

Whole-exome sequencing of endometriosis identifies frequent alterations in genes involved in cell adhesion and chromatin-remodeling complexes

Xiaolei Li^{1,†}, Yan Zhang^{2,†}, Luyang Zhao^{3,†}, Lingxiong Wang⁴, Zhiqiang Wu¹, Qian Mei¹, Jing Nie¹, Xiang Li¹, Yali Li³, Xiaobing Fu^{1,2,5}, Xiaoning Wang¹, Yuanguang Meng^{3,*} and Weidong Han^{1,2,*}

¹Department of Molecular Biology, Institute of Basic Medicine, School of Life Sciences, Chinese PLA General Hospital, Beijing, China, ²Department of Bio-therapeutic, Chinese PLA General Hospital, Beijing, China, ³Department of Gynecologic Oncology, Chinese PLA General Hospital, Beijing, China, ⁴Department of Cancer Center, Chinese PLA General Hospital, Beijing, China and ⁵Key Laboratory of Wound Healing and Cell Biology, Institute of Burns, the First Affiliated Hospital to the Chinese PLA General Hospital, Trauma Center of Postgraduate Medical School, Beijing, China

Received March 29, 2014; Revised May 16, 2014; Accepted June 23, 2014

Endometriosis is a complex and enigmatic disease that arises from the interplay among multiple genetic and environmental factors. The defining feature of endometriosis is the deposition and growth of endometrial tissues at sites outside of the uterine cavity. Studies to date have established that endometriosis is heritable but have not addressed the causal genetic variants for this disease. Here, we conducted whole-exome sequencing to comprehensively search for somatic mutations in both eutopic and ectopic endometrium from 16 endometriosis patients and five normal control patients using laser capture microdissection. We compared the mutational landscape of ectopic endometrium with the corresponding eutopic sample from endometriosis patients compared with endometrium from normal women and identified previously unreported mutated genes and pathway alterations. Statistical analysis of exome data identified that most genes were specifically mutated in both eutopic and ectopic endometrium cells. In particular, genes that are involved in biological adhesion, cell–cell junctions, and chromatin-remodeling complex(es) were identified, which partially supports the retrograde menstruation theory that proposes that endometrial cells are refluxed through the fallopian tubes during menstruation and implanted onto the peritoneum or pelvic organs. Conspicuously, when we compared exomic mutation data for paired eutopic and ectopic endometrium, we identified a mutational signature in both endometrial types for which no overlap in somatic single nucleotide variants were observed. These mutations occurred in a mutually exclusive manner, likely because of the discrepancy in endometriosis pathology and physiology, as eutopic endometrium rapidly regrows, and ectopic endometrial growth is inert. Our findings provide, to our knowledge, an unbiased view of the landscape of genetic alterations in endometriosis and vital information for indicating that genetic alterations in cytoskeletal and chromatin-remodeling proteins could be involved in the pathogenesis of endometriosis, thus implicating a novel therapeutic possibility for endometriosis.

INTRODUCTION

Endometriosis (MIM131200) is considered as a common gynecologic disease that affects 5–10% of women of reproductive age; however, in women with pelvic pain, unexplained infertility

or both, the frequency is 35–50% (1). It is characterized by the presence of endometrium-like tissue in abnormal locations, mainly on the pelvic peritoneum but also on the ovaries, in the rectovaginal septum, and, more rarely, in the pericardium, pleura and even the brain (2,3). The concept that endometriosis

*To whom correspondence should be addressed at: Department of Molecular Biology, Institute of Basic Medicine, School of Life Sciences, Chinese PLA General Hospital, Beijing, China and Department of Bio-therapeutic, Chinese PLA General Hospital, Beijing, China. Tel: +86 1066937463; Fax: +86 1066937516; Email: hanwdrsw69@yahoo.com (W. Han); Department of Gynecologic Oncology, Chinese PLA General Hospital, Beijing, China. Tel: +86 1066937917; Fax: +86 1066937516; Email: meng6512@vip.sina.com (Y. Meng)

[†]X. Li, Y. Zhang and L. Zhao contributed equally to this work.

is a tumor-like disease because of its metastatic potential, local tissue invasion and increased growth and vascularization of ectopic endometrial tissue is widespread and generally accepted (3). Disease severity is classified using the revised American Fertility Society (rAFS) system (4), which assigns affected individuals on a scale of Stages I–IV (defined as minimal-to-severe disease), which quantifies disease according to the amount of ectopic endometrial tissue present, its location and the amount of pelvic scarring.

Although endometriosis predisposition is likely multifactorial, a genetic component is evident (5–7). The genetic variants potentially underlying the hereditary component of endometriosis have been widely investigated via hypothesis-driven candidate gene studies and/or hypothesis-free genome-wide approaches, such as genome-wide association studies (GWASs) and genome-wide linkage studies (GWLSs) that reported a significant association of endometriosis with *7p15.2*, *9p21* and *10q23-26* loci, which are highly likely to be involved in the pathogenesis of endometriosis in women of Japanese and European ancestry (8–12). Nevertheless, a comprehensive catalog of genetic alternations is far from completed, and the corresponding efforts to identify causal coding variants and the key drivers of endometriosis pathogenesis remain poorly understood. None of the current classification schemes are entirely accurate, which suggests that a more complete understanding of the genetic changes that are relevant to the pathogenesis of endometriosis will be required for better classification of risk and, ultimately, better approaches to therapy. Although the study of the genetics of endometriosis is starting to bear fruit, it is still faced with many challenges inherent to the complexities of the disease and the genetic loci identified to date have not pinpointed the actual genetic variants that are functional and/or causal to endometriosis.

In the present study, to explore the contribution of functional coding variants to endometriosis, we surveyed the spectrum of somatic alternations in women with or without endometriosis by sequencing the exomes of endometrial cells from 16 endometriosis patients and 5 healthy women. We analyzed non-synonymous single-nucleotide variants (SNVs) across the genome by exome sequencing in both eutopic and ectopic endometrium samples and normal endometrium samples, and we also compared the exome of the ectopic and eutopic endometrium of the same patients with respect to endometrium from unaffected samples. Frequently mutated genes in endometriosis involved in cell adhesion, chromatin modification, cell cycle, DNA repair and the regulation of apoptosis were the most enriched biological pathways, suggesting that alternation of genes in these pathways may be involved in the pathogenesis of endometriosis and may provide opportunities for novel diagnostics and therapies.

RESULTS

Landscape of somatic mutations in the exomes of ectopic versus eutopic endometrium

To define the mutational spectrum of endometriosis, we performed whole-exome sequencing of genomic DNA from paired eutopic and ectopic microdissected endometrium using laser capture microdissection (LCM) and matched peripheral

blood samples that were obtained concurrently from 17 fertile women patients with surgically confirmed ovarian endometriosis. Of these, one eutopic endometrium sample on which whole-exome sequencing was performed lacked matched whole-exome sequence data for the ectopic endometrium owing to sequencing failure for this sample; therefore, it was excluded from subsequent analyses (Supplementary Material, Table S1 and Fig. S1). We enriched exonic sequences using Agilent's SureSelect Human All Exon kit technology for targeted exon capture, which targeted 50 Mb of sequence from exons and their flanking regions in ~22 000 genes. Subsequently, we performed whole-exome sequencing on the Illumina Genome Analyzer II platform with 100 bp by paired-end reads (Illumina, Inc.) and aligned the reads to the reference human genome (hg19) using MAQ and BWA software (Online Methods). On average, we sequenced the 30 Mb targeted exome regions of each sample to a mean depth of 30 × or greater, and, on average, 82% of the targeted regions were covered sufficiently for confident variant calling (defined as ≥ 6 ×; Fig. 1A, B and Supplementary Material, Table S2). A Monte-Carlo simulation indicated that the study detected 76% of the existing somatic mutations (Online Methods and Supplementary Material, Fig. S2). To eliminate common germline mutations, we removed any potential somatic mutation that was observed in either dbSNP132 or data from the 1000 Genomes Project. To determine which of these alterations were somatic (that is, endometriosis specific), we also removed those variation data that were present in exomes from the matched blood samples. Using stringent filtering criteria to avoid false-positive calls, which included an empirically determined threshold that accounted for read quality and depth and provided an optimal balance between positive predictive value (82.3%) and sensitivity (96.4%; Online Methods). In total, across the coding regions of the 16 cases, 18 879 somatic alterations were identified in eutopic endometrium, with a range of 82 to 4864 per sample, and 16 891 somatic alterations were identified in ectopic endometrium, with a range of 265 to 3915 per sample (Fig. 2A). All of the somatic alterations were absent in DNA from peripheral blood that was collected from 16 fertile women with endometriosis. Of these, 6421 were synonymous (silent) mutations, and 12 458 were predicted to alter protein structure (range: 62–3265 per sample) in eutopic endometrium, and in ectopic endometrium, 5641 were synonymous (silent) mutations and 11 250 were non-synonymous mutations in ectopic endometrium. Among the mutations, the number of C>T mutations was significantly greater than the numbers of other nucleotide substitutions, which was consistent with spontaneous deamination (13), resulting in a high prevalence of C>T/G>A transitions followed by C>A/G>T and A>G/T>C, and was similar between the eutopic and ectopic endometrium (Fig. 2B and E and Supplementary Material, Table S3), resembling the mutational spectra in gastric (14) and colorectal cancers (13,15) as well as endometrial carcinomas (16). Somatic mutation rates varied considerably among these endometrium samples. Some had mutation rates of <2 per 10⁶ bases, whereas a few had mutation rates of >100 per 10⁶ bases. The overall non-synonymous median mutation frequency was 26.2 mutations per megabase (range of 1.76–107.9 mutations per megabase) in eutopic endometrium and was 20 mutations per megabase (range of 5.8–87.1 mutations per megabase) in

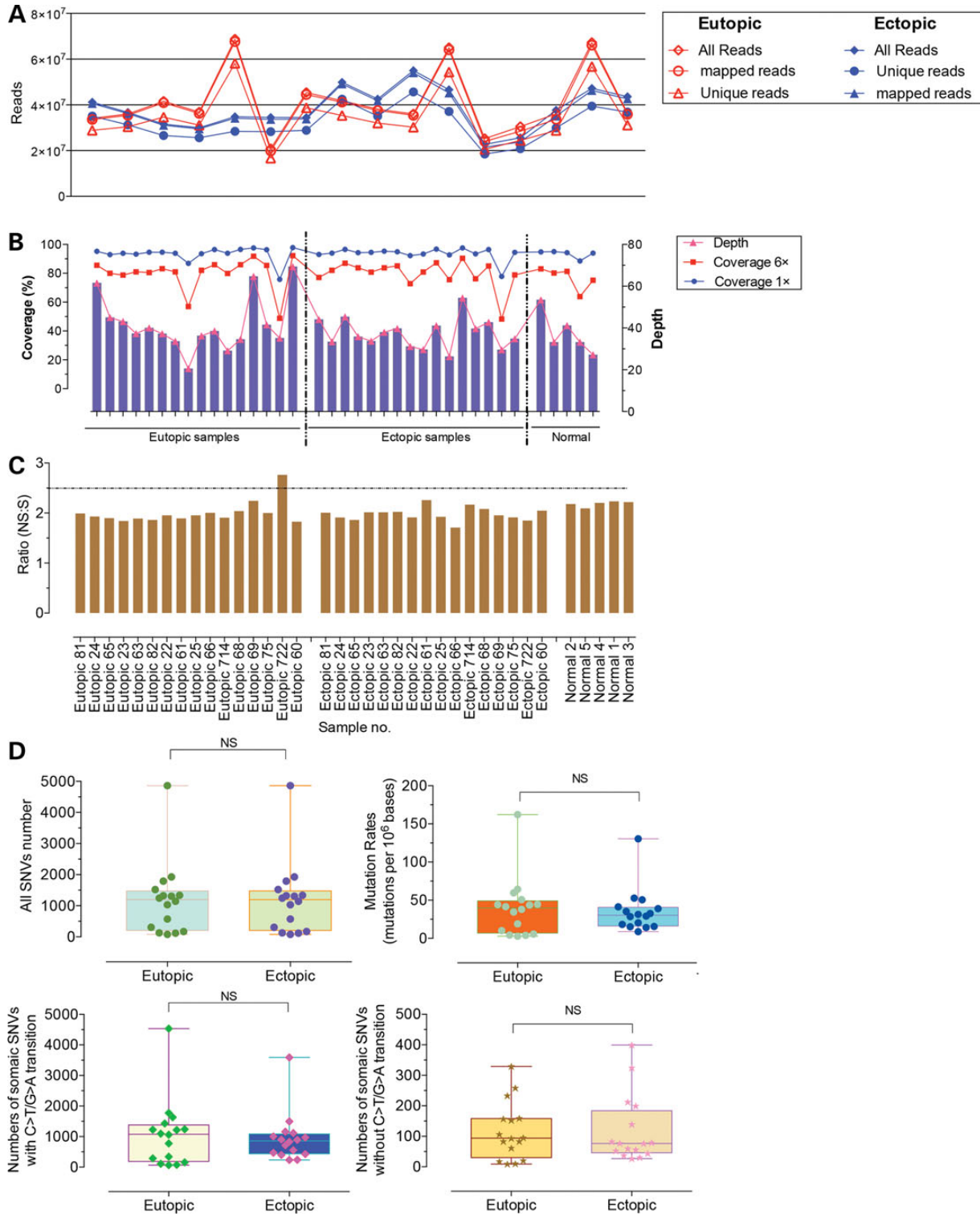


Figure 1. Sequencing characteristics and mutation number and type. **(A)** Exome-sequencing quality control. Total raw sequencing reads and enrichment for the targeted exome regions enrichments were similar across samples. **(B)** Coverage of the target coding regions analyzed at the indicated depth is plotted for each pair of eutopic and ectopic endometrium specimens, and normal endometrium. **(C)** Non-synonymous versus synonymous mutations in the endometrium samples (including eutopic endometrium and ectopic endometrium) were analyzed by whole-exome sequencing. **(D)** Plots demonstrating the number of somatic mutations (or the number of somatic mutations per megabase) in eutopic and ectopic endometrium samples from 16 individuals with endometriosis; plots also showing the fraction of somatic SNVs that were C>T/G>A transitions or other than C>T/G>A transitions in eutopic and ectopic endometrium from 16 endometriosis patients. The *P*-value indicates whether there was a significantly different mutation number in eutopic and ectopic endometrium and was calculated using the two-tailed Wilcoxon signed-rank test.

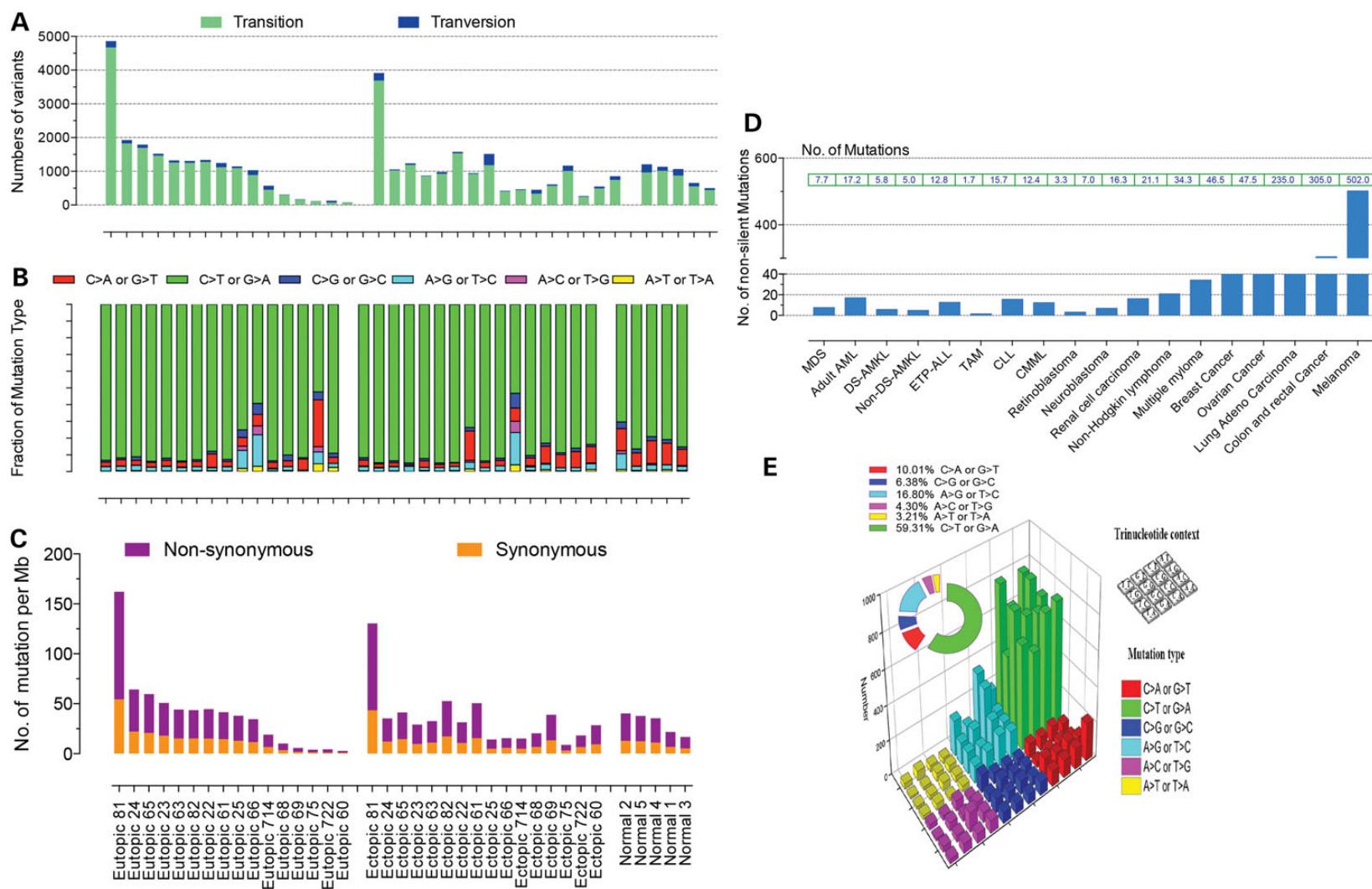


Figure 2. Mutation spectrum of whole-exome sequenced eutopic and ectopic endometrium from endometriosis. **(A)** For each endometrium sample, the total number of SNVs, transition variants and transversion mutations is shown. **(B and E)** The mutation spectra of single base pair substitutions in endometriosis as assessed by whole-exome sequencing. The number of each of the six base substitution classes resulting in non-synonymous changes in the whole-exome screen is shown in B. Lego plots of mutation frequencies across 16 endometriosis samples from whole-exome sequencing. Base substitutions were divided into six categories to represent the six possible base changes (each category represented by a different color). Substitutions were further divided by the 16 possible flanking nucleotides surrounding the mutated base as listed in the trinucleotide context legend. Inserted pie charts indicate the distribution of all mutations for a given middle mutated base across the evaluated (as shown in E). **(C)** Somatic mutation rate per megabase of the covered target sequence in the 16 cases with endometriosis and five normal control cases. **(D)** Number of non-silent mutations across other cancer types and number of non-silent mutations for the indicated cancer types. MDS, myelodysplastic syndromes; ETP-ALL, early T-cell precursor acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; CMML, chronic myelomonocytic leukemia.

ectopic endometrium (Fig. 2C). Overall, the rates of alternations that we observed in endometriosis, whether in the eutopic or ectopic endometrium samples, were significantly higher than had been previously determined rates for other tumors (13,17–20) (Fig. 2D) that have also been sequenced by whole-exome sequencing, but was similar to an ultramutated group of endometrial tumors with unusually high mutation rates (16). These findings may reflect differences in mitotic index, prior treatments or carcinogen exposure between the endometrium samples.

Identification of background mutations in normal endometrium cells

To determine whether normal endometrium cells could accumulate mutations over time, we also performed whole-exome sequencing of normal endometrium and their matched peripheral blood samples from five healthy women without endometriosis (Supplementary Material, Fig. S1 and Table S1). On average, over 77% of the targeted regions were covered sufficiently for confident variant calling (defined as $\geq 6\times$; Fig. 1B and Supplementary Material, Table S2). We determined somatic mutations using the same method that had been used in the 16 patients with endometriosis, and we identified 4562 somatic mutations, consisting of 1437 synonymous and 3125 non-synonymous, which were present in the DNA from normal endometrium cells from five healthy women but absent from their matched peripheral blood DNA. The observed mutation spectrum was very similar to that of samples from 16 fertile women patients with ovarian endometriosis. Both mostly had transitions, suggesting that they may represent mutations caused by methylcytosine residues deamination (Fig. 2A). The overall non-synonymous median mutation frequency was 24.7 mutations per megabase (range of 11.6–75.7 mutations per megabase) in these normal endometrium samples (Fig. 2A and C). Although the preexisting mutations would not be pathogenetically relevant, they would be present in every cell in the founding clone and appear as somatic mutations when the endometrium sample is sequenced. Although this hypothesis cannot be proven directly with current technologies, it strongly suggests that normal, self-renewing endometrium cells derived from healthy women may accumulate random, benign background mutations that would increase with the extension in endometrium update cycles and generally be irrelevant for endometriosis pathogenesis. We also investigated whether the high mutation frequency in endometriosis was associated with clinical variables, including age and disease stage. However, likely because of the limited sample size of our study, no significant associations were identified (Supplementary Material, Fig. S3A–D).

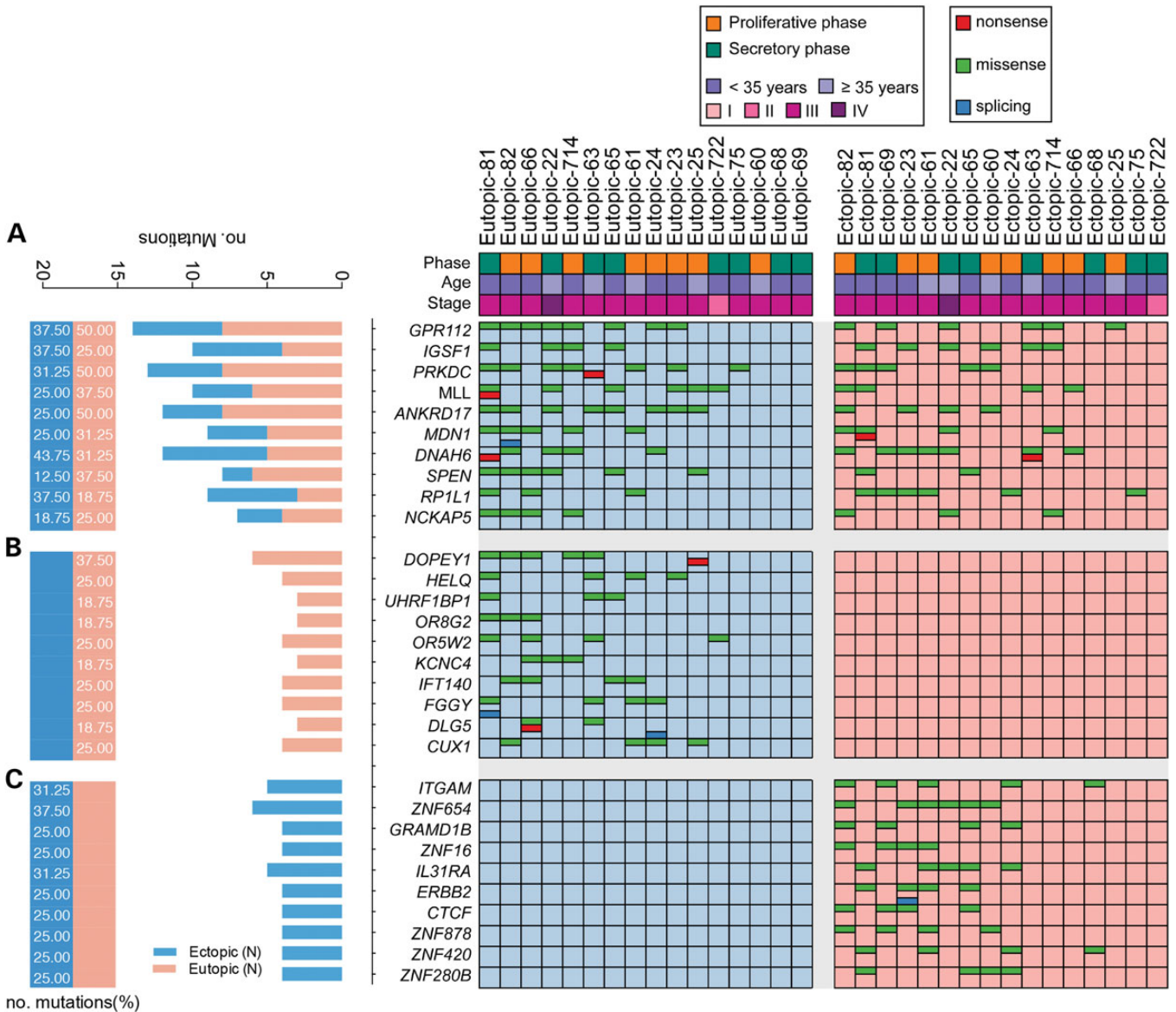
Variant allele frequencies within individual cases

By performing targeted sequencing with high read coverage to validate the variants detected by whole-exome sequencing, we were able to accurately estimate the variant allele frequencies for somatic mutations identified in each endometriosis case (including both eutopic and ectopic endometrium samples). The variant allele frequencies were calculated according to Kernel Density Analysis. Based on the variant allele frequency distribution, we estimated the clonal population number and size in each

endometrium sample. Recent studies have demonstrated the importance of clonal evolution in tumor progression and metastasis development (21,22). Clusters of mutations with similar variant allele frequencies within individual cases provide potential evidence of single founding clones in both eutopic and ectopic endometrium samples. According to Kernel Density Analysis, we identified one cluster of mutations (likely representing the founding clone in all cases without additional subclones derived from the founding clone) in each genome. Conspicuously, in endometriosis cases, the number of variants specific to each endometrium sample, whether eutopic or ectopic endometrium cells, was relatively small. We focused on mutations with variant allele frequencies of 20% in most endometrium samples, which strongly suggested that these somatic mutations were randomly distributed (Supplementary Material, Fig. S4A–C).

Analysis of data to identify genes affecting the risk of endometriosis and pathway analysis

Somatic mutations can either be ‘driver’ mutations that have a role in endometriosis or functionally inert ‘passenger’ changes. As described above, this screen yielded a ratio of 2.0:1 non-synonymous-to-synonymous changes (NS:S ratio), which is not higher than the 2.5:1 NS:S ratio that was predicted for non-selected, passenger mutations (23), suggesting that most of these alternations are likely to be ‘passenger’ mutations rather than ‘drivers’ in endometriosis (Fig. 1C). A previous study has developed a significantly mutated gene algorithm to detect biologically significant variants from cancer genome sequencing data in an unbiased manner. Mutations were evaluated across samples to distinguish somatic genetic variations, which appear to be targeted by driver rather than passenger mutations using the MutSig algorithm analysis (24–26). Therefore, to identify genes that are mutated at significant frequencies across our cohort, we used the MutSig, which takes into account gene size, trinucleotide context, gene structure, sample-specific mutation rate, non-silent to silent mutation ratios, clustering within genes, and base conversation across species. This tool compares mutation occurrence in each gene to what would be expected by chance given a background mutation frequency model that factors in the mutation spectra, silent mutations, mutation frequencies and regional mutation frequencies along the genome (27). To search for new recurrent mutations, we only considered genes that were mutated in at least two samples that had been subjected to whole-exome sequencing and were significantly more mutated than the local background mutation rate. However, in our cohort of 16 paired endometrium samples, we unexpectedly could not identify any significantly mutated genes in either eutopic or ectopic endometrium samples, which was likely because of the limited sample size of our study. This result does not in line with previous human cancer studies (13,15,16,27). Taken together, these results suggest that eutopic and ectopic endometrium cells have distinct mutational profiles, even within different endometrium samples from the same patient (Fig. 3A–C). One possible interpretation for this difference between endometriosis and other human cancer is because the endometrium has unique physiological and pathological characteristics, such as constant updating in the eutopic tissues and inert growth in ectopic tissues, which may delay



diagnosis because of symptom variability and confusion with other disorders (2,3), compared with other human cancers. Isolated or rare mutations that did not reach significance across the entire cohort were observed in many genes with known involvement in cell adhesion, such as laminin β -3 (*LAMB3*), protocadherin β -6 (*PCDHB6*) and laminin γ -3 (*LAMC1*), which participate in the pathogenesis of endometriosis (28,29). These data implied that all of the mutations were mutually exclusive, suggesting that the effect of a mutation in one gene can substitute for the other, and altering only one complex component may be sufficient to disrupt or alter the entire complex.

To search for potential recurrent mutated genes, we looked for genes with mutations in either the same or different mutated sites

in at least two endometrium samples. In total, we observed somatic mutations in 8135 genes in endometriosis, but no mutations in these genes appeared in our normal endometrium exome-seq reads, of which, 2090 were mutated in two or more eutopic endometrium samples and/or 1611 were mutated in two or more ectopic endometrium samples (Supplementary Material, Table S4). Of these genes, we also identified a large number of genes with novel mutations (636 genes), most of which occurred at high frequency found in both the eutopic and ectopic endometrium genomes (Supplementary Material, Table S5). Mutations in these genes may represent cooperating events that are capable of interacting with several different initiating mutations. We also observed ~724 genes that were

mutated in only two or more eutopic endometrium samples and 465 genes that were mutated in only two or more ectopic endometrium samples. We also determined that mutations in 222 genes, as has been previously described in a GWAS of endometriosis, also appeared in our whole-exome sequencing study, of which, 106 genes were mutated in two or more endometrium samples from endometriosis patients (Supplementary Material, Table S6). To further understand the impact of the mutations on gene function, we applied SIFT (30) and/or Polyphen 2 (31) to predict the probable functional impact of the validated non-synonymous mutations and determined that a large fraction of the alterations would affect protein function and were predicted to be deleterious (Supplementary Material, Table S7 and S8).

To shed light on the potential important underlying molecular mechanisms of endometriosis and explore the functional impact of the mutated genes, we performed a series of pathway analyses using genes that were recurrently mutated in both eutopic and ectopic endometrium samples from two or more individuals using ClueGO (32,33) and identified mechanisms known to be important in endometriosis such as cell adhesion ($P = 4.5 \times 10^{-5}$) and biological adhesion ($P = 4.6 \times 10^{-5}$), which ranked near the top of the list as enriched (Supplementary Material, Table S9, Fig. S5 and Fig. 4). These findings support the hypothesis that enhanced cellular motility and invasiveness has an important role in endometriosis etiology (34,35). Aside from these functional groupings, one of the enriched functional groupings was chromatin modification ($P = 3.8 \times 10^{-4}$), formed by *CHD6*, *MLL*, *MLL3* and *KDM5A*, which were significantly mutated in the cohort. We also examined genes and gene family mutation frequencies and observed frequent alternations in multiple endometrium samples in 18 other chromatin-remodeling genes (*ING5*, *RSF1*, *CREBBP*, *INO80*, *TTF1*, *CECR2*, *ARID1A*, *RPS6KA5*, *MLL5*, *BAZ1B*, *KDM2A*, *SMARCD3*, *KDM3A*, *KDM3B*, *SETD2*, *BAZZA*, *MLL4*, *BRD8*). Of these genes, four members of the MLL complex (*MLL*, *MLL3*, *MLL4*, *MLL5*) were identified, which encoded histone methyltransferase genes and are involved in histone H3 lysine 4 modification. We also detected alterations in other related genes in more than two endometriosis patients: histone demethylase genes *KDM5A* (*JARID1A*), *KDM2A* and *KDM3B*; two SWI/SNF-related chromatin-remodeling genes, *SMARCD3* and *ARID1A*; the histone acetyltransferase *CREBBP*; and other chromatin-remodeling genes, such as *CHD6* (encoding a component of the SNF2/RAD54 helicase family that remodels chromatin to allow cell-type-specific gene expression) and *SETD2* (encoding the only H3K36 trimethyltransferase in humans). In addition, alternations in chromatin remodeling and DNA replication genes have also been reported in various other tumors (36–39) [i.e., serous endometrial tumors (40,41)] in recent studies. These novel findings led us to speculate that altered epigenetic chromatin regulation and posttranslational modifications may play a pivotal role in the pathogenesis of endometriosis, further implicating frequent mutational disruption of these processes in the molecular pathogenesis of one of the deadliest forms of the disease progression of endometriosis.

Aside from those genes that were mutated in both eutopic and ectopic endometrium, we also discovered some genes that were only mutated in either eutopic or ectopic endometrium samples. To distinguish those genes that were specifically mutated in either the eutopic or the ectopic endometrium, we only

considered genes that were specifically mutated in at least two samples in only eutopic or ectopic endometrium samples but not mutated in normal endometrium. Using these criteria, we identified 724 genes and 465 genes were specifically mutated in eutopic and ectopic endometrium, respectively, but were not mutated in normal samples, indicating that gene mutation differences do exist between the eutopic and ectopic endometrium samples (Supplementary Material, Table S5). To decipher the potential functionally important underlying mechanisms in the eutopic and ectopic endometrium cells, we also performed a series of pathway analyses using genes that were specifically mutated in either eutopic or ectopic endometrium cells using ClueGO, DAVID (database for annotation, visualization and integrated discovery) Bioinformatics Resource 6.7 [(42) and (43)], and a DAVID functional annotation analysis identified that cellular membrane adhesion and cell–cell junction organization pathways were the most significantly enriched mutated genes (Supplementary Material, Table S10 and S11).

DISCUSSION

Endometriosis is defined as the presence of endometrial glands and stroma at extra-uterine sites. It is a multifactorial disorder responsible for infertility and pelvic pain that affects ~5–10% of women of reproductive age (44). Susceptibility to endometriosis is thought to depend on the complex interaction of genetic, immunologic, hormonal and environmental factors. Therefore, changes in normal cellular functions such as cellular proliferation, cell migration, invasiveness and cell survival could be involved in establishment of endometriosis. The multifactorial etiology allows taking into consideration different theories on its pathogenesis, which is currently still not completely understood.

In recent years, the development of non-invasive diagnostic tools based on microarrays could be of great benefit in the clinical management of endometriosis. To discover sufficiently sensitive and specific genetic biomarkers that may be suitable for non-invasive diagnosis of endometriosis and development of strategies for the prevention of deleterious sequelae and treatment, it is necessary to detect differences in genetic changes in endometrium that are consistently found across individual patients with endometriosis compared with women without endometriosis. An understanding of the genetic underpinning of endometriosis is prerequisite to the development of novel diagnostic and therapeutic approaches. Moreover, the genetic mutations occurring in sporadic endometriosis-associated ovarian carcinoma have been elucidated by studies both in human samples and in murine models of disease (45,46), highlighting the involvement of *K-Ras*, *PTEN*, *HNF-1* and *ARID1A*, among others. As with other complex diseases, genetic factors contribute substantially to endometriosis development. The ultimate purpose of identifying individual genetic signatures is to eventually develop targets for improved diagnostics. The recent development in genetic analyses has increased the likelihood of pinpointing generalizable, endometriosis-specific markers. Thus, the main objectives of our study were to survey the genetic alternations by exome sequencing in the ectopic endometrium compared with eutopic endometrium in women with endometriosis and to gain new insight into the underlying biology of this enigmatic disease. We found that

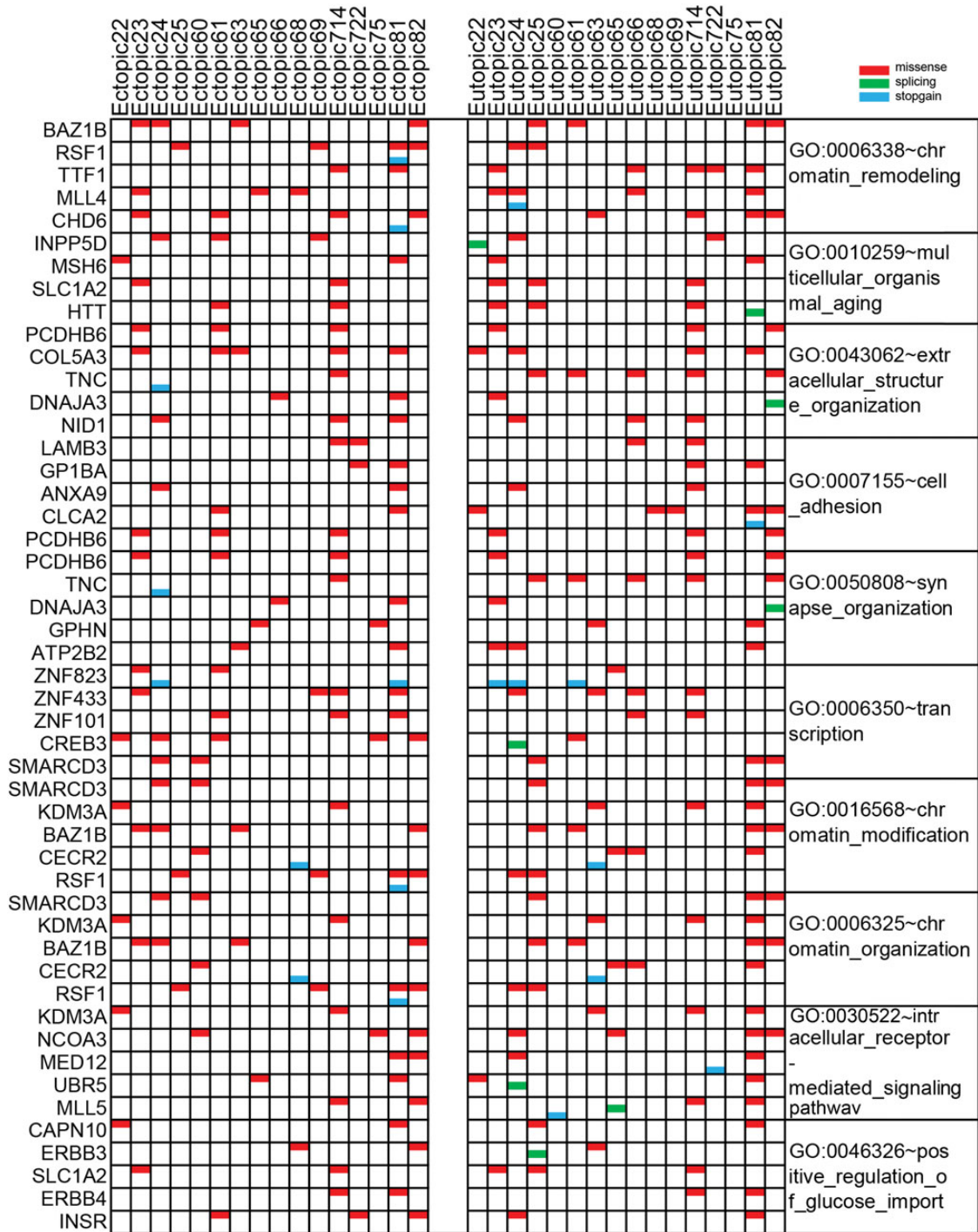


Figure 4. Integrative analysis of genomic alterations in endometriosis. The heatmap shows the pattern of case-specific, concurrent and mutually exclusive mutations in 16 eutopic endometrium and ectopic endometrium samples. Each column represents an individual affected case, and each row denotes a specific gene assigned to one of 10 functional categories (labeled on the right). Significant and selected endometriosis related somatic mutations, splicing and stop-gain in endometriosis are indicated. Mutation subtypes are denoted by color. If multiple mutations were found in a gene in a single sample, only one is shown.

both previously reported and newly identified genes coding for proteins involved in biological adhesion were recurrently mutated in endometriosis. In addition, we identified mutations in genes involved in chromatin modification, cell cycle, DNA

repair and the regulation of apoptosis. The frequent alternation of genes in these pathways may be involved in the pathogenesis of endometriosis and provide opportunities for novel diagnostics and therapies.

Compared with previously published studies

The application of high-throughput technologies for the analysis of the genetic background of subjects affected by endometriosis is revealing several genetic abnormalities of a different nature, likely at the basis of the different familial and sporadic cases of this recurrent pathology or implicated in disease susceptibility. In endometriosis, three major studies have been published to date in two different populations (European and Japanese). GWAS meta-analyses, conducted on thousands of endometriosis cases, has identified several endometriosis risk loci, including *WNT4*, *CDKN2BAS* and *FNI* (8,9,12). Our study differs from these studies; we applied the following strategies in the present study. First, we focused on patients with deep endometriosis and strictly selected fertile women with macroscopically normal pelvic cavities as controls for a preliminary screening. Correlation of basic science with a well-defined clinical population is the key to successful translational research that will lead to the development of new diagnostic and targeted therapeutic approaches (47). Second, we used LCM to harvest the desired cell type. LCM has been proven to be a tissue microdissection procedure that allows accurate, single-cell, tissue sampling from small target tissues such as endometriotic lesions (48). The purity of epithelial cell purification was confirmed by a histological review of laser capture microdissected cells as previously described (49,50). The use of LCM ensures an accurate and reliable acquisition of cells of the desired type from specific microscopic regions of tissue sections under direct visualization, which in turn, permits the molecular genetic analysis of pure populations of epithelial cells taken from lesion samples. The procedure minimizes or even eliminates any possibility of contamination. Because epithelial cells and stromal cells often have different properties (51), the use of LCM increases more specificity to our findings. In the present study, we chose to focus on epithelial cells only, using LCM to avoid stromal and host tissue cells in our preparations. Stromal cells are more difficult to separate from the underlying host tissues and inflammatory cells, even with current technology. However, there is strong evidence in the literature to support the importance of stromal cells and the host tissue microenvironment (51). It is likely that some, if not many, of the differences we found between eutopic and ectopic endometriosis from endometriosis patients could be ascribed to hormonal microenvironment differences: the ‘soil’ versus ‘seed’ analysis of this complex problem. Third, whole-exome sequencing is a powerful tool for discovering novel genetic variants that predispose to disease. To examine the genetics basis of endometriosis, we analyzed the mutation alternation in endometrium samples from 16 independent cases with endometriosis and 5 normal female cases by sequencing the exomes. To our knowledge, this is the first effort to date where endometriosis has been studied by exome sequencing to identify novel endometriosis-predisposing genes.

In this context, we found that endometriotic cells from endometriosis patients possess a unique genetic fingerprint compared with healthy endometriotic cells. Moreover, the existence of similarities and differences among frequently mutated genes in eutopic or ectopic endometrium might reflect that most of the mutations present in endometrium genomes from endometriosis patients already existed in the eutopic endometrium cell that was ‘transformed’ by the initiating mutation, and nearly all of these

mutations are probably benign and irrelevant for pathogenesis. Consistent with this hypothesis, we observed that eutopic and ectopic endometrium exomes have similar numbers of total aforementioned mutations (Fig. 1D), and eutopic exomes contain unique mutations that rarely occur in ectopic endometrium cells. Conversely, ectopic cells also contain unique mutations that are almost never present in eutopic endometrium cells. As mentioned above, there were many mutant genes that were shared between the eutopic endometrium cells and ectopic endometrium cells, suggesting that these mutations might cooperate with a variety of initiating mutations (Supplementary Material, Table S5). To a certain extent, our results suggest the following model for the pathogenesis of endometriosis: the disease commences when menstrual tissue fails to exit the body and instead travels back to the ovaries and/or peritoneal cavity, partially supporting the Sampson’s theory of retrograde menstruation, which claims that ectopic endometrial tissues originated from eutopic tissues that traveled to the most common site of implantation, the peritoneal cavity, through the fallopian tubes during menstrual shedding, where they adhere to the peritoneal cavity wall, invade the extracellular matrix, proliferate and form endometriotic lesions (52). In this study, we revealed similarities and differences of mutated genes between eutopic and ectopic endometrium cells, which could support the hypothesis that the endometrium cells that travel to other sites under certain conditions have abnormal genetics, which could contribute to ‘transforming’ eutopic endometrium into ectopic endometrium tissues. Taken together, our results represent a significant advance in our understanding of the pathogenesis of endometriosis; however, questions remain. It is unclear why only some women develop endometriosis, although retrograde menstrual flow likely occurs in most women of reproductive age.

Pathways likely to be involved in endometriosis pathogenesis

Our study has confirmed previously reported findings and newly identified genes coding for proteins involved in the pathogenesis of endometriosis. A number of molecules, phenotypically relevant to adhesion, attachment, invasion and migration, are involved in the regeneration, growth and functions of the endometrium (53–55). Notably, ectopic endometrial-tissue fragments have mechanisms by which they can attach to and invade peritoneal surfaces (56,57). Some cell-adhesion molecules, but not all, for the adhesion of endometrium to the peritoneum have been identified by gene expression analysis. The endometrial tissue from controls and women with endometriosis showed constitutive expression of laminin γ -2 (*LAMC2*), apolipoprotein (*ApoE*), integrin β -2 (*ITGB2*), integrin β -7 (*ITGB7*), laminin γ -1 (*LAMC1*) and junctional adhesion molecule-1 (*JAM-1*). Gene expression of *ApoE* and *JAM-1* was decreased in both proliferative and secretory phases in the endometrium from endometriosis patients compared with controls. Additionally, *LAMC1* expression level was reduced in the endometrium from women with endometriosis compared with control endometrium in the proliferative phase (29). In addition, a higher mRNA expression of *LAMC2* was observed in ectopic endometrium from women with endometriosis compared with eutopic endometrium from endometriosis patients (28). Intriguingly, when we reviewed our exome-sequencing data, out of the six molecules studied above, we found three genes (*LAMC1*,

LAMC2 and *JAMI*) with genetic alterations in our study (Supplementary Material, Table S9). It is tempting to speculate that either gene expression changes or genetic alternations of these adhesion, attachment and invasion proteins in endometrium from endometriosis patients may independently or collaboratively contribute to the pathogenesis of endometriosis, although we did not detect the expression levels of these genes in the present study. These results suggest that abnormal of molecules with these properties may have a role in anchoring endometrial cells at ectopic sites, thus initiating the development of endometriosis.

Endometriosis also results from elevated cellular proliferation or decreased apoptosis of the retrograded endometrium in response to appropriate stimuli. Based on the implantation theory, the survival of the refluxed endometrium in the pelvic cavity may be essential for the initial development of endometriosis. A previous study has revealed that apoptosis was decreased in eutopic endometrium from endometriosis patients compared with that from women without endometriosis during the late secretory phase (58). Additionally, microarray analysis of samples from women with endometriosis has identified gene expression changes involved in several important signaling pathways, including the rat sarcoma (RAS)/RAF/mitogen activated protein kinase (MAPK) pathway, the wingless type murine mammary tumor virus integration site family (WNT)-related pathway and the phosphoinositide-3-kinase (PI3K) pathway (59–62). Alterations in these signaling pathways contribute to the transformed phenotype including cell growth, proliferation, differentiation, cell survival, adhesion and cell motility (63–65), which may be necessary for the initial development and growth of endometriosis. In addition, other studies in the last decade have contributed to the consolidation of the hypothesis that endometriosis develops from stem/progenitor cells, through different approaches, both in experimental models of disease and in specimens collected from diseased women during surgical procedures (66,67). In this context, one recent study has investigated the differential expression of a panel of 13 stemness-related genes (*BMI1*, *ERAS*, *TCL1*, *UTF1*, *OCT4*, *SOX2*, *SOX15*, *NONOG*, *SALL4*, *DPPA2*, *GDF3*, *ZFP42* and *KLF4*) in human endometrial and endometriotic tissues, revealing an overall increased expression of these stemness-related genes in endometriotic samples, suggesting that stem cells could play a role in the pathogenesis of endometriosis (68). To interrogate whether genetic alternations also appeared in such genes, we reviewed our exome-sequencing data. However, few genes that have been previously reported by gene expression analysis were identified. There are several possible explanations for these results. First, this is the first study to investigate the spectrum of somatic alternations of epithelial cells in both eutopic and ectopic endometrium from endometriosis patients, using the combined methods of whole-exome sequencing and LCM. Second, it is easy to see that there are vast differences, at least at the genetic level, between the ectopic and eutopic endometrium, despite the fact that endometriosis is defined as the ectopic presence of endometrial glands and stroma (see Supplementary Material, Table S5). Because these vast differences are only a snapshot in the long process of the pathogenesis of endometriosis, many mutated genes, especially those involved in the initiation of the disease, may not be captured. This is especially true because our exome sequencing did not achieved a high coverage or provide enough information. In addition, our filtration analysis cannot capture any genetic changes. Some, but not all, genetic changes

that we observed are surely linked to endometriosis progression. Other mutations may also have occurred as a result of ectopic relocation of endometrial cells. Third, the heterogeneity observed in endometriosis is to be expected because, like tumors, heterogeneity among endometriotic lesions may be a result of ‘a high level of redundancy, and hence increased chances of survival and growth’ (69). Fourth, genetic alternations in those genes, which are involved in the regulation of signaling pathways (i.e., MAPK, PI3K and WNT) or in the control of stem cell properties, may be harmful for their structural and functional integrity. Endometrial cells with these gene mutations will not grow, adhere and develop endometriotic implants. In addition, the presence of cells with stemness-related gene mutations in these tissue implants cannot proliferate and do not lead to the formation of cysts or endometriosis. Although further studies are necessary to clarify the functional roles of these genes and proteins in eutopic endometrium from patients with endometriosis, the present findings suggest that biological adhesion and chromatin-remodeling pathways may be involved in the initial development of endometriosis and may be the basis for a potential novel strategy for prevention or effective targeted therapy.

SUMMARY

In summary, our study performed whole-exome analysis of both eutopic and ectopic endometrium cells from 16 endometriosis patients. Although many mutations in endometriosis have been discovered by sequencing more limited parts of exome, candidate whole-exome sequencing in our study did not achieved a high coverage or provide enough information to confidently predict the full spectrum of mutations that are required for disease in an individual patient or determine the global exomic character of a disease. More sequencing data are needed to identify additional rare mutations; however, our data suggest that most of the somatic events in endometriosis exomes are random, preexisting, background mutations in endometrium cells that acquired the key initiating mutation. Only a tiny fraction of total mutations in each endometriosis genome are therefore likely to be relevant to pathogenesis and potential targeted therapy. Our preliminary analysis provides a more comprehensive catalog of genetic alternations in endometriosis and important clues to understanding endometriosis pathogenesis that may help develop novel diagnostics and therapeutics for this disease. Notably, our results suggest that endometriosis is a disease that is caused not by hundreds of mutations but by only a few. Additionally, the comprehensive approach to acquiring data in this study provided an opportunity to investigate whether any particular new pathways play a role in endometriosis. From this analysis, we identified mutations in genes that are involved in the endometriosis process, including chromatin remodeling, biological adhesion, cell–cell junctions, histone modification and the immune system. The results may enhance our biological insight of endometriosis and will hopefully direct us towards new strategies to improve patient care.

MATERIALS AND METHODS

Method summaries

Eutopic endometrium, ectopic endometrium and their matched blood biospecimens were obtained from 16 endometriosis

patients, and normal endometrium samples were collected from five healthy women with appropriate consent from Institutional Review Boards (IRBs). In total, 2 μ g of DNA was sheared by sonication to a fragment length of 200–300 bp (peak on electrophoresis). Illumina DNA sequencing libraries were constructed from each sample using adapters suitable for ‘indexing’, which is a method for identifying samples during a multiplexed sequencing run with several samples per lane of the sequencing flow cell. Exome enrichment was performed using the Illumina TruSeq exome enrichment kit. Sequencing was performed by the Mayo Clinic Genomics Core on an Illumina HiSeq2000 instrument to 100 bp in length in paired-end mode (each read representing 100 bp \times 2).

DNA extraction and sample collection

All of the samples were obtained under IRB approval and with documented informed consent. All of the samples were fresh-frozen primary resections from individuals with endometriosis who had not been treated previously with chemotherapy or radiation. A board-certified pathologist examined hematoxylin and eosin-stained slices and selected cases with glandular-like structures in their ectopic endometrium samples. Cells surrounding the eutopic endometrium and ectopic endometrium glandular cavity from 16 endometriosis patients were isolated by LCM using the Arcturus PixCell II microscope and CapSure Macro caps (LCM 0211; Arcturus, Carlsbad, CA). Peripheral blood samples from corresponding patients were collected concurrently for white blood cell DNA extraction. Frozen sections of normal endometrium cells from five healthy women were also collected through LCM as the control samples, and the corresponding peripheral blood samples were also obtained. DNA was extracted using salt precipitation or phenol-chloroform extraction and quantified using PicoGreen dsDNA Quantitation Reagent (Invitrogen).

Whole-genome amplification (WGA) and quality control

WGA was achieved on the LCM samples using the PEPLI-g Mini Kit according to the manufacturer’s manual (QIAGEN GmbH). A reaction in 50 μ l was performed at 30°C for 16 h and terminated at 65 °C for 5 min. The concentration of DNA products was measured using the Qubit Quantization Platform (Invitrogen Life Science). Multiplex PCR containing eight housekeeping genes located on different chromosomes was performed to assess amplified DNA product coverage. The DNA products that amplified successfully with at least six housekeeping genes were selected for further library construction procedures.

Library preparation and sequencing

For each qualified DNA sample, 2 μ g of DNA was randomly fragmented by Covaris, and the library fragment size was mainly distributed between 150 and 200 bp. Adapters were ligated to both ends of the resulting fragments. The adapter-ligated templates were purified by the Agencourt AMPure SPRI beads, and fragments with insert sizes of \sim 180 bp were selected. Purified DNA was amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the Agilent SureSelect Human All Exon 50 Mb Kit for enrichment. Hybridized

fragments were bound to the streptavidin beads, whereas non-hybridized fragments were washed out after hybridization. Captured LM-PCR products were subjected to the Agilent 2100 Bioanalyzer to determine the effect of enrichment. Each captured library was then loaded on a HiSeq2000 platform, and we performed high-throughput sequencing for each captured library with the paired-end 100-bp read option to ensure that each sample met the desired average sequencing depth. Raw image files were processed by Illumina base calling Software 1.7 for base calling with default parameters, according to the manufacturer’s instructions.

Detecting somatic mutations

We aligned the whole-exome sequencing reads to the NCBI human reference genome (hg19) using Burrows–Wheeler Aligner (BWA) (70). After marking the duplicated read pairs by Mark duplication from the Picard tools package, we then performed local realignment around known indel intervals using the Genome Analysis Toolkit (71). For each endometrium sample, we called somatic mutations using the matched blood sample as a control by VarScan (v2.2) (72) after the base quality score recalibration. The raw mutations were screened via applying several heuristic rules: (i) both endometrium and matched blood samples should be covered by more than six (and $<$ 100) reads at the compared genomic positions; (ii) any base supporting variant for a given genomic position should have a quality value of no $<$ 20 in both endometrium cells and blood samples; (iii) any read supporting variant should have a mapping quality of no $<$ 20; and (iv) variants should be supported by at least 10% of the total reads from the endometrium, and no $>$ 1% variant-supporting reads in blood samples were detected at the same genome position. We then annotated the high quality somatic mutations and obtained the mutated genes. To eliminate previously described germline variants, somatic mutations were cross-referenced against the dbSNP (version 132) database and SNP datasets from the 1000 Genomes Project. Any mutations that were present in these datasets were filtered out, and the remaining mutations were subjected to subsequent analyses.

Mutation annotation

Somatic point, insertion and deletion mutations were annotated using information from publicly available databases, including the UCSC Genome Browser’s UCSC Genes track, miRBase release 15, dbSNP build 132, UCSC Genome Browser’s ORegAnno track, UniProt release 2011_03 and COSMIC v51. ANNOVAR was used to separate SNVs into different functional categories according to their genic location and their expected effect on encoded gene products, based on information from the RefSeq database information.

Mutation significance analysis

For the purpose of discovering recurrently mutated genes, we used the MutSig algorithm, as described. In short, this method builds a background model of mutational processes, which takes genome-wide variability in mutation rates into account. This was achieved by considering different covariates that affect mutation rates: GC content (measured on 100 kb windows); local relative replication time; open versus closed

chromatin status as determined by HiC (fine-scale mapping of nuclear three-dimensional DNA contacts); and gene expression and local gene density measured in a 1 Mb window. For each gene, we defined a set of nearest neighbors according to these covariates and estimated the background mutation rate from non-coding (in flanking sequences and introns) and silent mutations of these neighbors. We then assigned a score based on the ratio between the non-silent coding mutation rate of the gene and the non-coding and silent mutation rate of the given gene and its neighbors. Furthermore, we performed an independent significance analysis that was restricted to events that had been previously reported in the COSMIC database.

Power analysis of whole-exome sequencing

The power of detecting somatic mutations at each nucleotide position in whole-exome sequencing was estimated by Monte-Carlo simulation ($n = 1000$) on the basis of the observed mean coverage depth for each exon in germline and endometrium samples and the observed endometrium content for each sample, which were estimated using the observed mutation allele frequencies. For the samples with no observed somatic mutations, the average endometrium content of the informative samples was employed. Simulations were performed across a total of 192 424 exons.

Pathway analyses

Pathway analyses were performed as reported. First, a gene list was selected. We processed different combinations of three lists: (i) all relevant genes (those with mutations leading to non-synonymous substitutions or affecting exon junctions); (ii) damaged genes (those with mutations predicted to be damaging); and (iii) recurrently altered genes (those with relevant mutations in both eutopic and ectopic endometrium in more than two samples; and those with relevant mutations only either in eutopic or ectopic endometrium in more than two samples). Several sources of gene sets were selected including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Database for Annotation, Visualization and Integrated Discovery (DAVID), and Cytoscape (see URLs). The sets included groups based on molecular function, cellular localization, biological processes and signaling pathways. The resulting lists were examined for enrichment in terms from GO (biological process) and KEGG pathways. For the latter, pathways associated with diseases were filtered out as reported. Enrichment analysis was based on a hypergeometric test. P values were adjusted using Benjamini–Hochberg's FDR; only $FDR < 0.1$ was considered. A correction for genes in overlapping clusters was applied.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

URLS

The 1000 Genomes Project, <http://www.1000genomes.org/>; KEGG (Kyoto Encyclopedia of Genes and Genomes) database,

<http://www.genome.jp/kegg/>; DAVID (database for annotation, visualization and integrated discovery), <http://david.abcc.ncifcrf.gov/summary.jsp>; Cytoscape database, <http://www.cytoscape.org>; MutSig Algorithm, <http://www.broadinstitute.org/cancer/cga/mutsig>; Polyphen-2, <http://genetics.bwh.harvard.edu/pph2/>; SIFT, <http://sift.jcvi.org/>.

AUTHORS' CONTRIBUTIONS

W. Han and Y. Meng designed and supervised the study. X. Fu and X. Wang directed the study. Y. Li and Y. Meng provided the clinical specimens and the clinical annotation. L. Wang, Z. Wu, Q. Mei, X. Li and J. Nie conducted pathological review of clinical specimens. L. Zhao prepared DNA samples. Y. Zhang analyzed and interpreted data. X. Li, Y. Zhang, L. Zhao and W. Han wrote the manuscript with the assistance and final approval from all authors.

ACKNOWLEDGEMENTS

We are grateful to our colleagues for generous support, helpful discussions and review of the manuscript. We thank members of the BGI-biotech samples Platform, Genetic Analysis Platform and Genome Sequencing Platform for sequencing services. We also want to thank Hanjie Wu (BGI-Shenzhen), Na Lu (BGI-Shenzhen), Fuqiang Li (BGI-Shenzhen) and Hongmei Zhu (BGI-Health, Tianjin), who participated in the analysis of exome sequencing data analysis. We are also grateful for the physicians and hospital staff, whose efforts in collecting these samples were essential to this research.

Conflict of Interest statement: None declared.

FUNDING

This work was supported by the Grants from the National Natural Science Foundation of China (nos 31201033, 31270820 and 81230061) and is partially supported by the Grant from the National Basic Science and Development Program of China (no. 2012CB518103). This work is also supported in part by the Beijing Nova Program (no. Z141107001814104). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1. Stern, R.C., Dash, R., Bentley, R.C., Snyder, M.J., Haney, A.F. and Robboy, S.J. (2001) Malignancy in endometriosis: frequency and comparison of ovarian and extraovarian types. *Int. J. Gynecol. Pathol.*, **20**, 133–139.
2. Bulun, S.E. (2009) Endometriosis. *N. Engl. J. Med.*, **360**, 268–279.
3. Giudice, L.C. and Kao, L.C. (2004) Endometriosis. *Lancet*, **364**, 1789–1799.
4. Anonymous. (1985) Revised American Fertility Society classification of endometriosis: 1985. *Fertil. Steril.*, **43**, 351–352.
5. Bischoff, F. and Simpson, J.L. (2004) Genetic basis of endometriosis. *Ann. N. Y. Acad. Sci.*, **1034**, 284–299.
6. Zondervan, K.T., Cardon, L.R. and Kennedy, S.H. (2001) The genetic basis of endometriosis. *Curr. Opin. Obstet. Gynecol.*, **13**, 309–314.
7. Stefansson, H., Geirsson, R.T., Steinthorsdottir, V., Jonsson, H., Manolescu, A., Kong, A., Ingadottir, G., Gulcher, J. and Stefansson, K. (2002) Genetic factors contribute to the risk of developing endometriosis. *Hum. Reprod.*, **17**, 555–559.

8. Nyholt, D.R., Low, S.K., Anderson, C.A., Painter, J.N., Uno, S., Morris, A.P., MacGregor, S., Gordon, S.D., Henders, A.K., Martin, N.G. *et al.* (2012) Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat. Genet.*, **44**, 1355–1359.
9. Painter, J.N., Anderson, C.A., Nyholt, D.R., Macgregor, S., Lin, J., Lee, S.H., Lambert, A., Zhao, Z.Z., Roseman, F., Guo, Q. *et al.* (2011) Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.*, **43**, 51–54.
10. Painter, J.N., Nyholt, D.R., Morris, A., Zhao, Z.Z., Henders, A.K., Lambert, A., Wallace, L., Martin, N.G., Kennedy, S.H., Treloar, S.A. *et al.* (2011) High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to CYP2C19. *Fertil. Steril.*, **95**, 2236–2240.
11. Treloar, S.A., Wicks, J., Nyholt, D.R., Montgomery, G.W., Bahlo, M., Smith, V., Dawson, G., Mackay, I.J., Weeks, D.E., Bennett, S.T. *et al.* (2005) Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. *Am. J. Hum. Genet.*, **77**, 365–376.
12. Uno, S., Zembutsu, H., Hirasawa, A., Takahashi, A., Kubo, M., Akahane, T., Aoki, D., Kamatani, N., Hirata, K. and Nakamura, Y. (2010) A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat. Genet.*, **42**, 707–710.
13. Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C. *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153–158.
14. Zang, Z.J., Cutcutache, I., Poon, S.L., Zhang, S.L., McPherson, J.R., Tao, J., Rajasegaran, V., Heng, H.L., Deng, N., Gan, A. *et al.* (2012) Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.*, **44**, 570–574.
15. Cancer Genome Atlas, N. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330–337.
16. Cancer Genome Atlas Research Network, Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R. *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**, 67–73.
17. Sato, Y., Yoshizato, T., Shiraiishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H. *et al.* (2013) Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.*, **45**, 860–867.
18. Ojesina, A.I., Lichtenstein, L., Freeman, S.S., Pedamallu, C.S., Imaz-Rosshandler, I., Pugh, T.J., Cherniack, A.D., Ambrogio, L., Cibulskis, K., Bertelsen, B. *et al.* (2014) Landscape of genomic alterations in cervical carcinomas. *Nature*, **506**, 371–375.
19. Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A. *et al.* (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**, 1801–1806.
20. Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varella, I., Lin, M.L., Ordonez, G.R., Bignell, G.R. *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, **463**, 191–196.
21. Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E., McLellan, M.D., Dooling, D., Abbott, R. *et al.* (2012) Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.*, **366**, 1090–1098.
22. Welch, J.S., Ley, T.J., Link, D.C., Miller, C.A., Larson, D.E., Koboldt, D.C., Wartman, L.D., Lamprecht, T.L., Liu, F., Xia, J. *et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell*, **150**, 264–278.
23. Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N. *et al.* (2006) The consensus coding sequences of human breast and colorectal cancers. *Science*, **314**, 268–274.
24. Cancer Genome Atlas Research, N. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
25. Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C. *et al.* (2011) The genomic complexity of primary human prostate cancer. *Nature*, **470**, 214–220.
26. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467–472.
27. Cancer Genome Atlas Research, N. (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**, 519–525.
28. Locci, R., Nisolle, M., Angioni, S., Foidart, J.M. and Munaut, C. (2013) Expression of the gamma 2 chain of laminin-332 in eutopic and ectopic endometrium of patients with endometriosis. *Reprod. Biol. Endocrinol.*, **11**, 94.
29. Sundqvist, J., Andersson, K.L., Scarselli, G., Gemzell-Danielsson, K. and Lalitkumar, P.G. (2012) Expression of adhesion, attachment and invasion markers in eutopic and ectopic endometrium: a link to the aetiology of endometriosis. *Hum. Reprod.*, **27**, 2737–2746.
30. Ng, P.C. and Henikoff, S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.
31. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
32. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z. and Galon, J. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **25**, 1091–1093.
33. Bindea, G., Galon, J. and Mlecnik, B. (2013) CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*, **29**, 661–663.
34. Vercellini, P., Vigano, P., Somigliana, E. and Fedele, L. (2013) Endometriosis: pathogenesis and treatment. *Nat. Rev. Endocrinol.*, **10**, 261–275.
35. Delbandi, A.A., Mahmoudi, M., Shervin, A., Akbari, E., Jeedi-Tehrani, M., Sankian, M., Kazemnejad, S. and Zarnani, A.H. (2013) Eutopic and ectopic stromal cells from patients with endometriosis exhibit differential invasive, adhesive, and proliferative behavior. *Fertil. Steril.*, **100**, 761–769.
36. Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C. *et al.* (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, **463**, 360–363.
37. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F. *et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.*, **44**, 760–764.
38. Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li, X., Zhou, L. *et al.* (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.*, **43**, 875–878.
39. Jones, S., Wang, T.L., Shih Ie, M., Mao, T.L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L.A. Jr., Vogelstein, B. *et al.* (2010) Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science*, **330**, 228–231.
40. Le Gallo, M., O'Hara, A.J., Rudd, M.L., Urick, M.E., Hansen, N.F., O'Neil, N.J., Price, J.C., Zhang, S., England, B.M., Godwin, A.K. *et al.* (2012) Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat. Genet.*, **44**, 1310–1315.
41. Church, D.N., Briggs, S.E., Palles, C., Domingo, E., Kearsey, S.J., Grimes, J.M., Gorman, M., Martin, L., Howarth, K.M., Hodgson, S.V. *et al.* (2013) DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum. Mol. Genet.*, **22**, 2820–2828.
42. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
43. Huang da, W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C. *et al.* (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, **35**, W169–W175.
44. Giudice, L.C. (2010) Clinical practice. Endometriosis. *N. Engl. J. Med.*, **362**, 2389–2398.
45. Dinulescu, D.M., Ince, T.A., Quade, B.J., Shafer, S.A., Crowley, D. and Jacks, T. (2005) Role of K-ras and Pten in the development of mouse models of endometriosis and endometrioid ovarian cancer. *Nat. Med.*, **11**, 63–70.
46. Wiegand, K.C., Shah, S.P., Al-Agha, O.M., Zhao, Y., Tse, K., Zeng, T., Senz, J., McConechy, M.K., Anglesio, M.S., Kalloger, S.E. *et al.* (2010) ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.*, **363**, 1532–1543.

47. Murphy, A.A. (2002) Clinical aspects of endometriosis. *Ann. N. Y. Acad. Sci.*, **955**, 1–10.
48. Bonner, R.F., Emmert-Buck, M., Cole, K., Pohida, T., Chuaqui, R., Goldstein, S. and Liotta, L.A. (1997) Laser capture microdissection: molecular analysis of tissue. *Science*, **278**, 1481, 1483.
49. Wu, Y., Basir, Z., Kajdacsy-Balla, A., Strawn, E., Macias, V., Montgomery, K. and Guo, S.W. (2003) Resolution of clonal origins for endometriotic lesions using laser capture microdissection and the human androgen receptor (HUMARA) assay. *Fertil. Steril.*, **79**(Suppl. 1), 710–717.
50. Saare, M., Soritsa, D., Vaidla, K., Palta, P., Remm, M., Laan, M., Karro, H., Soritsa, A., Salumets, A., D'Hooghe, T. *et al.* (2012) No evidence of somatic DNA copy number alterations in eutopic and ectopic endometrial tissue in endometriosis. *Hum. Reprod.*, **27**, 1857–1864.
51. Cooke, P.S., Buchanan, D.L., Young, P., Setiawan, T., Brody, J., Korach, K.S., Taylor, J., Lubahn, D.B. and Cunha, G.R. (1997) Stromal estrogen receptors mediate mitogenic effects of estradiol on uterine epithelium. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 6535–6540.
52. Sampson, J.A. (1927) Peritoneal endometriosis due to the menstrual dissemination of endometrial tissue into the peritoneal cavity. *Am. J. Obstet. Gynecol.*, **14**, 422–469.
53. Kyama, C.M., Overbergh, L., Mihalyi, A., Meuleman, C., Mwenda, J.M., Mathieu, C. and D'Hooghe, T.M. (2008) Endometrial and peritoneal expression of aromatase, cytokines, and adhesion factors in women with endometriosis. *Fertil. Steril.*, **89**, 301–310.
54. Quenby, S., Anim-Somuah, M., Kalumbi, C., Farquharson, R. and Aplin, J.D. (2007) Different types of recurrent miscarriage are associated with varying patterns of adhesion molecule expression in endometrium. *Reprod. Biomed. Online*, **14**, 224–234.
55. Singh, H. and Aplin, J.D. (2009) Adhesion molecules in endometrial epithelium: tissue integrity and embryo implantation. *J. Anat.*, **215**, 3–13.
56. Witz, C.A., Montoya-Rodriguez, I.A. and Schenken, R.S. (1999) Whole explants of peritoneum and endometrium: a novel model of the early endometriosis lesion. *Fertil. Steril.*, **71**, 56–60.
57. Witz, C.A., Takahashi, A., Montoya-Rodriguez, I.A., Cho, S. and Schenken, R.S. (2000) Expression of the alpha2beta1 and alpha3beta1 integrins at the surface of mesothelial cells: a potential attachment site of endometrial cells. *Fertil. Steril.*, **74**, 579–584.
58. Dmowski, W.P., Ding, J., Shen, J., Rana, N., Fernandez, B.B. and Braun, D.P. (2001) Apoptosis in endometrial glandular and stromal cells in women with and without endometriosis. *Hum. Reprod.*, **16**, 1802–1808.
59. Li, M.Q., Li, H.P., Meng, Y.H., Wang, X.Q., Zhu, X.Y., Mei, J. and Li, D.J. (2012) Chemokine CCL2 enhances survival and invasiveness of endometrial stromal cells in an autocrine manner by activating Akt and MAPK/Erk1/2 signal pathway. *Fertil. Steril.*, **97**, 919–929.
60. Aghajanova, L., Tatsumi, K., Horcajadas, J.A., Zamah, A.M., Esteban, F.J., Herndon, C.N., Conti, M. and Giudice, L.C. (2011) Unique transcriptome, pathways, and networks in the human endometrial fibroblast response to progesterone in endometriosis. *Biol. Reprod.*, **84**, 801–815.
61. Wu, Y., Kajdacsy-Balla, A., Strawn, E., Basir, Z., Halverson, G., Jailwala, P., Wang, Y., Wang, X., Ghosh, S. and Guo, S.W. (2006) Transcriptional characterizations of differences between eutopic and ectopic endometrium. *Endocrinology*, **147**, 232–246.
62. Sanchez, A.M., Vigano, P., Quattrone, F., Pagliardini, L., Papaleo, E., Candiani, M. and Panina-Bordignon, P. (2014) The WNT/beta-catenin signaling pathway and expression of survival promoting genes in luteinized granulosa cells: endometriosis as a paradigm for a dysregulated apoptosis pathway. *Fertil. Steril.*, doi: 10.1016/j.fertnstert.2014.02.040.
63. Cantley, L.C. (2002) The phosphoinositide 3-kinase pathway. *Science*, **296**, 1655–1657.
64. Chang, L. and Karin, M. (2001) Mammalian MAP kinase signalling cascades. *Nature*, **410**, 37–40.
65. Reya, T. and Clevers, H. (2005) WNT signalling in stem cells and cancer. *Nature*, **434**, 843–850.
66. Maruyama, T. and Yoshimura, Y. (2012) Stem cell theory for the pathogenesis of endometriosis. *Front. Biosci.*, **4**, 2854–2863.
67. Forte, A., Cipollaro, M. and Galderisi, U. (2014) Genetic, epigenetic and stem cell alterations in endometriosis: new insights and potential therapeutic perspectives. *Clin. Sci.*, **126**, 123–138.
68. Forte, A., Schettino, M.T., Finicelli, M., Cipollaro, M., Colacurci, N., Cobellis, L. and Galderisi, U. (2009) Expression pattern of stemness-related genes in human endometrial and endometriotic tissues. *Mol. Med.*, **15**, 392–401.
69. Kitano, H. (2003) Cancer robustness: tumour tactics. *Nature*, **426**, 125.
70. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
71. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
72. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L. and Wilson, R.K. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome res.*, **22**, 568–576.