermethylated changes compared to normal endometrium (Additional file 2: Figure S1A). On average, 4.7% of the genome became hypomethylated, and 1.5% became hypermethylated in cancer (Additional file 1: Table S2). EAC and UPSC did not exhibit dramatic differences in this analysis.

Our main goal was to identify local differentially methylated regions (DMRs) between cancer and normal endometrium in a genome-wide fashion. To this end, we

developed and applied M&M, a novel algorithm that detects DMRs by integrating MeDIP-seq and MRE-seq data [19]. In total, we identified 27,009 EAC-associated DMRs and 15,676 UPSC-associated DMRs, with 6,606 DMRs in common between the two types of endometrial cancer (EC-shared DMRs) (Figure 1A and Additional file 1: Table S3). A complete list of these DMRs and links to the WashU Epigenome Browser were provided on the accompanying website (Methods). In EAC, 68% of DMRs (18,294) were hypermethylated and 32% (8,715) were hypomethylated relative to normal endometrium. In contrast, 40% (6,296) of the UPSC-associated DMRs were hypermethylated and 60% (9,380) were hypomethylated (Figure 1B).

EAC and UPSC shared 4,597 hypermethylated and 2,009 hypomethylated DMRs (EC-shared DMRs). Previous studies have reported that two DNA methyltransferases (*DNMT1* and *DNMT3B*) were more highly expressed in EAC but repressed in UPSC [14,17]. Upregulation of the DNMTs was confirmed by qRT-PCR in the three EAC samples we assayed. However, in the UPSC samples, *DNMT1*, *DNMT3A* and *DNMT3B* were not repressed, but had variable expression across samples (Additional file 2: Figure S1B). Consistent with our result, mRNA-seq data from TCGA also did not support previous studies: the three DNA methyltransfrases exhibited significantly higher expression in patients with
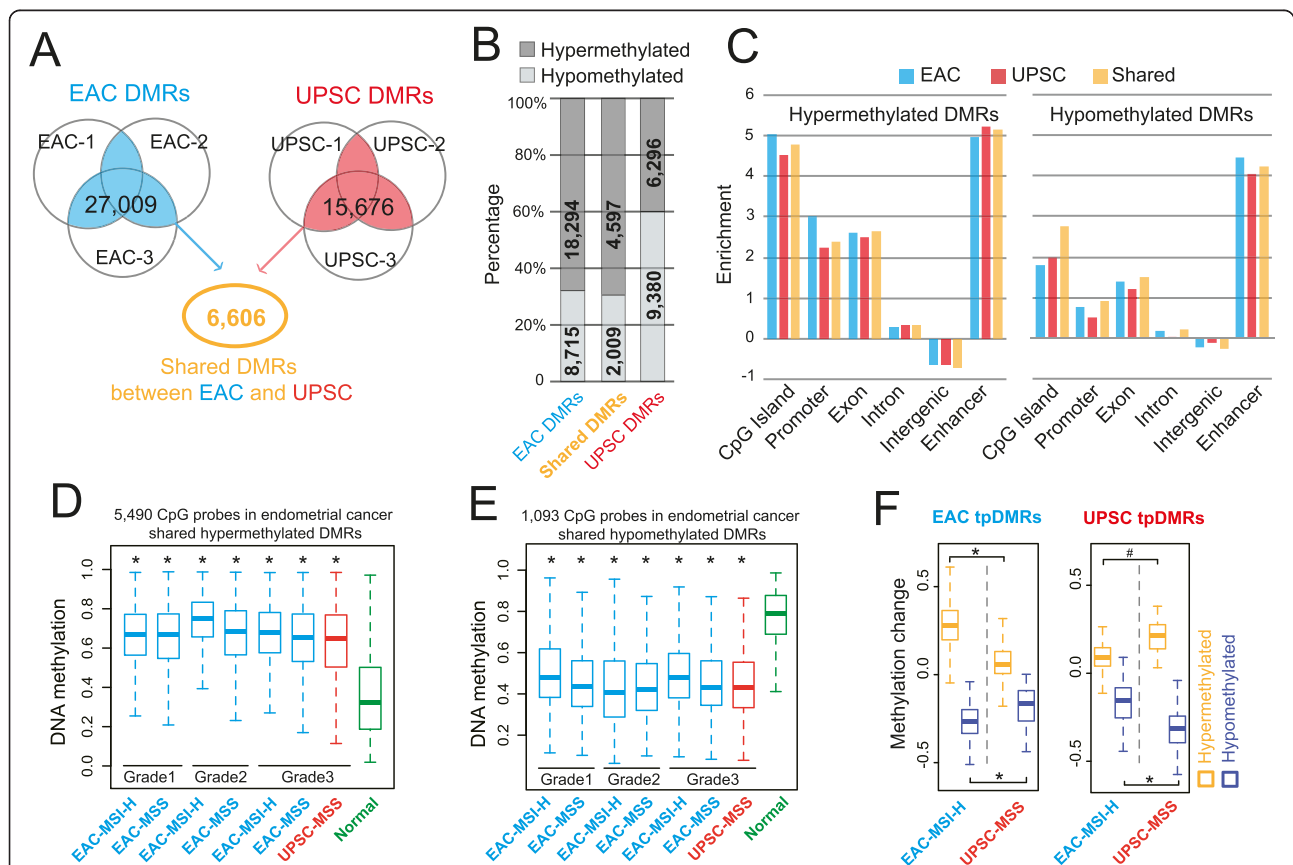
Figure 1 Identification, annotation, and validation of DMRs in endometrioid adenocarcinoma (EAC) and uterine papillary serous carcinoma (UPSC). (A) 27,009 DMRs were identified in at least 2 EAC samples (blue: EAC-DMRs), 15,676 DMRs were identified in at least 2 UPSC samples (red: UPSC-DMRs). 6,606 DMRs were identified in both EAC and UPSC as endometrial cancer (EC) shared DMRs (orange: EC-shared DMRs). **(B)** Percentage of hypermethylated (dark gray) and hypomethylated (light gray) EAC-, UPSC-, and EC-shared DMRs. **(C)** Genomic feature enrichment for hypermethylated DMRs (left) and hypomethylated DMRs (right). **(D)** The DNA methylation level of 5,490 CpGs located within 2,454 hypermethylated EC-shared DMRs in pre-classified (grade, microsatellite state, and subtype) endometrial cancer samples (blue: EAC; red: UPSC) and normal controls (green) in TCGA Infinium 450K data. DNA methylation level of each CpG site was averaged within the pre-classified group. Each boxplot represents the distribution of averaged methylation levels of 5,490 CpGs in cancer and controls. MSI-H: Microsatellite instability-high. MSS: Microsatellite stable. Asterisk: *P* < 1e-21 (Mann–Whitney *U* test). **(E)** The DNA methylation level of 1,093 CpGs located within 576 hypomethylated EC-shared DMRs in the same cohort (blue: EAC; red: UPSC) and controls (green) as in **(D)**. Asterisk: *P* < 1e-21 (Mann–Whitney *U* test). **(F)** Validation of EAC and UPSC type-preferred DMRs (tpDMRs) by TCGA Infinium 450K data. **Left**: DNA methylation change of 2,002 EAC tpDMRs in grade 3 MSI-H-EAC (blue) and MSS-UPSC (red), compared to normal controls. **Right**: DNA methylation change of 147 UPSC tpDMRs in same samples of UPSC (red) and EAC (blue). DNA methylation changes were calculated by subtracting the averaged methylation level of the normal controls from the methylation level of each DMR in each cancer sample. Asterisk: *P* < 1e-21, octothorpe: *P* <1e-3 (Mann–Whitney *U* test).

UPSC as compared with non-tumor tissue (Additional file 2: Figure S1C).

The genomic distribution of cancer-associated DMRs was highly non-random (Additional file 1: Table S4). Over 70% of hypermethylated DMRs and 35% of hypomethylated DMRs were located in open chromatin regions and transcription factor binding sites, as defined by ENCODE [35] (Additional file 2: Figure S2A). Hypermethylated DMRs were strongly enriched for gene-related features, including CpG islands, promoters, and exons (Figure 1C). This is consistent with the notion that transcription-related regions, in particular CpG islands, are frequent targets of DNA methylation changes in cancer. However, these DMRs (CpG islands at promoters, exons, and UTRs) only accounted for 40% of the total hypermethylated DMRs in EACs. 60% of hypermethylated DMRs and 88% of hypomethylated DMRs were located in introns and intergenic regions (Additional file 2: Figure S2B). Strikingly, when we examined the relationship between these DMRs and annotated enhancers, we found that both hypermethylated and hypomethylated DMRs were strongly enriched for enhancer elements (Figure 1C). This suggests that in addition to promoters and CpG islands, enhancer elements can be a preferential target of DNA methylation alteration in cancers.

Next, we validated that the endometrial cancer-associated DMRs were recurrent by using Infinium 450K data produced by The Cancer Genome Atlas (TCGA). The Infinium 450K platform contains probes that span 6,583 CpG sites in 40% of 6,606 EC-shared DMRs (Additional file 2: Figure S2C). On average, over 85% EC-shared DMRs exhibited significant DNA methylation differences (Methods) between cancer patients and normal controls (Additional file 2: Figure S2D). Patients of all pathologic groups displayed consistent DNA methylation levels within these regions, robustly confirming our initial findings (Figure 1D and 1E). Importantly, these recurrent EC-shared DMRs exhibited hypo- or hypermethylation in low-grade EAC patients to the same degree as that of high-grade patients (Figure 1D and 1E), suggesting that abnormal DNA methylation is shared across tumors of different grades, and can potentially be detected in early stage of endometrial carcinogenesis. Thus, these EC-shared, recurrent DNA methylation changes could represent a unique epigenetic signature for endometrial cancers.

In addition, we defined cancer type-preferred DMRs (tpDMRs, see Methods). 3,443 EAC tpDMRs were present in all three EAC samples but not in any UPSC sample, and another 720 UPSC tpDMRs were present in all three UPSC samples but not in any EAC sample. Of these, 57% of EAC tpDMR regions and 21% of UPSC tpDMR regions contained Infinium 450K CpG probes (Additional file 2: Figure S2C), which allowed validation of our discovery using TCGA data. Among these regions, 86% of EAC tpDMRs and 89% UPSC tpDMRs showed significant DNA methylation differences (Methods) between cancer patents and normal controls (Additional file 2: Figure S2D). Additionally, EAC and UPSC tpDMRs were enriched in proximity to genes with different gene ontology terms (Additional file 2: Figure S3A). Among TCGA patient samples, tpDMRs were confirmed and found to exhibit cancer type-specific hypo- or hypermethylation status, suggesting that these cancer type-preferred DMRs are recurrent with respect to each cancer type (Figure 1F).

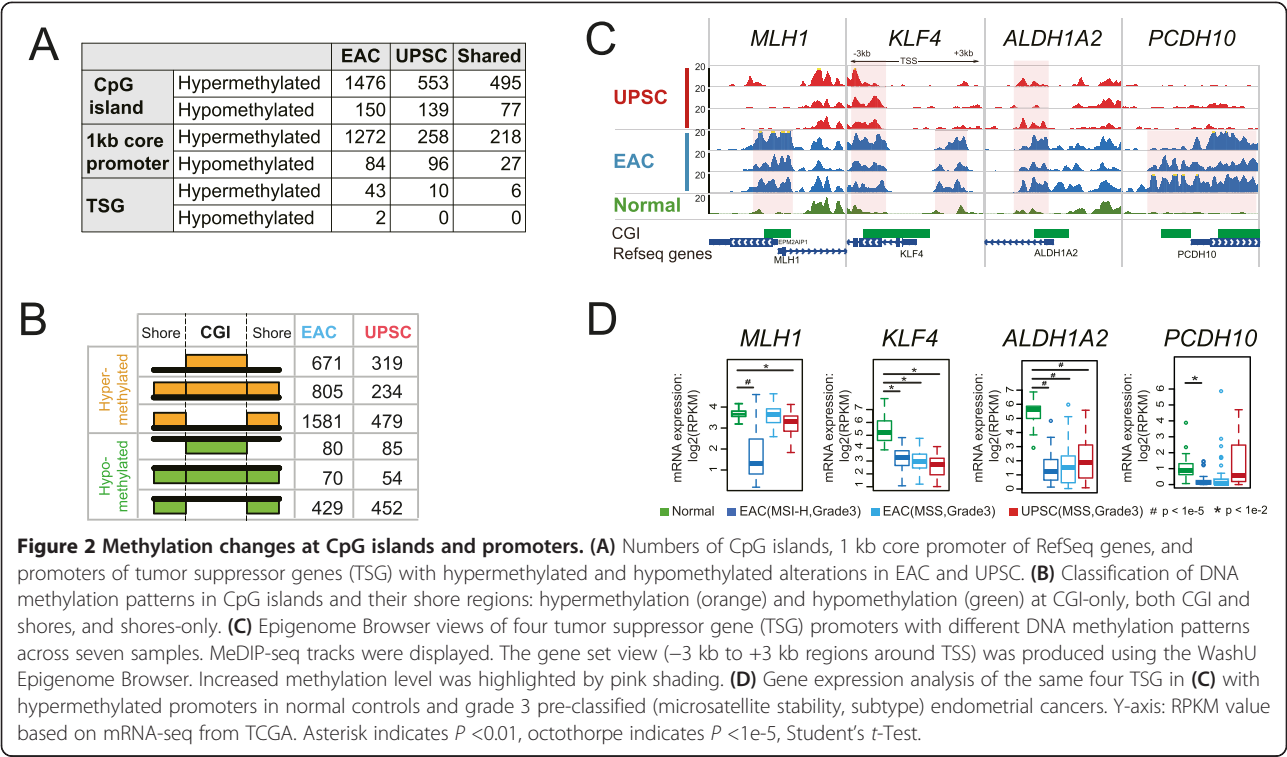### Distinct DNA methylation patterns at CpG islands, shores, and gene promoters

There are about 28,000 CpG islands (CGIs) in the human genome. About 70% of gene promoters are associated with CGIs, and ~50% of CGIs contain annotated transcription start sites (TSS). The majority of promoter CGIs are unmethylated in somatic cells, although a significant portion of gene body CGIs exhibit a conserved, tissue-specific DNA methylation pattern [22,35,36]. DNA methylation of CpG islands is often correlated with the repression of the associated genes [37]. CpG islands frequently gain methylation in cancer [38]. Hypermethylation of CGIs, especially those associated with tumor suppressor genes, is considered a hallmark of many types of cancers [39,40].

Examination of CGI DNA methylation changes in the two types of endometrial cancers revealed distinct signatures for EAC and UPSC. Compared to normal endometrium, we defined 1,476 and 553 hypermethylated CGIs for EAC and UPSC respectively (Figure 2A). 495 CGIs were hypermethylated in both types of endometrial cancer. 150 CGIs in EAC and 139 CGIs in UPSC exhibited reduced DNA methylation compared to normal endometrial cells. The two types of endometrial cancer can be clearly distinguished based on their CGI DNA methylation pattern (Additional file 2: Figure S2E).

Compared to promoter CGIs, non-promoter CGIs were more likely to undergo DNA methylation changes in endometrial cancers (Table 1). 532 CGIs (in EAC) and 193 CGIs (in UPSC) out of ~6,000 intergenic CGIs, and 448 CGIs (in EAC) and 223 CGIs (in UPSC) out of ~8,000 gene body CGIs, were hypermethylated. In contrast, of the ~13,000 promoter CGIs, only 496 and 137 were hypermethylated in EAC and UPSC respectively. Hypomethylated CGIs were also predominantly located in non-promoter regions, with 52 (in EAC) and 47 (in UPSC) in intergenic regions, and 94 (in EAC) and 85 (in UPSC) in gene bodies. Only 4 and 7 promoter CGIs were hypomethylated in EAC and UPSC respectively.

Next, we examined DNA methylation alterations in CpG island shores (the flanking 1 kb around CGIs). Abnormal DNA methylation of CGI shores was first reported in colon cancer and found to affect expression of the nearby genes [8]. In endometrial cancer, CGI

Figure 2 Methylation changes at CpG islands and promoters. (A) Numbers of CpG islands, 1 kb core promoter of RefSeq genes, and promoters of tumor suppressor genes (TSG) with hypermethylated and hypomethylated alterations in EAC and UPSC. **(B)** Classification of DNA methylation patterns in CpG islands and their shore regions: hypermethylation (orange) and hypomethylation (green) at CGI-only, both CGI and shores, and shores-only. **(C)** Epigenome Browser views of four tumor suppressor gene (TSG) promoters with different DNA methylation patterns across seven samples. MeDIP-seq tracks were displayed. The gene set view (−3 kb to +3 kb regions around TSS) was produced using the WashU Epigenome Browser. Increased methylation level was highlighted by pink shading. **(D)** Gene expression analysis of the same four TSG in **(C)** with hypermethylated promoters in normal controls and grade 3 pre-classified (microsatellite stability, subtype) endometrial cancers. Y-axis: RPKM value based on mRNA-seq from TCGA. Asterisk indicates $P < 0.01$, octothorpe indicates $P < 1e-5$, Student's *t*-Test.

shores exhibited dynamic DNA methylation changes that were independent of changes in the CpG islands (Figure 2B). We further classified DNA methylation changes with respect to their patterns in both CGIs and corresponding shores. In EAC, 805 of 1,476 hypermethylated CGIs also displayed hypermethylation in their shore regions, while 234 of 553 UPSC hypermethylated CGIs had accompanying hypermethylated shores. Strikingly, 1,581 EAC CGIs and 479 UPSC CGIs without significant DNA methylation changes had flanking hypermethylated shores. Additionally, 429 CGIs in EAC and 452 CGIs in UPSC were specifically hypomethylated in their shore regions but not in CGIs themselves (Figure 2B). In EAC, 791 genes, including lncRNA genes, were found around these 1,581 CGI shore-only hypermethylated CGIs, and 116 genes were close to 429 shore-only hypomethylated CGIs.

We then examined the DNA methylation status of 1 kb core promoters around the TSS of annotated RefSeq genes (Figure 2A). 1272 and 84 genes were hypermethylated or hypomethylated, respectively, at their core promoters in EAC. These numbers were 258 and 96 in UPSC. This result highlights the difference between the two types of endometrial cancers. 218 gene promoters were hypermethylated in both tumor types, while 27 gene promoters were hypomethylated in both. Gene function enrichment analysis of these promoters returned no significantly enriched terms related to endometrial tissue function (Additional file 2: Figure S3B).

We observed 43 tumor suppressor genes (TSGs) [41] with hypermethylated promoters in EAC, which is consistent with our expectation that the silencing of tumor suppressor genes is an important component of carcinogenesis. In contrast, in UPSC we found only 10 tumor suppressor genes with hypermethylated promoters (Additional file 2: Figure S4 and S5). As expected, the majority of these tumor suppressor genes were repressed in EAC (Additional file 2: Figure S4 and S5). Surprisingly, some of these tumor suppressor genes were also repressed in UPSC, even though the promoter regions were

Table 1 DNA methylation alterations of CpG islands in EAC and UPSC

		Intergenic regions	Promoter	Gene-body
CpG islands (CGI) in human genome		6236	13212	8269
EAC	Hypermethylated CGI	532	496	448
	Hypomethylated CGI	52	4	94
UPSC	Hypermethylated CGI	193	137	223
	Hypermethylated CGI	47	7	85

unmethylated as in normal endometrium. For example, a well-known marker, the promoter of the *MLH1* gene, is usually highly methylated in microsatellite instability-high (MSI-H) EAC, and was unmethylated in UPSC (Figure 2C and Additional file 2: Figure S5C). However, RNA-seq data from TCGA suggested that expression of *MLH1* was also repressed in some UPSC patients (Figure 2D). As expected, hypermethylation of tumor suppressor gene promoters was strongly associated with gene repression. For example, *KLF4* had a hypermethylated CGI shore and gene body and was repressed in three groups of endometrial cancer patients (Figure 2D). Similarly, the promoter of *ALDH1A2* was hypermethylated and expression of this gene was repressed in both EAC and UPSC. In contrast, the promoter of *PCDH10* was specifically methylated in EAC, and this gene was respressed in EAC only (Figure 2C and 2D). Unexpectedly, we found the core promoter regions of two tumor suppressor genes, *CDH1* and *SFN*, were demethylated in three EAC samples (Additional file 2: Figure S6A). Interestingly, these two genes were significantly upregulated in all three groups of endometrial cancer patients from TCGA (Additional file 2: Figure S6B). *CDH1* and *SFN* are important tumor suppressor genes whose loss of function has been implicated in many types of cancers [42-45]. Thus, their demethylation and increased expression underscores the complexity of regulatory events in endometrial cancer.

Loss of X chromosome DNA methylation and *XIST* expression in UPSC

One prominent signature distinguishing EAC and UPSC was DNA methylation differences on the X chromosome. More than 95% of DMRs on chromosome X were hypomethylated in UPSC (Figure 3A and Additional file 1: Table S5). In contrast, only ~16% of chromosome X DMRs were hypomethylated in EAC. Plotting MeDIP-seq signal and MRE-seq signal at a large scale (500 kb resolution) revealed loss of DNA methylation across the whole X chromosome in UPSC, as evidenced by reduced MeDIP-seq signal (~20%) and increased MRE-seq signal (~60%) (Figure 3B). EAC also exhibited loss of DNA methylation across the X chromosome, but to a much lesser degree than UPSC. In contrast, similar analysis of any autosome (e.g., chromosome 10) revealed no global change of DNA methylation in either EAC or UPSC (Additional file 2: Figure S7). We further examined expression of *XIST*, the lncRNA critical for establishing inactivation of the X chromosome, both in published data and by performing qRT-PCR on our samples. *XIST* showed reduced expression in both EAC and UPSC. Interestingly, in microsatellite stable (MSS) type EAC samples, the expression level of *XIST* was inversely correlated with the tumor grade (cor = −0.19) (Figure 3C). qRT-PCR also confirmed that *XIST* was down-regulated in four of six

endometrial cancer samples in our study (Additional file 2: Figure S1D). Finally, we examined the methylation level of 9,620 CpGs (~0.8% of total CpGs on the X chromosome) in TCGA endometrial cancer samples profiled on the Infinium 450K platform (Figure 3D). As expected, compared to normal controls (mean 0.52), the patients with grade 3 UPSC cancer type showed significantly lower DNA methylation level (mean 0.45, p-value < 1e-7), and the patients with grade 3 MSI-H EAC cancer type did not show difference on these CpG sites (mean 0.51, p-value = 0.21). Interestingly, the TCGA samples with grade 3 MSS EAC cancer type also show lightly reduced DNA methylation level on the X chromosome (mean 0.48, p-value < 1e-5). These results suggest that global hypomethylation of X chromosome and repressed *XIST* expression constitute a strong signature for UPSC cancers.

DNA methylation changes of non-coding RNA genes

Our genome-wide DNA methylome maps also allowed us to comprehensively assess DNA methylation changes around non-coding RNA genes, including miRNA and lncRNA. Epigenetic changes in promoters of non-coding RNA genes could result in changes of their expression levels, which subsequently could result in expression changes of downstream protein coding genes in endometrial cancer [46] and in changes of the epigenetic landscape [47].

We examined DNA methylation patterns in promoter regions around the TSS of miRNA gene clusters [48]. In EAC, promoters of 24 miRNA gene clusters exhibited increased DNA methylation, while 6 miRNA gene clusters exhibited reduced DNA methylation. These two numbers in UPSC were 9 and 2 (Additional file 2: Figure S8A). Some of these miRNAs were reported to have significant expression changes in endometrial cancer [46,49-51]. By examining TCGA miRNA-seq data, we found that expression levels of certain miRNAs correlated with DNA methylation changes in their promoters (Additional file 2: Figure S8B and S8C). The microRNA cluster *MIR200B-MIR200A-MIR429* and cluster *MIR200C-MIR141* showed significant demethylation in endometrial cancer, and the expression levels of these five microRNAs were significantly increased in three groups of endometrial cancer patients (Additional file 2: Figure S8C and S8D). The *MIR200* family was considered a tumor suppressor because of their inhibitory role in epithelial-mesenchymal transition (EMT) in bladder cancer and breast cancer cell lines [52-54]. Moreover, high expression of this cluster was tightly correlated with survival rate in those tumor types, and overexpression of these miRNAs inhibited ovarian cancer cell motility [55]. In addition, the *MIR200* family was reported to show significant upregulation in endometrial carcinomas [50,56,57]. The exact functions of *MIR200* family members in gynecologic cancers, especially
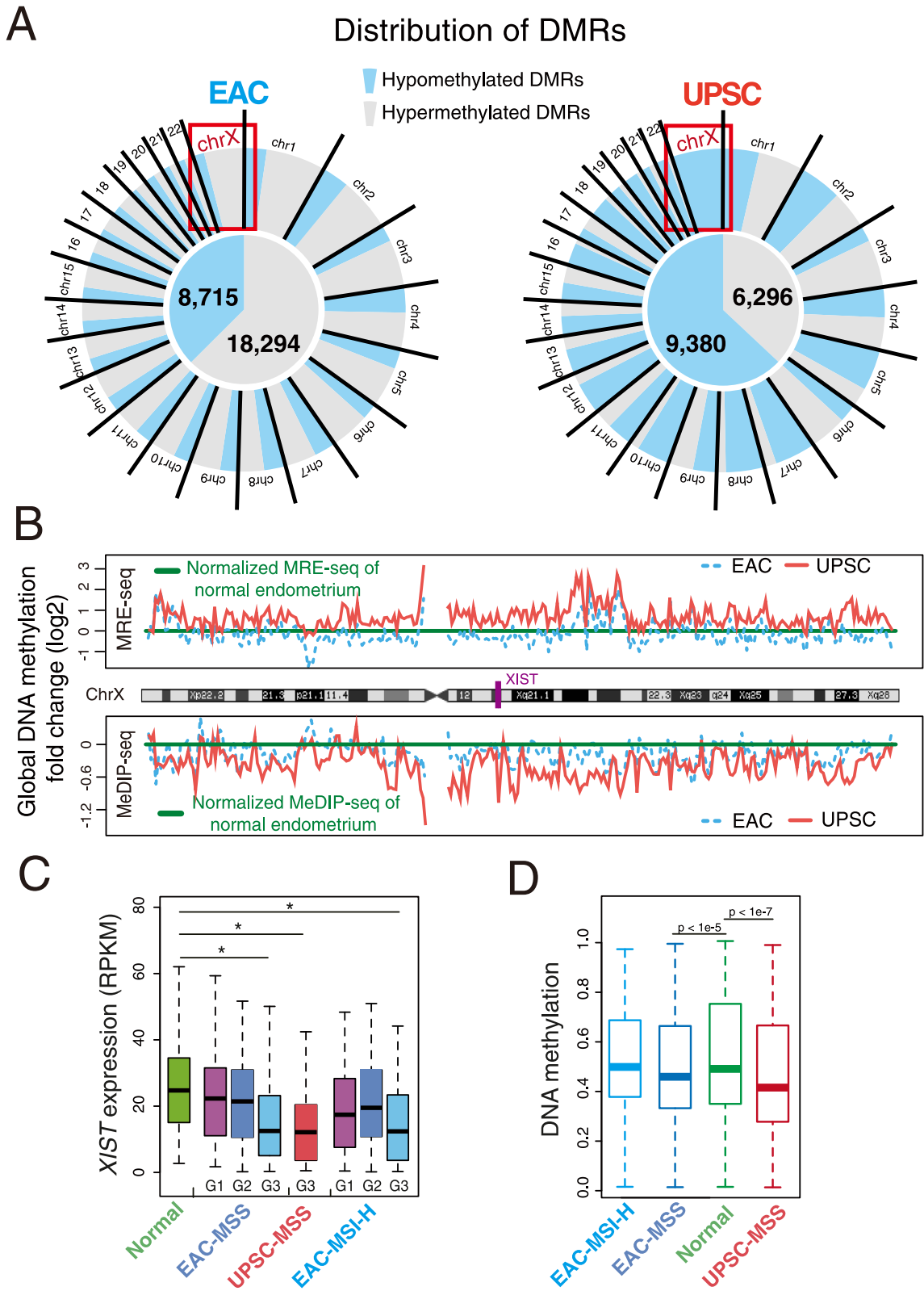
Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 Loss of DNA methylation on the X chromosome. (A) Distribution of hypermethylated (gray) and hypomethylated DMRs (blue) on all chromosomes in EAC (left) and UPSC (right). **(B)** Global DNA methylation changes on the X chromosome. MeDIP-seq and MRE-seq RPKM values of seven samples were calculated at 500 kb resolution across the X chromosome. The averaged RPKM fold changes (cancer/normal) of each type (3 EACs and 3 UPSCs) were log2-transformed and plotted along with the X chromosome coordinate. **(C)** *XIST* expression in normal controls and pre-classified (grade: G1, G2, G3; microsatellite state, and subtype) endometrial cancer samples. Y-axis: calculated RPKM value based on mRNA-seq from TCGA. Asterisk indicates *P* <0.01, Student's *t*-test. **(D)** DNA methylation distribution of 9,620 CpG sites on the X chromosome in pre-classified grade 3 (microsatellite state, and subtype) uterine corpus endometrial cancer samples (light blue: MSI-H type EAC patients; dark blue: MSS type EAC patients; red: MSS type UPSC patients. Infinium 450K platform) and normal controls (green, Infinium 450K platform). For each CpG site, the averaged DNA methylation level was calculated within the pre-classified group, and each boxplot represents distribution of the averaged methylation level of all 9,620 CpGs of the cancer groups and the normal controls. CpG sites with no value in any sample were removed. MSI-H: Microsatellite instability-high. MSS: Microsatellite stability. The Mann–Whitney *U* test was performed respectively for each cancer group when compared to normal controls.

in endometrial carcinogenesis, still need to be determined. Interestingly, even with strongly hypermethylated promoters, some miRNAs exhibited upregulated expression level, including *MIR25*, *MIR93*, *MIR99B*, *MIR324*, *MIR106B*, and *MIR3074* (Additional file 2: Figure S8E). The regulatory relationship between promoter hypermethylation and microRNA expression is likely complex and needs further investigation.

Hundreds of lncRNA genes [58] also underwent DNA methylation changes in our endometrial samples (Additional file 2: Figure S8A). Specifically, 621 and 168 lncRNA genes had hypermethylated promoters in EAC and UPSC respectively, compared to normal endometrium. Of these, 144 were in common between the two cancer types. The numbers of hypomethylated lncRNA genes were 207 for EAC and 245 for UPSC, with 68 in common. With the exception of classic examples such as *XIST*, *H19*, and *HOTAIR*, the widespread regulatory roles of lncRNAs are only recently being revealed [59]. The roles of lncRNAs and consequences of their abnormal DNA methylation in endometrial carcinogenesis remain to be elucidated. However, interesting candidate genes emerged from our analysis. For example, maternally expressed gene 3 (*MEG3*), a tumor suppressor non-coding RNA [60], was highly methylated at its promoter region in EAC. This was associated with significantly downregulated *MEG3* mRNA (Additional file 2: Figure S9). Re-expression of *MEG3* was found to induce *TP53* activation, and inhibited tumor cell proliferation in culture and colony formation in soft agar [60-64].

DNA methylation changes were enriched in functional promoter and enhancer elements
The majority of the differentially methylated regions were located in intergenic or intragenic regions, while a small fraction (11% in EACs, and 5% in UPSCs) overlapped directly with annotated gene promoters. Previous work reported that normal tissue developmental processes could be used to model tumor developmental states [65]. Here we examined endometrial cancer DMRs in the context of their epigenetic status in embryonic stem cell (ESC) H1 to

map the dynamics of DNA methylation changes [19]. We defined four DNA methylation patterns across ESC, normal endometrium, and endometrial cancer (Figure 4A): **MMU**, **M**ethylated in ES cells, **M**ethylated in normal endometrium, but **U**nmethylated (or hypomethylated) in cancer; **MUM**, **M**ethylated in ES cells, **U**nmethylated in normal endometrium, and **M**ethylated in cancer; **UMU**, **U**nmethylated in ES cells, **M**ethylated in normal endometrium, and **U**nmethylated in cancer; and **UUM**, **U**nmethylated in ES cells, **U**nmethylated in normal endometrium, and **M**ethylated in cancer. We reasoned that the genomic distribution and functions of genes associated with different patterns might help to reveal insights into endometrial carcinogenesis in the context of normal development. We annotated the potential regulatory function of these different categories of DMRs with the chromatin state maps defined by chromHMM using nine cell lines [66]. Interestingly, the majority of the **UUM** DMRs were annotated as promoters, whereas the majority of the **MUM** DMRs were annotated as enhancers (Figure 4B).

We then examined the functional enrichment of genes near these four different categories of DMRs (Figure 4C and Additional file 2: Figure S10A). Significantly, genes near pattern **MUM** enhancer DMRs in both types of cancer were strongly enriched for terms related to uterine development and disease. Consistent with our previous annotation of tissue-specific enhancers [19], this result suggested that these DMRs might encode uterine-specific regulatory elements that became unmethylated and activated during normal development and differentiation. Their hypermethylation in endometrial cancer might contribute to the loss of endometrial tissue type identity and to the gain of stem cell-like characteristics often observed in cancers. In contrast, pattern **MMU** DMRs enriched for functions specific to tissues unrelated to endometrium. For example, in EAC, genes related to hindbrain morphogenesis and mammary gland epithelium development were enriched near **MMU** DMRs. In UPSC, we found enrichment of genes related to lung, muscle, and kidney development. In addition, we found genes related to estrogen response, consistent
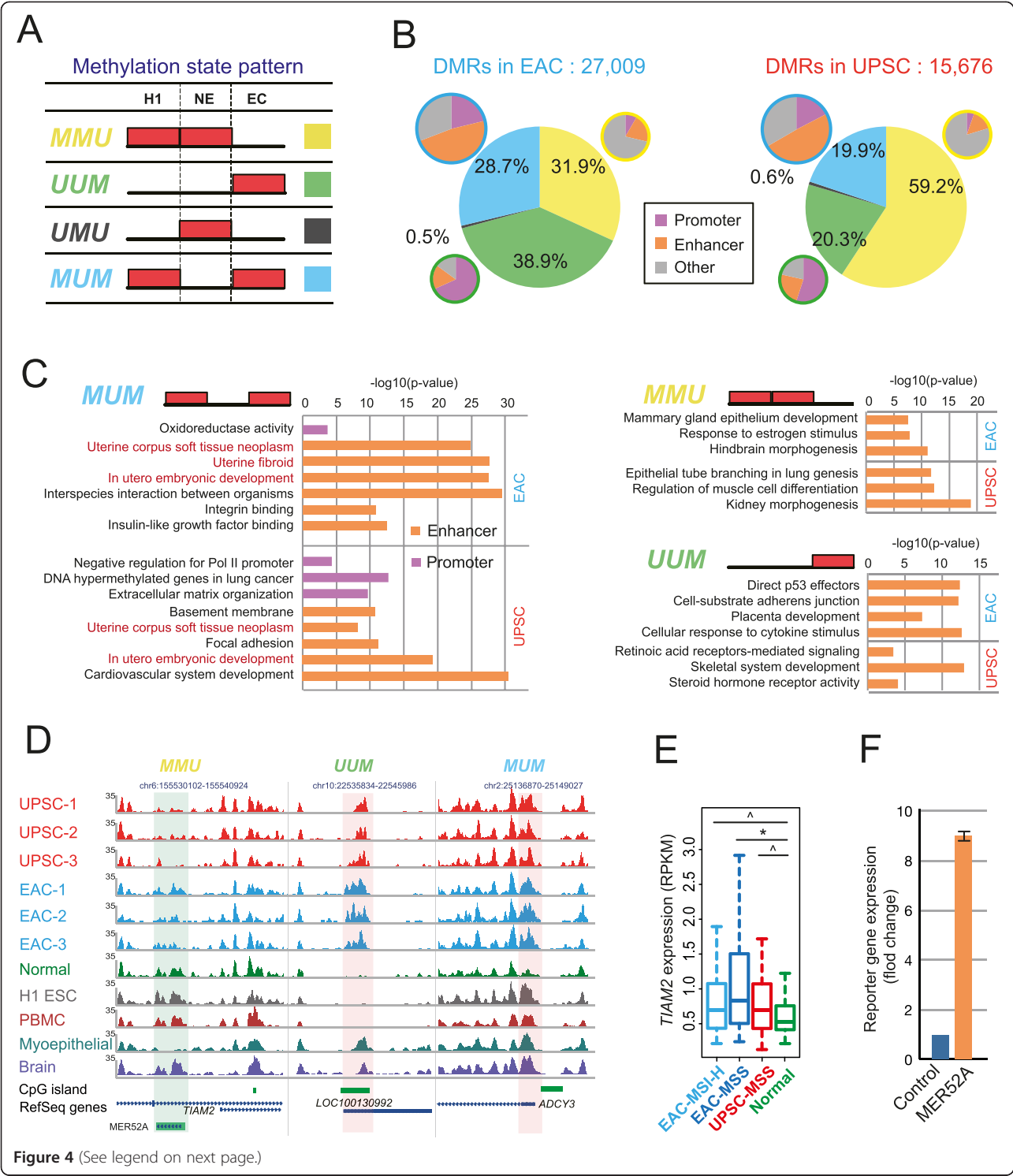
Figure 4 (See legend on next page.)

(See figure on previous page.)

Figure 4 DNA methylation alterations are enriched in functional promoters and enhancers. (A) Four DNA methylation state patterns across embryonic stem cells (H1), normal endometrium (NE), and endometrial cancer (EC). Red block: methylated state; empty: unmethylated state. **(B)** Fractions of DMRs with different DNA methylation state patterns in EAC (left) and UPSC (right). DMRs with patterns ***MMU*** (yellow), ***UUM*** (green), and ***MUM*** (blue) were annotated by chromHMM for their potential regulatory functions (purple: promoter; orange: enhancer state; gray: other states). **(C)** Functional enrichment of DMRs in pattern ***MMU***, ***UUM***, and ***MUM*** by GREAT analysis (purple: DMRs annotated as promoters by chromHMM; orange: DMRs annotated as enhancers by chromHMM). X-axis denotes negative log10-transformed p-value. **(D)** Epigenome Browser views of representative DMRs with different methylation state patterns. MeDIP-seq signal tracks were displayed. Increased and decreased methylation levels were highlighted by pink and green shading, respectively. **Left** (pattern ***MMU***): a DMR overlapped with transposable element *MER52A* was methylated in normal endometrium, H1 ESC, PBMC, and fetal brain, but unmethylated in breast myoepithelial cells. In EAC and UPSC, the element was hypomethylated, indicated by decreased MeDIP-seq signal. **Middle** (pattern ***UUM***): a DMR overlapped with CGI of lncRNA *LOC100130992* was unmethylated in normal endometrium, H1 ESC, and PBMC, but methylated in breast myoepithelial and fetal brain. The element was hypermethylated in EAC and UPSC. **Right** (pattern ***MUM***): a DMR overlapped with CGI-shore of *ADCY3* was methylated in all other tissues except in normal endometrium. In EAC and UPSC, this element was hypermethylated. **(E)** *TIAM2* expression in normal controls and grade 3 pre-classified (microsatellite state, and subtype) endometrial cancer patients samples. Y-axis: calculated RPKM value based on mRNA-seq from TCGA. Asterisk indicates *P* <0.01, caret indicates *P* <0.05, Student's *t*-Test. **(F)** The *MER52A* element enhanced luciferase reporter gene expression in HEK-293T cells. Blue: mini promoter; orange: *MER52A* + mini promoter.

with the strong connection between EAC and estrogen [67,68]. This data suggested a very different but complementary mechanism for cancer gene deregulation: these ***MMU*** DMRs might represent a broad array of tissue- or cell type-specific regulatory sites important for development and differentiation. Their demethylation in endometrial cancer might simultaneously deregulate tissue-specific genes important for other tissues, thereby contributing to the loss of endometrium specificity. The ***UUM*** pattern, on the other hand, was associated with functions related to carcinogenesis in general, including RNA polymerase II, nucleic acid-binding transcription factors, tumor suppressor genes, direct effectors of *TP53*, cell-substrate adherent factors, and agents of cellular response to cytokine stimuli (Figure 4C and Additional file 2: Figure S10A). Finally, very few DMRs exhibited the ***UMU*** pattern, which was predicted to be associated with embryonic development (Additional file 2: Figure S10A).

We highlighted a few examples with different methylation patterns (Figure 4D). A DMR with pattern ***MMU*** overlapped a copy of transposable element *MER52A*. This DMR was predicted to be an enhancer by chromHMM [66] and was located in the intron of *TIAM2* and 2.7 kb upstream of a gene-body alternative promoter. The DMR was highly methylated in ESC and normal endometrium, but was hypomethylated in endometrial cancer, particularly in UPSC. This element was also hypomethylated in breast myoepithelial cells but methylated in blood cells and brain tissues, consistent with the hypothesis that ***MMU*** pattern was associated with tissue-specific enhancers (Figure 4D). *TIAM2* regulates focal adhesion, and knockdown of this gene leads to a reduced rate of cell migration [69]. By examining TCGA mRNA-seq data, we found that the expression level of *TIAM2* was upregulated in both EAC and UPSC, compared to normal controls (Figure 4E). Importantly, we did not find DNA methylation changes at the promoter region of *TIAM2*. Based on a luciferase reporter assay, this *MER52A* element

increased reporter gene expression about 9-fold in HEK-293T cells (Figure 4F). These results suggested that demethylation of this transposable element could activate its enhancer function and increase *TIAM2* expression, potentially contributing to cancer cell migration and endometrial carcinogenesis. In another example, a pattern ***UUM*** DMR was identified in the promoter region of lncRNA *LOC100130992*. The DNA methylation status of this promoter was tissue-specific: methylated in breast myoepithelial cells and brain tissue, but unmethylated in H1 ESC, normal endometrium cells, and peripheral blood monocytes (Figure 4D). Numerous promoter DMRs with pattern ***MUM*** were predicted to play roles in normal endometrial functions. For example, the *ADCY3* promoter was specifically hypomethylated in normal endometrium, but its associated CpG island shore was highly methylated in EAC and UPSC. *ADCY3* encodes adenylyl cyclase 3, which catalyzes the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). cAMP stimulates vascular endothelial growth factor (*VEGF*) secretion by human endometrial stromal cells [70] and mediates progesterone-dependent decidualization of the human endometrium [71]. *ADCY3* expression was indeed repressed in both EAC and UPSC (Additional file 2: Figure S10B).

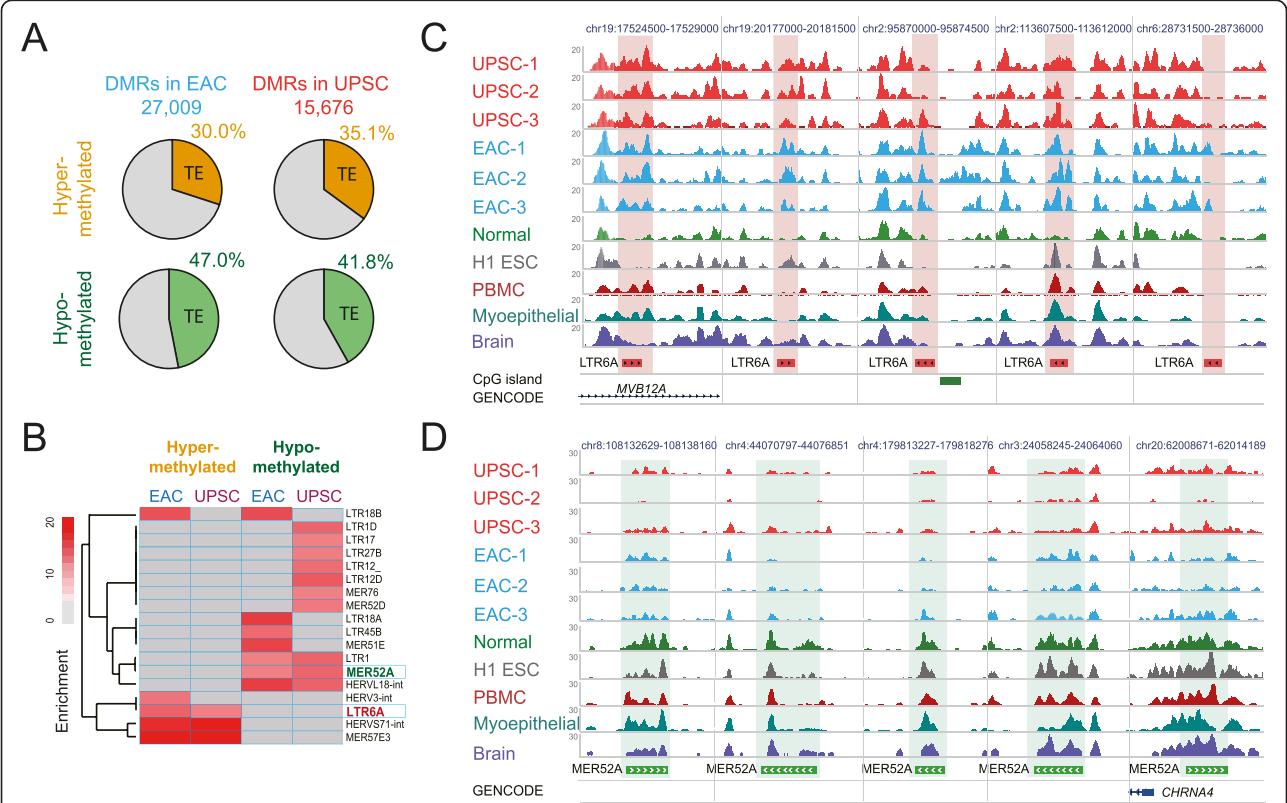## Bidirectional DNA methylation alteration on TE-derived enhancers

We noted that the *TIAM2* intragenic enhancer was within a transposable element. TEs make up nearly half of the human genome, but they have long been considered "junk DNA" that does not systematically contribute to normal cellular function [25,72]. Deregulation of TEs has, however, been proposed to be an important contributor to carcinogenesis. TEs, especially LINE-1 retrotransposons, are hypomethylated in many types of cancers [73-78]. We systematically evaluated the DNA methylation state of all TEs in the human genome that could be uniquely mapped

by our sequencing technology. Surprisingly, more than 30% of hypermethylated DMRs and more than 40% of hypomethylated DMRs in both types of endometrial cancer were within TEs (Figure 5A). These changes in methylation status were enriched for specific TE families and subfamilies. For example, the LINE-1 family was hypomethylated in both EAC and UPSC (Additional file 2: Figure S11), consistent with observations made in other cancer types [73-75]. We calculated enrichment of TE subfamilies in endometrial cancer DMRs. Four subfamilies were significantly more likely to be hypermethylated in EAC or UPSC (enrichment > 5-fold), whereas 13 subfamilies were significantly more likely to be hypomethylated in EAC or UPSC (enrichment > 5-fold) (Figure 5B).

We further investigated individual TE copies that exhibited a change in DNA methylation. For the purpose of illustration, we used *LTR6A* and *MER52A*. We computed methylation levels of individual copies of *LTR6A* and *MER52A* in EAC and UPSC (RPKM values from MeDIP-seq and MRE-seq data normalized to the values

in normal endometrium). About 70% of *LTR6A* elements exhibited increased MeDIP-seq and decreased MRE-seq signal, reflecting hypermethylation (Additional file 2: Figure S12A). In contrast, around 70% of *MER52A* elements had decreased MeDIP-seq and increased MRE-seq (Additional file 2: Figure S12B). The most highly hypermethylated *LTR6A* copies were found to be differentially methylated across different tissues, and were annotated as enhancers by chromHMM (Figure 5C). In contrast, the most highly hypomethylated *MER52A* copies exhibited strong methylation in all five normal tissues (including H1 ESC, breast, blood, brain, and endometrium) (Figure 5D).

Our results indicate that TEs may be important regulatory elements, particularly enhancers, in carcinogenesis. Transposable elements have been shown to wire gene regulatory networks during evolution [25,27-30], and TE subfamilies with tissue-specific hypomethylation also function as enhancers, which may associate with tissue identity [23]. How these transposable elements contribute

Figure 5 DNA methylation alterations of transposable elements. (A) Percentage of DMRs overlapping transposable elements in EAC and UPSC. **(B)** DMR enrichment within transposable element subfamilies. Enrichments > 5-fold were shaded in red. **(C)** Epigenome browser view of five genomic copies of *LTR6A* and their methylation levels across multiple samples. Only MeDIP-seq signal tracks were included. Increased methylation levels were highlighted by pink shading. These elements were unmethylated in normal endometrium, differentially methylated across H1 ESC, PBMC, breast myoepithelial cells, and fetal brain, and hypermethylated in EAC and UPSC. **(D)** Epigenome browser view of five genomic copies of *MER52A* and their methylation levels across multiple samples. MeDIP-seq signal tracks were displayed. Decreased methylation levels were highlighted by green shading. These elements were methylated (normal endometrium, H1 ESC, PBMC, and fetal brain) or partially methylated (breast myoepithelial cells) in normal cells, but were hypomethylated in EAC and UPSC.

to endometrial carcinogenesis remains to be elucidated. Moreover, in contrast to the common belief that most TEs are methylated and repressed in somatic cells [75,79,80], the discovery that a significant fraction of TEs were unmethylated in normal tissues but methylated in cancer suggests that some TEs play important roles in maintaining normal cell states and sustaining regular cellular physiological processes. These TE-mediated processes were likely disrupted in cancer cells. The widespread abnormal DNA methylation of transposable elements in cancer could reflect rewiring of gene regulatory networks during carcinogenesis. In normal tissues, some TEs were unmethylated and acted as tissue-specific enhancers. When these TEs were hypermethylated in cancer cells, they were silenced and this could have contributed to the down-regulation or silencing of target genes. On the other hand, hypomethylation of normally methylated TEs could function as cryptic enhancers or promoters [20], contributing to the upregulation of target oncogenes.

Discussion

Changes in DNA methylation have been shown to play roles in carcinogenesis and cancer progression in many malignancies [4,8,9] including endometrial cancer [11,12]. Previous studies have identified many genes exhibiting abnormal DNA methylation changes in endometrial cancer [11-14,16,81]; however, these studies, including the most recent and comprehensive mapping of DNA methylation in endometrial cancers by TCGA [18], were all focused on selected genomic regions (e.g., promoters and CpG islands) and CpG sites. Our study systematically investigated the complete DNA methylome in two clinically distinct types of endometrial cancer, EAC and UPSC, as well as normal endometrium, providing the first whole-genome DNA methylation map for this important disease. We observed significant DNA methylation changes in both types of endometrial cancer, which displayed both shared and cancer subtype-preferred DNA methylation signatures. Endometrial cancer-shared DMRs defined a specific signature of endometrial cancer that is confirmed by TCGA endometrial cancer methylation data (Figure 1D and 1E). In UPSC, which previously was thought to have infrequent DNA methylation changes and low *DNMT* expression, we identified numerous novel DNA methylation alterations (95% of 15,657 UPSC DMRs) in non-promoter regions (intergenic and intragenic). These results demonstrate the importance of applying sequencing-based, whole-genome approaches (i.e., MeDIP-seq and MRE-seq) to comprehensively map DNA methylation changes in cancers. In contrast, the Infinium HumanMethylation450 BeadChip, a popular genome-wide platform used by TCGA and many others, only interrogates about 1.7% (485,000) of the 28 million CpGs in the human genome. In our study, more than 40% of the hypermethylated

DMRs and 75% of the hypomethylated DMRs we identified were not covered by any Infinium probes; additionally only about 25% of the hypermethylated DMRs and 9% of the hypomethylated DMRs had more than two probes (Figure 6).

Despite being much more comprehensive, sequencing-based DNA methylome technologies are still cost inhibitory for directly profiling the large cohort of specimens that is often required to obtain power to detect most epigenetic events significantly associated with cancer. Thus, we advocate a strategy that combines deep profiling of a small number of samples (in this study, three samples per cancer type) followed by validation in a large cohort (in this study, we used publically available TCGA data). The strong validation results (Figure 1D-1F) suggest that our strategy is able to capture recurrent abnormal DNA methylation and provide a much more complete picture of the epigenetic landscape in endometrial cancer.

CpG islands and gene promoters are classic targets of DNA methylation changes in carcinogenesis. In endometrial cancer, we confirmed that hypermethylation was strongly enriched in transcription-related regions, including CGIs and promoters. Over 65% of CGIs with altered methylation were located in intergenic and intragenic regions, suggesting that alternative promoters may be frequent targets of DNA methylation change in cancers [20,22]. In addition, CGI shores were frequently affected by DNA methylation alterations (Figure 2A and Additional file 2: Figure S6A) [8,34]. Similar to CGIs, epigenetic changes at these potential regulatory regions may also regulate expression of nearby genes (Figure 2D, Additional file 2: Figure S6B). Epigenetic silencing of tumor suppressor genes (TSGs) is an important mechanism of carcinogenesis. In our study, many TSGs had hypermethylation in their core promoter regions, but interestingly, two tumor suppressor genes (*CDH1* and *SNF*) were found to have hypomethylated promoters and increased mRNA expression. Hypomethylation of the *S*NF promoter was also reported in lung and prostate cancer [42,43]. However, the methylation pattern of *CDH1* seems to be much more complex in endometrial cancer. Hypermethylated *CDH1* promoter and repressed *CDH1* expression were reported in some endometrial cancer cases [82,83]. Our approach revealed a hypomethylated DMR located in the 5' shore region of the *CDH1* promoter CpG island. This discovery was confirmed by TCGA Infinium 450K data: significant demethylation was detected by 2 CpG probes located in the DMR, but 8 nearby CpG probes within the CpG island did not detect any DNA methylation change (Additional file 2: Figure S6C). These results suggest that a demethylated promoter and increased expression of *CDH1* might be a common feature of endometrioid adenocarcinomas. The role of *CHD1* in endometrial cancer remains to be determined.
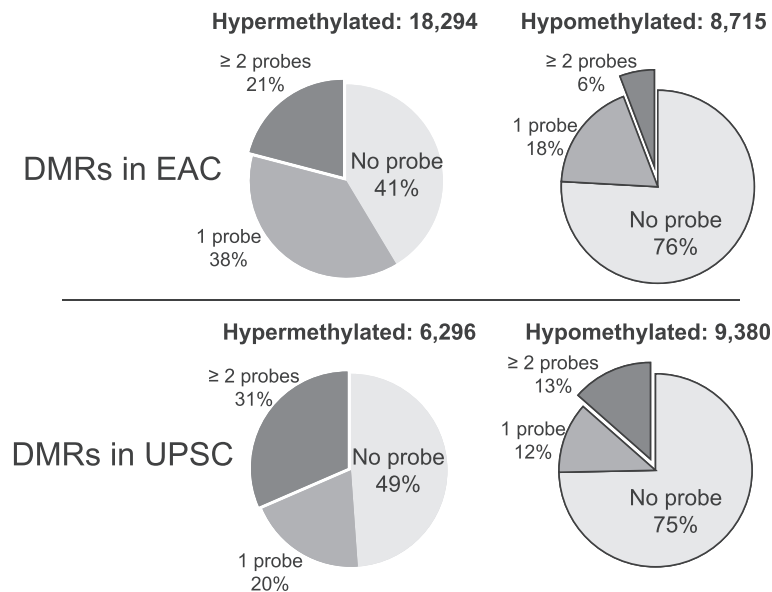
Figure 6 Infinium 450K array probe representation for DMRs. For EAC DMRs, 41% of hypermethylated and 76% of hypomethylated regions were not represented by Infinium 450K probes (light gray). For the UPSC DMRs, 49% of hypermethylated and 75% of hypomethylated regions did not contain Infinium 450K probes (light gray).

During human embryonic development, one copy of the X chromosome is randomly inactivated in female cells. The X-inactivation process is initiated by the expression of *XIST*, a mono-allelically expressed long noncoding RNA. High allelic expression of *XIST* on the inactive X establishes the inactivation state [84-88]. DNA methylation and repressive histone modifications, such as H3K27me3 and H3K9me3, play essential roles in maintaining the inactivated state [85,89-93]. The link between X chromosome epigenetic changes and female reproductive cancers was evident in a subset of breast and ovarian cancers [94]. In UPSC, we observed significant demethylation across the entire X chromosome, accompanied by down-regulated *XIST* expression. Interestingly, the expression level of *XIST* was also decreased in EAC and correlated with tumor grade (Figure 3C). In contrast to UPSC, we did not observe a significant global hypomethylation pattern in our three EAC samples. TCGA data revealed no global changes in the MSI-H type EAC, but a small but significant global hypomethylaton in the MSS type of EAC and in UPSC (Figure 3D). The Infinium 450K array only interrogates ~0.8% CpG sites on the X chromosome and is biased towards CpG sites in promoters and CpG islands. Thus, further validation is necessary to confirm the status of X chromosome DNA methylation. However, our results suggest that this significant sex chromosome demethylation could be a distinguishing signature between EAC and UPSC and might be associated with microsatellite stability. In ovarian cancer, a strong association was found between down regulated *XIST* expression and decreased time to recurrence in patients treated with Taxol

[95]. Although the mechanisms underlying *XIST* down-regulation and the biological significance of DNA methylation loss remain to be determined, we hypothesize that repressed *XIST* expression and DNA demethylation might contribute to the aggressive behavior in endometrial cancer, especially in UPSC.

Up to 80% of the human genome is transcribed but only ~2% encodes proteins [96-98]. In the past decade, numerous small and long non-coding RNAs (lncRNA) have been shown to be important regulators in both normal development and disease [59,99]. We observed DNA methylation alterations at many miRNA promoters. The expression of some miRNAs was correlated with DNA methylation states. For example, the *MIR200A-MIR200B-MIR429* cluster, which are tumor suppressor genes in many cancers, showed strong promoter demethylation in EAC and UPSC, with significantly upregulated expression (Additional file 2: Figure S8C and S8D). We also observed DNA methylation alterations at promoters of hundreds of lncRNA genes. The precise biological function of these miRNAs and lncRNAs and the significance of their DNA methylation changes in endometrial cancers remain to be determined. However, their identification underscores the importance of having a comprehensive catalog of cancer-associated DNA methylation changes across the whole genome.

DNA methylation exhibits strong tissue specificity, which is tightly programmed during embryonic development [100-104]. Alteration of DNA methylation in cancer radically changes the normal epigenetic landscape and results in loss of tissue specificity or cell identity [40]. For example, colorectal cancer DMRs had a striking overlap

with tissue-specific DMRs located in CGI shore regions, possibly disrupting tissue-specific regulatory programs [8]. By including DNA methylome data from embryonic stem cells, we identified a subset of endometrial cancer DMRs showing specific demethylation in normal endometrium, but high methylation in both H1 ESC and endometrial cancer (DNA methylation pattern *MUM*). These DMRs, which accounted for 30% of EAC DMRs and 20% of UPSC DMRs, showed a strong functional enrichment for genes related to uterine embryonic development and disease. Moreover, more than half of these *MUM* DMRs and ~25% of *MMU* and *UUM* DMRs were predicted to be distal enhancers (Figure 4B) that were deregulated in endometrial cancers. These results highlight that in addition to gene promoters, enhancers are a major target of DNA methylation changes in endometrial cancer.

Interestingly, a significant fraction of deregulated enhancers were derived from transposable elements (TEs): more than 30% of hypermethylated and 40% of hypomethylated DMRs harbored TEs. TEs have classically been understood to have high methylation in normal somatic cells and undergo global demethylation in cancers [10,73-75]. However, recent data have demonstrated that TEs play many important roles in normal biology, especially in wiring gene regulatory networks during evolution [25,27-30], and contributing to the establishment of tissue-specific gene regulation in development [23]. A breast-specific hypomethylated LTR element, *MER52A*, was found in the intron of the oncogenic *TIAM2* gene. This element was methylated in normal endometrium but demethylated in endometrial cancer (Figure 4D), and its enhancer function was validated using a reporter gene assay (Figure 4F). Our study suggested that DNA methylation in TEs might have a much more profound impact in cancer than previously believed. Consistent with our recent discovery [23], we identified unmethylated TEs in normal endometrium that may serve as tissue-specific enhancers. Some of these TE-derived, tissue-specific enhancers were hypermethylated and presumably silenced in endometrial cancer samples. In contrast, some normally methylated TEs (and TE-derived enhancers, possibly specific for other tissues) became hypomethylated in endometrial cancer. Disruption of normal DNA methylation of TE-derived and tissue-specific enhancers might be a novel mechanism to facilitate loss of tissue identity, acquisition of new cell type phenotypes, and to contribute to tumor development and progression (Figure 7). Complex rewiring of gene networks by epigenetic alterations might play a critical role in endometrial carcinogenesis. However, additional data, including a comprehensive annotation of the enhancer landscape in normal endometrial cells and in endometrial cancers, are needed to further elucidate this important mechanism.

Conclusions

We systematically investigated the complete DNA methylome in two clinically distinct types of endometrial cancer, EAC and UPSC, as well as normal endometrium, providing the first whole-genome DNA methylation map for this common and deadly disease. From these datasets, we identified tens of thousands of DMRs specific to the two cancer subtypes (endometrioid adenocarcinoma, or EAC, and uterine papillary serous carcinoma, or UPSC), and common to both cancer subtypes. We estimate that more than 2/3 of the DMRs we identified could not be discovered by using current array-based platforms. Many methylation changes were found in CpG island shores and were associated with expression changes of nearby genes. We observed large-scale DNA demethylation of chromosome X in UPSC accompanied by decreased *XIST* expression. Most significantly, we discovered that the majority of DMRs harbor regulatory functions including promoters and enhancers that are important to developmental and pathological changes of the uterus. Among these, remethylation of transposable elements in cancers might provide a novel mechanism to deregulate normal endometrium-specific enhancers derived from specific transposable elements. Our results demonstrate that DNA methylation changes are an important signature of endometrial cancer and regulate gene expression by affecting not only proximal promoters, but also distal enhancers, including those derived from transposable elements.
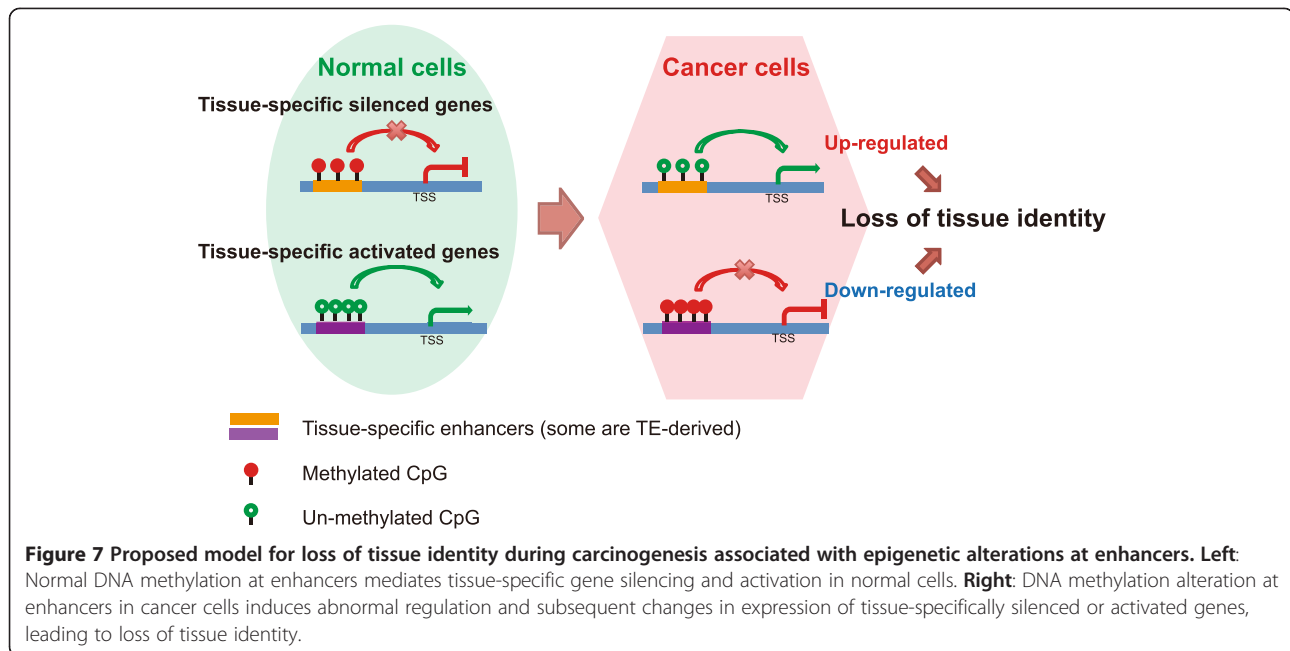
Methods
Website
A complete list of endometrial cancer associated DMRs and links to the WashU Epigenome Browser [105,106] are provided at the following website:

http://epigenome.wustl.edu/Cancer_Epigenome/.

Sample collection
All primary endometrial tumors and normal endometrium specimens analyzed were collected as part of IRB-approved studies (Washington University Medical Center Human Research Protection Office protocols HRPO-91-507, –93-0828, –92-242, and –10-1245), with participants' written informed consent. Histologic grading and typing were performed by gynecologic pathologists. Staging was determined using 1988 criteria from the International Federation of Gynecology and Obstetrics. Details of sample information were described in Additional file 1: Table S6. Tissue specimens and blood were obtained at the time of surgery and stored at –70°C until nucleic acids were extracted. All primary tumors evaluated had ≥ 70% neoplastic cellularity. Normal endometrium cells were collected from 10 healthy

Figure 7 Proposed model for loss of tissue identity during carcinogenesis associated with epigenetic alterations at enhancers. Left:
Normal DNA methylation at enhancers mediates tissue-specific gene silencing and activation in normal cells. **Right**: DNA methylation alteration at
enhancers in cancer cells induces abnormal regulation and subsequent changes in expression of tissue-specifically silenced or activated genes,
leading to loss of tissue identity.

donors in routine gynecological examinations and were
pooled together.

Library construction, sequencing and mapping

Genomic DNA from tumor tissues and normal endomet-
rium was extracted using the DNeasy Tissue kit (Qiagen,
Valencia, CA). MeDIP and MRE sequencing libraries were
constructed as previously described [22]. Sequencing
reads were aligned to hg19 with Bowtie [107]. MRE reads
were normalized to account for differences in enzyme effi-
ciency and scoring consisted of tabulating reads with
CpGs at each fragment end [22].

MeDIP-seq data and MRE-seq data of normal tissues,
including H1 ESC, PBMC, breast myoepithelial, and fetal
brain, were obtained from previous work [19]. This in-
formation was listed in Additional file 1: Table S7.

Affymetrix SNP array 6.0 processing

Affymetrix Genome-Wide SNP Array 6.0 was used for de-
tection of copy number changes in this study. Genomic
DNA was extracted from endometrial cancer samples and
matched peripheral blood mononuclear cells (PBMC)
using the DNeasy Tissue kit (Qiagen, Valencia, CA). The
Affymetrix Genome-Wide SNP Array 6.0 standard
protocol was performed as recommended by the manu-
facturer. 500 ng of total genomic DNA were analyzed
in the Genome Technology Access Center at Washing-
ton University in St. Louis. Data were analyzed by
Genotyping Console™ (GTC) Software 4.1.1 following
the manufacturer's instructions.

RNA isolation

RNA was extracted from frozen endometrial cancer tissues
and pooled normal endometrium using Trizol reagent (Life
Technologies) following the manufacturer's instructions
with an additional phenol/chloroform extraction to exclude
protein contamination. Trace DNA was removed by treat-
ment with TURBO DNA-free Kit (Life Technologies).
Quantity and quality of isolated RNA was measured and
evaluated by UV spectrophotometer and gel electro
phoresis.

Quantitative real time PCR (qRT-PCR) analysis

The qRT-PCR analyses were performed using the Super-
Script VILO cDNA Synthesis Kit (Life Technologies) with
iTaq Universal SYBR Green Supermix (Bio-Rad). 500 ng
total RNA was used in a 20 ul reverse transcription reac-
tion. The cDNA obtained was diluted to a total volume of
100 ul and stored at −20℃. The primers for human
DNMT1, DNMT3A, DNMT3B [108], *DNMT2* [109], *XIST*
[110] and candidate human housekeeping gene 18S rRNA
[111] were used for amplification of the target genes in
normal endometrium and endometrial cancer tissues. All
primers were synthesized by Integrated DNA Technolo-
gies. The qRT-PCR was performed in a 20 ul reaction con-
sisting of 2 ul diluted cDNA, 0.2 uM of each primer and
10 ul iTaq Universal SYBR Green Supermix. All amplifica-
tions were carried out in a Bio-Rad CFX96 Real-Time
PCR Detection (Bio-Rad) with denaturation at 95℃ for
30 s, followed by 40 cycles at 95℃ for 5 s and 60℃ for
30 s. A melting curve analysis was performed for each run
to confirm the specificity of amplification and lack of pri-
mer dimers. The qRT-PCR experiments were always run

in triplicate. The relative mRNA expression levels of target genes were quantified using the 2-ΔΔCT equation for endometrial normal and cancer tissues. Mean CT of normal endometrium was used as the calibrator sample.

DMR identification

The methylMnM package (http://epigenome.wustl.edu/MnM/) was utilized to identify differentially methylated regions (DMRs) in the R 2.15 environment between the DNA methylome of normal endometrium and the DNA methylome of each cancer sample. Default parameters were used, and a statistical cutoff of q-value < 1e-5 was applied to select DMRs from each pair-wise comparison at a resolution of 500 bp. EAC or UPSC DMRs were defined such that the same genomic region must have been called a DMR and have the same direction of DNA methylation change in at least two out of the three cancer vs. normal pairwise comparisons. Endometrial cancer shared DMRs (EC-shared DMRs) were defined as DMRs that were both EAC DMRs and UPSC DMRs and had same direction of DNA methylation change in both types of cancers compared to the pooled normal sample. EAC type-preferred DMRs (tpDMRs) were defined as DMRs in all three EAC samples but not in any UPSC sample; UPSC tpDMRs were defined as DMRs in all three UPSC samples but not in any EAC sample. The same procedure and statistical cutoff were used to define DMRs between H1 ESC and normal endometrium.

Genomic features

RepeatMasker annotations, CpG islands, and RefSeq Gene coding loci features were all downloaded from the UCSC Genome Browser [112,113]. 1 KB core promoters were defined as 1 kb around the most 5' transcription start site (500 bp upstream and 500 bp downstream of TSS) of any RefSeq gene record. microRNA loci were downloaded from mirBASE. The microRNA gene cluster TSSs were download from mirStart (http://mirstart.mbc.nctu.edu.tw/). lincRNA loci were download from the Human lincRNA Catalog (http://www.broadinstitute.org/genome_bio/human_lincrnas/).

TCGA DNA methylation data

Processed DNA methylation data of uterine corpus endometrial carcinomas (Infinium HumanMethylation450 BeadChip platform) was downloaded from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/). Quantile normalization was performed across all samples. The beta-value of each probe within DMRs was isolated for further analysis in the R 2.15 environment. Sample histology information was obtained from the supplementary materials of Kandoth et al. [18]. Sample details were described in Additional file 1: Table S8.

TCGA RNA expression data

Processed mRNA-seq and miRNA-seq data (normalized read counts for each gene, analysis version 2) of uterine corpus endometrial carcinomas (Illumina GA and Illumina HiSeq platform) were downloaded from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/). RPKM (Reads per kilobase per million) values were computed for each gene using TCGA's mRNA sequencing data (analysis version 2). The lengths of transcripts were obtained from the NCBI gene bank. Quantile normalization was performed across all samples. RPM (Reads per million) values for each microRNA gene were computed using TCGA's microRNA sequencing data. All analyses were performed in the R environment (Version 2.15). Sample details were provided in Additional file 1: Table S8.

Validation of DMRs using TCGA Infinium 450K data

Based on TCGA Infinium 450K data, for any DMR identified in our study that overlaped at least one Infinium probe, we computed the average methylation level (aML) in each cancer type group (described in Additional file 1: Table S8) using the beta value of overlapping Infinium probes. Specifically:

$$aML = \sum_i^n \left(\sum_j^w BVij \Big/ w \right) \Big/ n$$

where n is the number of samples in a specific cancer type group, w is the number of available Infinium CpG probes within the corresponding DMR, and $BVij$ is beta-value of the j th Infinium CpG probe within that DMR in the i th sample.

DNA methylation change (DMC) were calculated as:

$$DMC = aML_c - aML_n$$

where aML_c is the averaged methylation level of a DMR in a specific cancer type group; aML_n is the averaged methylation level of a DMR in the normal control group.

Based on TCGA Infinium 450K data, all DMRs belonging to EC-shared DMRs,

EAC tpDMRs, and UPSC tpDMRs, were validated by performing the Man-Whitney U test to examine the significance of DNA methylation difference between cancer group and normal controls. P-values were corrected for multiple testing by the *Benjamini-Hochberg* method. Validated DMRs were defined by the criteria: 1) the DMR has an adjusted p-value of < 0.05 between the cancer group and normal controls; 2) the DMR has an averaged DNA methylation change of over 0.05 in the cancer group compared to normal controls; 3) the direction of DNA methylation change between cancer and control samples in TCGA Infinium 450K data must fit our predication.

ENCODE chromHMM chromatin state annotation

ChromHMM annotation of nine ENCODE cell lines [66] were obtained from the UCSC Genome Browser [114]. The nine cell lines are: H1 ESC, GM12878, K562, HepG2, HUVEC, HMEC, HSMM, NHEK, and NHLF. The genome was split to 500 bp windows. For each 500 bp window, we assigned overlapping annotations of "promoter", "enhancer", and "insulator" states from these chromHMM maps. In each cell line, the state of each 500 bp window was assigned to the dominant chromHMM state, defined as the state that occupied over 50% of a given window. After merging the data from 9 cell lines, "promoter" regions were defined as the windows have a dominant "promoter" chromHMM state in at least one cell line. "Enhancer" regions were defined as windows have a dominant "enhancer" chromHMM state in at least one cell line but no "promoter" chromHMM state in any cell line.

ENCODE TFBS and DHS data

ENCODE TFBS data and Dnase I Hypersensitive sites (DHS) data were downloaded from the UCSC Genome Browser ENCODE data portal [114]. TFBS data contained 4,380,444 regions; DHS data contained 1,281,988 regions.

Enrichment calculation

The binding site enrichment score (*ES*) for each genomic feature, DHS, and transcription factor with respect to DMRs was calculated as:

$$ES = \frac{n\_hit/n\_DMR}{N\_hit/N\_all}$$

Where n_hit is the number of DMRs that contain specific a genomic feature, experimentally annotated DHS, or TFBS; n_DMR is the total number of DMRs; N_hit is the number of genomic windows with a specific genomic feature, annotated DHS, or TFBS; N_all is the number of 500 bp windows in the human genome (hg19).

Gene ontology enrichment analysis

Gene Ontology (GO) analyses for biological processes were performed using the GREAT package [115]. Gene regulatory domains were defined by default as regions spanning 5 kb upstream to 1 kb downstream of the TSS (regardless of other nearby genes). Gene regulatory domains were extended in both directions to the nearest gene's basal domain but no more than a maximum extension in one direction. Only categories that were below a false discovery rate of 0.05 were reported.

Reporter gene assay

Candidate transposable element sequences were amplified from genomic DNA using Pfu-polymerase (Agilent) and primers containing KpnI- or BglII- restriction sites. PCR products were gel-purified using the Qiagen Gel purification kit and then digested by the corresponding restriction enzymes (NEB). The digested PCR products were cloned into the pGL4.23[luc2/minP]-vector (Promega, E8411) using T4-ligase (NEB) and transformed into chemically competent DH5α-cells. The positive clones were verified by enzyme digestion and sequencing. 800 ng of reporter plasmid (or empty pGL4.23[luc2/minP]-vector control) was cotransfected into HEK-293T cell line with the CMV promoter driven *Renilla* luciferase plasmid (200 ng). The luciferase activity was measured 48 hours post transfection and normalized by the relative *Renilla* control. Genomic locations of candidates and primer information were summarized in Additional file 1: Table S9.

Availability of supporting data

The data set supporting the results of this article are available at the accompanying website (http://epigenome.wustl.edu/Cancer_Epigenome/). Sequencing data has been deposited in the NCBI's Gene Expression Omnibus repository under the accession number GSE51565.

Additional files

Additional file 1: Supplementary tables and figures legend.

Additional file 2: Supplementary figures.

Author details
[1]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA.
[2]Shanghai International Joint Cancer Institute, The Second Military Medical

University, Shanghai 200433, P. R. China. [3]Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin Province 130024, P. R. China. [4]Department of Mathematics and Division of Biostatistics, Washington University in Saint Louis, Saint Louis, MO 63130, USA. [5]Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins University, Baltimore, MD 21231, USA. [6]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. [7]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Washington University School of Medicine, St Louis, MO 63124, USA. [8]The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA.

References

1. Siegel R, Naishadham D, Jemal A: **Cancer statistics, 2012.** *CA Cancer J Clin* 2012, **62**(1):10–29.
2. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I: **Endometrial cancer.** *Lancet* 2005, **366**(9484):491–505.
3. del Carmen MG, Birrer M, Schorge JO: **Uterine papillary serous cancer: a review of the literature.** *Gynecol Oncol* 2012, **127**(3):651–661.
4. Lapeyre JN, Walker MS, Becker FF: **DNA methylation and methylase levels in normal and malignant mouse hepatic tissues.** *Carcinogenesis* 1981, **2**(9):873–878.
5. Laird PW, Jaenisch R: **DNA methylation and cancer.** *Hum Mol Genet* 1994, **3 Spec No:**1487–1495.
6. Cheung HH, Lee TL, Rennert OM, Chan WY: **DNA methylation of cancer genome.** *Birth Defects Res C Embryo Today* 2009, **87**(4):335–350.
7. Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G: **DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer.** *Genome Res* 2011, **21**(7):1017–1027.
8. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, Feinberg AP: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41**(2):178–186.
9. McLendon RFA, Bigner D, Van Meir EG, Brat DJ: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**(7216):1061–1068.
10. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, Camargo AA, Stevenson BJ, Ecker JR, Bafna V, Strausberg RL, Simpson AJ, Ren B: **Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer.** *Genome Res* 2012, **22**(2):246–258.
11. Esteller M, Catasus L, Matias-Guiu X, Mutter GL, Prat J, Baylin SB, Herman JG: **hMLH1 promoter hypermethylation is an early event in human endometrial tumorigenesis.** *Am J Pathol* 1999, **155**(5):1767–1772.
12. Salvesen HB, MacDonald N, Ryan A, Jacobs IJ, Lynch ED, Akslen LA, Das S: **PTEN methylation is associated with advanced stage and microsatellite instability in endometrial carcinoma.** *Int J Cancer* 2001, **91**(1):22–26.
13. Sasaki M, Kotcherguina L, Dharia A, Fujimoto S, Dahiya R: **Cytosine-phosphoguanine methylation of estrogen receptors in endometrial cancer.** *Cancer Res* 2001, **61**(8):3262–3266.
14. Liao X, Siu MK, Chan KY, Wong ES, Ngan HY, Chan QK, Li AS, Khoo US, Cheung AN: **Hypermethylation of RAS effector related genes and DNA methyltransferase 1 expression in endometrial carcinogenesis.** *Int J Cancer* 2008, **123**(2):296–302.
15. Moreno-Bueno G, Hardisson D, Sanchez C, Sarrio D, Cassia R, Garcia-Rostan G, Prat J, Guo M, Herman JG, Matias-Guiu X, Esteller M, Palacios J: **Abnormalities of the APC/beta-catenin pathway in endometrial cancer.** *Oncogene* 2002, **21**(52):7981–7990.
16. Risinger JI, Maxwell GL, Berchuck A, Barrett JC: **Promoter hypermethylation as an epigenetic component in Type I and Type II endometrial cancers.** *Ann N Y Acad Sci* 2003, **983**:208–212.
17. Xiong Y, Dowdy SC, Xue A, Shujuan J, Eberhardt NL, Podratz KC, Jiang SW: **Opposite alterations of DNA methyltransferase gene expression in endometrioid and serous endometrial cancers.** *Gynecol Oncol* 2005, **96**(3):601–609.
18. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA: **Integrated genomic characterization of endometrial carcinoma.** *Nature* 2013, **497**(7447):67–73.
19. Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, Cheng JB, Li D, Stevens M, Lee HJ, Xing X, Zhou J, Sundaram V, Elliott G, Gu J, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecek T, Weiss A, O'Geen H, Farnham PJ, Maire CL, Ligon KL, Madden PA, Tam A, Moore R, Hirst M, Marra MA, Zhang B, Costello JF, Wang T: **Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm.** *Genome Res* 2013, **23**(9):1522–1540.
20. Nagarajan RP, Zhang B, Bell RJ, Johnson BE, Olshen AB, Sundaram V, Li D, Graham AE, Diaz A, Fouse SD, Smirnov I, Song J, Paris PL, Wang T, Costello JF: **Recurrent epimutations activate gene body promoters in primary glioblastoma.** *Genome Res* 2014, **24**(5):761–774.
21. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, et al: **Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications.** *Nat Biotechnol* 2010, **28**(10):1097–1105.
22. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF: **Conserved role of intragenic DNA methylation in regulating alternative promoters.** *Nature* 2010, **466**(7303):253–257.
23. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecek T, Weiss A, O'Geen H, Farnham PJ, Madden PA, Mungall AJ, Tam A, Kamoh B, Cho S, Moore R, Hirst M, Marra MA, Costello JF, Wang T: **DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape.** *Nat Genet* 2013, **45**(7):836–841.
24. Stevens M, Cheng JB, Li D, Xie M, Hong C, Maire CL, Ligon KL, Hirst M, Marra MA, Costello JF, Wang T: **Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods.** *Genome Res* 2013, **23**(9):1541–1553.
25. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D: **Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53.** *Proc Natl Acad Sci U S A* 2007, **104**(47):18613–18618.
26. Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H: **Transposase-derived transcription factors regulate light signaling in Arabidopsis.** *Science* 2007, **318**(5854):1302–1305.
27. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET: **Evolution of the mammalian transcription factor binding repertoire via transposable elements.** *Genome Res* 2008, **18**(11):1752–1762.
28. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G: **Transposable elements have rewired the core regulatory network of human embryonic stem cells.** *Nat Genet* 2010, **42**(7):631–634.
29. Lynch VJ, Leclerc RD, May G, Wagner GP: **Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals.** *Nat Genet* 2011, **43**(11):1154–1159.
30. Chuong EB, Rumi MA, Soares MJ, Baker JC: **Endogenous retroviruses function as species-specific enhancer elements in the placenta.** *Nat Genet* 2013, **45**(3):325–329.
31. Lowe CB, Bejerano G, Haussler D: **Thousands of human mobile element fragments undergo strong purifying selection near developmental genes.** *Proc Natl Acad Sci U S A* 2007, **104**(19):8005–8010.
32. Vasavi Sundaram YC, Zhihai M, Daofeng L, Xiaoyun X, Peter E, Snyder MP, Ting W: **Widespread contribution of transposable elements to the innovation of gene regulatory networks.** *Genome Res*, in press.
33. Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR: **Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns.** *Genome Biol* 2012, **13**(10):R84.
34. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: **Differential methylation of tissue- and cancer-specific CpG island shores**

distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 2009, **41**(12):1350–1353.

35. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, *et al*: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**(7414):57–74.
36. Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A: **Cell type-specific DNA methylation at intragenic CpG islands in the immune system.** *Genome Res* 2011, **21**(7):1074–1086.
37. Deaton AM, Bird A: **CpG islands and the regulation of transcription.** *Genes Dev* 2011, **25**(10):1010–1022.
38. Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD: **DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas.** *Cancer Res* 1986, **46**(6):2917–2922.
39. Esteller M: **Epigenetic gene silencing in cancer: the DNA hypermethylome.** *Hum Mol Genet* 2007, **16**(Spec No 1):R50–R59.
40. Kulis M, Esteller M: **DNA methylation and cancer.** *Adv Genet* 2010, **70**:27–56.
41. Zhao M, Sun J, Zhao Z: **TSGene: a web resource for tumor suppressor genes.** *Nucleic Acids Res* 2013, **41**(Database issue):D970–D976.
42. Hsu A, Wong CP, Yu Z, Williams DE, Dashwood RH, Ho E: **Promoter de-methylation of cyclin D2 by sulforaphane in prostate cancer cells.** *Clin Epigenetics* 2011, **3**:3.
43. Shiba-Ishii A, Noguchi M: **Aberrant stratifin overexpression is regulated by tumor-associated CpG demethylation in lung adenocarcinoma.** *Am J Pathol* 2012, **180**(4):1653–1662.
44. Grady WM, Willis J, Guilford PJ, Dunbier AK, Toro TT, Lynch H, Wiesner G, Ferguson K, Eng C, Park JG, Kim SJ, Markowitz S: **Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer.** *Nat Genet* 2000, **26**(1):16–17.
45. Negraes PD, Favaro FP, Camargo JL, Oliveira ML, Goldberg J, Rainho CA, Salvadori DM: **DNA methylation patterns in bladder cancer and washing cell sediments: a perspective for tumor recurrence detection.** *BMC Cancer* 2008, **8**:238.
46. Huang YW, Liu JC, Deatherage DE, Luo J, Mutch DG, Goodfellow PJ, Miller DS, Huang TH: **Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer.** *Cancer Res* 2009, **69**(23):9038–9046.
47. Tan Y, Zhang B, Wu T, Skogerbo G, Zhu X, Guo X, He S, Chen R: **Transcriptional inhibiton of Hoxd4 expression by miRNA-10a in human breast cancer cells.** *BMC Mol Biol* 2009, **10**:12.
48. Chien CH, Sun YM, Chang WC, Chiang-Hsieh PY, Lee TY, Tsai WC, Horng JT, Tsou AP, Huang HD: **Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data.** *Nucleic Acids Res* 2011, **39**(21):9345–9356.
49. Tsuruta T, Kozaki K, Uesugi A, Furuta M, Hirasawa A, Imoto I, Susumu N, Aoki D, Inazawa J: **miR-152 is a tumor suppressor microRNA that is silenced by DNA hypermethylation in endometrial cancer.** *Cancer Res* 2011, **71**(20):6450–6462.
50. Ratner ES, Tuck D, Richter C, Nallur S, Patel RM, Schultz V, Hui P, Schwartz PE, Rutherford TJ, Weidhaas JB: **MicroRNA signatures differentiate uterine cancer tumor subtypes.** *Gynecol Oncol* 2010, **118**(3):251–257.
51. Torres A, Torres K, Pesci A, Ceccaroni M, Paszkowski T, Cassandrini P, Zamboni G, Maciejewski R: **Diagnostic and prognostic significance of miRNA signatures in tissues and plasma of endometrioid endometrial carcinoma patients.** *Int J Cancer* 2013, **132**(7):1633–1645.
52. Davalos V, Moutinho C, Villanueva A, Boque R, Silva P, Carneiro F, Esteller M: **Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis.** *Oncogene* 2012, **31**(16):2062–2074.
53. Wiklund ED, Bramsen JB, Hulf T, Dyrskjot L, Ramanathan R, Hansen TB, Villadsen SB, Gao S, Ostenfeld MS, Borre M, Peter ME, Orntoft TF, Kjems J, Clark SJ: **Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer.** *Int J Cancer* 2011, **128**(6):1327–1334.
54. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ: **The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.** *Nat Cell Biol* 2008, **10**(5):593–601.
55. Hu X, Macdonald DM, Huettner PC, Feng Z, El Naqa IM, Schwarz JK, Mutch DG, Grigsby PW, Powell SN, Wang X: **A miR-200 microRNA cluster as**

prognostic marker in advanced ovarian cancer. *Gynecol Oncol* 2009, **114**(3):457–464.
56. Snowdon J, Zhang X, Childs T, Tron VA, Feilotter H: **The microRNA-200 family is upregulated in endometrial carcinoma.** *PLoS One* 2011, **6**(8):e22828.
57. Dai Y, Xia W, Song T, Su X, Li J, Li S, Chen Y, Wang W, Ding H, Liu X, Li H, Zhao Q, Shao N: **MicroRNA-200b is overexpressed in endometrial adenocarcinomas and enhances MMP2 activity by downregulating TIMP2 in human endometrial cancer cell line HEC-1A cells.** *Nucleic Acid Ther* 2013, **23**(1):29–34.
58. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL: **Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.** *Genes Dev* 2011, **25**(18):1915–1927.
59. Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs: insights into functions.** *Nat Rev Genet* 2009, **10**(3):155–159.
60. Zhang X, Zhou Y, Mehta KR, Danila DC, Scolavino S, Johnson SR, Klibanski A: **A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells.** *J Clin Endocrinol Metab* 2003, **88**(11):5119–5126.
61. Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y, Zhou Y, Klibanski A: **Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions.** *Endocrinology* 2010, **151**(3):939–947.
62. Miyoshi N, Wagatsuma H, Wakana S, Shiroishi T, Nomura M, Aisaka K, Kohda T, Surani MA, Kaneko-Ishino T, Ishino F: **Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q.** *Genes Cells* 2000, **5**(3):211–220.
63. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C, Klibanski A: **Activation of p53 by MEG3 non-coding RNA.** *J Biol Chem* 2007, **282**(34):24731–24742.
64. Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce CM, Patel T: **microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer.** *Oncogene* 2011, **30**(47):4750–4756.
65. Zhang B, Chen B, Wu T, Xuan Z, Zhu X, Chen R: **Estimating developmental states of tumors and normal tissues using a linear time-ordered model.** *BMC Bioinformatics* 2011, **12**:53.
66. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**(7345):43–49.
67. Smith DC, Prentice R, Thompson DJ, Herrmann WL: **Association of exogenous estrogen and endometrial carcinoma.** *N Engl J Med* 1975, **293**(23):1164–1167.
68. Persson I, Weiderpass E, Bergkvist L, Bergstrom R, Schairer C: **Risks of breast and endometrial cancer after estrogen and estrogen-progestin replacement.** *Cancer Causes Control* 1999, **10**(4):253–260.
69. Rooney C, White G, Nazgiewicz A, Woodcock SA, Anderson KI, Ballestrem C, Malliri A: **The Rac activator STEF (Tiam2) regulates cell migration by microtubule-mediated focal adhesion disassembly.** *EMBO Rep* 2010, **11**(4):292–298.
70. Popovici RM, Irwin JC, Giaccia AJ, Giudice LC: **Hypoxia and cAMP stimulate vascular endothelial growth factor (VEGF) in human endometrial stromal cells: potential relevance to menstruation and endometrial regeneration.** *J Clin Endocrinol Metab* 1999, **84**(6):2245–2248.
71. Brar AK, Frank GR, Kessler CA, Cedars MI, Handwerger S: **Progesterone-dependent decidualization of the human endometrium is mediated by cAMP.** *Endocrine* 1997, **6**(3):301–307.
72. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**(6998):471–476.
73. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T: **Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer.** *Int J Cancer* 2009, **124**(1):81–87.
74. Ogino S, Nosho K, Kirkner GJ, Kawasaki T, Chan AT, Schernhammer ES, Giovannucci EL, Fuchs CS: **A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer.** *J Natl Cancer Inst* 2008, **100**(23):1734–1738.

75. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, Voravud N, Sriuranpong V, Mutirangura A: Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* 2004, **23**(54):8841–8846.

76. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ: Landscape of somatic retrotransposition in human cancers. *Science* 2012, **337**(6097):967–971.

77. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* 1992, **52**(3):643–645.

78. Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM: Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 1988, **333**(6168):87–90.

79. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP: A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004, **32**(3):e38.

80. Florl AR, Schulz WA: Peculiar structure and location of 9p21 homozygous deletion breakpoints in human cancer cells. *Gene Chromosome Canc* 2003, **37**(2):141–148.

81. Zhuang J, Jones A, Lee SH, Ng E, Fiegl H, Zikan M, Cibula D, Sargent A, Salvesen HB, Jacobs IJ, Kitchener HC, Teschendorff AE, Widschwendter M: The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. *PLoS Genet* 2012, **8**(2):e1002517.

82. Fiolka R, Zubor P, Janusicova V, Visnovsky J, Mendelova A, Kajo K, Lasabova Z, Plank L, Danko J: Promoter hypermethylation of the tumor-suppressor genes RASSF1A, GSTP1 and CDH1 in endometrial cancer. *Oncol Rep* 2013, **30**(6):2878–2886.

83. Yi TZ, Guo J, Zhou L, Chen X, Mi RR, Qu QX, Zheng JH, Zhai L: Prognostic value of E-cadherin expression and CDH1 promoter methylation in patients with endometrial carcinoma. *Cancer Invest* 2011, **29**(1):86–92.

84. Puck JM, Willard HF: X inactivation in females with X-linked disease. *N Engl J Med* 1998, **338**(5):325–328.

85. Mohandas T, Sparkes RS, Shapiro LJ: Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 1981, **211**(4480):393–396.

86. Norris DP, Patel D, Kay GF, Penny GD, Brockdorff N, Sheardown SA, Rastan S: Evidence that random and imprinted Xist expression is controlled by preemptive methylation. *Cell* 1994, **77**(1):41–51.

87. Wang X, Miller DC, Clark AG, Antczak DF: Random X inactivation in the mule and horse placenta. *Genome Res* 2012, **22**(10):1855–1863.

88. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF: A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 1991, **349**(6304):38–44.

89. Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y: Role of histone H3 lysine 27 methylation in X inactivation. *Science* 2003, **300**(5616):131–135.

90. Kaslow DC, Migeon BR: DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci U S A* 1987, **84**(17):6210–6214.

91. Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E: X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* 2000, **225**(2):294–303.

92. Hellman A, Chess A: Gene body-specific methylation on the active X chromosome. *Science* 2007, **315**(5815):1141–1143.

93. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL: Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* 2001, **107**(6):727–738.

94. Pageau GJ, Hall LL, Ganesan S, Livingston DM, Lawrence JB: The disappearing Barr body in breast and ovarian cancers. *Nat Rev Cancer* 2007, **7**(8):628–633.

95. Huang KC, Rao PH, Lau CC, Heard E, Ng SK, Brown C, Mok SC, Berkowitz RS, Ng SW: Relationship of XIST expression and responses of ovarian cancer to chemotherapy. *Mol Cancer Ther* 2002, **1**(10):769–776.

96. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M: Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004, **306**(5705):2242–2246.

97. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais

E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR: RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007, **316**(5830):1484–1488.

98. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, *et al*: Landscape of transcription in human cells. *Nature* 2012, **489**(7414):101–108.

99. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 2009, **11**(3):228–234.

100. Bird A: DNA methylation patterns and epigenetic memory. *Genes Dev* 2002, **16**(1):6–21.

101. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E, Oyolu CB, Schroth GP, Absher DM, Baker JC, Myers RM: Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* 2009, **19**(6):1044–1056.

102. Monk M, Boubelik M, Lehnert S: Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 1987, **99**(3):371–382.

103. Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M: Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 2010, **42**(12):1093–1100.

104. Yagi S, Hirabayashi K, Sato S, Li W, Takahashi Y, Hirakawa T, Wu G, Hattori N, Hattori N, Ohgane J, Tanaka S, Liu XS, Shiota K: DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res* 2008, **18**(12):1969–1978.

105. Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, Koebbe BC, Nielsen C, Hirst M, Farnham P, Kuhn RM, Zhu J, Smirnov I, Kent WJ, Haussler D, Madden PA, Costello JF, Wang T: The Human Epigenome Browser at Washington University. *Nat Methods* 2011, **8**(12):989–990.

106. Zhou X, Lowdon RF, Li D, Lawson HA, Madden PA, Costello JF, Wang T: Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat Methods* 2013, **10**(5):375–376.

107. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, **10**(3):R25.

108. Xue ZT, Sjogren HO, Salford LG, Widegren B: An epigenetic mechanism for high, synergistic expression of indoleamine 2,3-dioxygenase 1 (IDO1) by combined treatment with zebularine and IFN-gamma: potential therapeutic use in autoimmune diseases. *Mol Immunol* 2012, **51**(2):101–111.

109. Ehrlich M, Woods CB, Yu MC, Dubeau L, Yang F, Campan M, Weisenberger DJ, Long T, Youn B, Fiala ES, Laird PW: Quantitative analysis of associations between DNA hypermethylation, hypomethylation, and DNMT RNA levels in ovarian tumors. *Oncogene* 2006, **25**(18):2636–2645.

110. Ogawa Y, Sun BK, Lee JT: Intersection of the RNA interference and X-inactivation pathways. *Science* 2008, **320**(5881):1336–1341.

111. Valente V, Teixeira SA, Neder L, Okamoto OK, Oba-Shinjo SM, Marie SK, Scrideli CA, Paco-Larson ML, Carlotti CG Jr: Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. *BMC Mol Biol* 2009, **10**:17.

112. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002, **12**(6):996–1006.

113. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Dreszer TR, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, Kent WJ: The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* 2013, **41**(Database issue):D64–D69.

114. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, Diekhans M, Fujita PA, Goldman M, Gravell RC, Harte RA, Hinrichs AS, Kirkup VM, Kuhn RM, Learned K, Maddren M, Meyer LR, Pohl A, Rhead B, Wong MC, Zweig

AS, Haussler D, Kent WJ: ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res* 2012, **40**(Database issue):D912–D917.

115. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010, **28**(5):495–501.