MUC4 gene polymorphisms associate with endometriosis development and infertility

Cherry Ying-Yi Chang[1,2], Hui-Wen Chang[3], Chih-Mei Chen[4], Chia-Ying Lin[4], Chih-Ping Chen[5], Chih-Ho Lai[6], Wei-Yong Lin[4,7], Hsing-Ping Liu[8], Jim Jinn-Chyuan Sheu[4,9,10,§], Fuu-Jen Tsai[4,11,§]

[1]Department of Obstetrics and Gynecology, China Medical University, Taichung, Taiwan

[2]Institute of Environmental Health, China Medical University, Taichung, Taiwan

[3]Department of Pathology, China Medical University, Taichung, Taiwan

[4]Human Genetic Center, China Medical University Hospital, Taichung, Taiwan

[5]Department of Obstetrics and Gynecology, Mackay Memorial Hospital, Taipei, Taiwan

[6]School of Medicine, China Medical University, Taichung, Taiwan

[7]Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan

[8]Graduate Institute of Acupuncture Science, China Medical University, Taichung, Taiwan

[9]School of Chinese Medicine, China Medical University, Taichung, Taiwan

[10]Department of Health and Nutrition Biotechnology, Asia University, Taichung, Taiwan

[11]School of Post-Baccalaureate Chinese Medicine, China Medical University, Taichung, Taiwan

[§]Corresponding authors

Email addresses:

CYYC: changyinyi@hotmail.com

HWC: huiwencmu@gmail.com

CMC: t6783@mail.cmuh.org.tw

CYL: poppyrabbit97@yahoo.com.tw

CPC: chihpingcmu@gmail.com

CHL: chl@mail.cmu.edu.tw

WYL: wylin@www.cmuh.org.tw

HPL: hpliu@mail.cmu.edu.tw

JJCS: jimsheu@mail.cmu.edu.tw

FJT: d0704@mail.cmuh.org.tw

Abstract

Background

Mucin 4 (MUC4) plays an important role in protecting and lubricating epithelial surface of reproductive tracts, but its role in the pathogenesis of endometriosis is largely unknown.

Methods

To correlate *MUC4* polymorphism with risk of endometriosis and endometriosis-related infertility, we performed a case-control study of 140 patients and 150 healthy women. Six unique SNPs (rs882605, rs1104760, rs2688513, rs2246901, rs2258447 and rs2291652) were selected for this study. DNA fragments containing the target SNP sites were amplified by PCR using the *Taqman* SNP genotyping assay system to evaluate allele frequency and distribution of genotype in *MUC4* polymorphisms.

Results

Both T/G genotype of rs882605 and frequency of haplotype T-T (rs882605 and rs1104760) were higher in patients than in controls and were statistically significant. The frequency of C allele at rs1104760, C allele at rs2688513, G allele at rs2246901 and A allele at rs2258447 were associated with advanced-stage of endometriosis. Moreover, G allele at rs882605 was verified as a key genetic factor for infertility in patients. Protein sequence analysis indicated that amino acid substitutions by genetic variations at rs882605, rs2688513 and rs2246901 locate in the putative functional loops and the VWFD domain in MUC4 sequence.

Conclusions

MUC4 polymorphisms are associated with endometriosis development and endometriosis-related infertility in Taiwanese population.

Background

Endometriosis is a common chronic gynecologic disease characterized by the presence of endometrial gland and stroma outside the uterine cavity, affecting approximately 10% reproductive age women (1, 2). The common clinical symptoms include pelvic pain, heavy menstrual bleeding, pelvic adhesion, bloating and fatigue. Notably, the prevalence of endometriosis is 0.5-5% in fertile and 25-40% in infertile women (3), suggesting infertility as one possible consequence of endometriosis. To date, the implantation theory is widely accepted, stating that endometrial tissues pass through fallopian tube, then attach and grow on pelvic tissue. However, the hypothesis can not explain the existence of endometriosis outside the pelvis, and how endometriosis progress and invade to other tissues. Additional factors like genetic or immune differences were suggested as possible contributors to trigger the formation of endometriosis (4-6). Family history and genomewide linkage studies also support genetic predisposition during the development of endometriosis (7-10). These studies provide molecular evidence demonstrating endometriosis as a genetic disease, and it is desirable to explore more genetic variations associated with endometriosis.

Similar to malignant diseases, extensive growth of endometrial cell on peritoneal surface and invasion of pelvic organ are very common during the development of endometriosis. This process is frequently associated with several mechanisms involved in angiogenesis and cellular adhesion. In fact, women who have endometriosis appear more at risk of developing several different kinds of ovarian cancers (8, 11-13). Epidemiology study showed the prevalence of endometriosis in patients with endometrioid and clear cell ovarian carcinoma is 19 and 35.9, respectively (14). These findings

suggest that endometriosis and certain types of ovarian cancer may share several common genetic alterations during pathogenesis. Genes that regulate cell mobility and invasion in ovarian cancers are therefore possible candidates to play roles in endometriosis.

Mucins are high molecular weight glycoproteins with function of protecting and lubricating epithelial surface of respiratory, gastrointestinal and reproductive tracts (15). Among the mucin proteins, MUC4 and MUC1 are the major ones expressed in the endometrial epithelium (16, 17). In cancer study, these two mucins have been shown to be aberrantly expressed in various malignancies, and validated as a novel target for cancer diagnosis and therapy (18-20). Distinct from MUC1, the extracellular domain of MUC4 can interact with HER2 on the cell surface and modulate the downstream cell growth signaling via stabilizing/enhancing the activity of cell growth receptor complexes (18, 21, 22). Consequently, changes of cytoarchitectures and cellular signaling may lead to the increase of cell mobility and tumor cell invasion.

The above findings provide us clues to hypothesize that genetic variations in the extracellular domain of MUC4, especially those resulting in amino acid substitutions, may play roles involved in the development of endometriosis. With endometriosis as a possible cause of infertility in women, we also would like to study the association of *MUC4* SNPs with the susceptibility to endometriosis-related infertility.

Methods

Study population

A total of 140 individuals who underwent surgery due to benign diseases and pathology-proven endometriosis were identified at China Medical University Hospital from 1998 to 2008 and enrolled in this study. In general, these patients were diagnosed with ovarian cyst by sonography and suffered from several clinical symptoms related to endometriosis including dysmenorrhea, lower abdominal pain, infertility or abnormal menstruation. Study patient subjects who failed to be pathology-proven endometriosis were excluded in this study. For the control group, blood samples of 150 healthy women were selected from a pool of persons who received regular heath checkup at the same hospital and identified as normal based on the examines conducted. A total of 142 control subjects were frequency-matched for age profile with patients (Supplementary Table 1). Control subjects who showed one of the endometriosis-associated symptoms, even though the results of their health checkups showed normal, were excluded in this study. This study was approved by the Institutional Review Board at China Medical University, with informed consent obtained from each patient.

Clinical stages and association study

Clinical information on patients was collected from clinical notes, including clinical stage, lesion size, location, drug treatment and fertility (Supplementary Table 1). Definition of endometriosis staging was based on the classification of the American Society for Reproductive Medicine: Stage 1:

minimal; Stage 2: mild; Stage 3: moderate and Stage 4: severe (23).

Genomic DNA extraction and genotyping of SNPs in *MUC4*

Genomic DNA was extracted from peripheral blood leukocytes according to standard protocols (Genomic DNA kit; Qiagen, Valencia, CA, USA). DNA fragments containing the target SNP sites were amplified by PCR using the *Taqman* SNP genotyping assay system from Applied Biosystems, Inc (Carlsbad, CA, USA). Probe search and design are available on its website (https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=ABGTKeywordSearch). Supplementary Table 2 lists probe IDs for the six SNPs tested. PCR amplification conditions consisted of initial denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 10 s, 56°C for 10 s, and 72°C for 20 s, with one additional cycle of 72°C for 5 min. Genetic variations were detected by reading the fluorescence signals of PCR products. A positive signal indicates a perfect match between the probe and the tested DNA, thus identifying the allele types. Ten percent of study subjects were randomly chosen and genotyped in duplicate to confirm the concordance of the genotyping results. In our study, the call rates for these SNP probes were above 94% (Supplementary Table 3).

Statistic analysis

The allelic frequency and genotype frequency distributions for the six polymorphisms of endometriosis patients and controls were performed by χ^2 analysis using SPSS software (version

10.0, SPSS Inc. Chicago, Illinois, US). An unordered, two degree-of-freedom, and two-sided test was used for the statistical analyses of our genotyping results. A P value < 0.05 was considered statistically significant. Allelic and genotype frequencies are expressed as percentages of total alleles and genotypes. Odds ratios (OR) were calculated from allelic and genotype frequencies with 95% Confident Interval (95% CI). The major (also the wild type) allele was used as the reference for the allelic analyses. For the genotype analyses, the homozygous major allele genotype was used as the referent group. Adherence to the Hardy-Weinberg equilibrium (HWE) constant was tested, using χ^2 test with one degree of freedom. To study the association of the six SNPs with clinical stages and reproductive ability, Fisher's exact tests, instead of χ^2 tests, were used due to small number of subjects tested.

Haplotypes of each individual were determined by Bayesian statistical method available in the program PHASE 2.1 (24). This approach incorporates a priori expectations of haplotypic structure based on population genetics and coalescence theory. Lewontin's D' (|D'|) and the linkage disequilibrium (LD) coefficient r^2 were determined between selected pairs of biallelic loci (25). Haploview version 3.2 (Whitehead Institute for Biomedical Research, Cambridge, MA) was used to examine the structure of the LD block (26). This program uses two-marker expectation maximization to estimate the maximum likelihood values of the four gamete frequencies from which the D' and log of odds (LOD) values are derived. The genetic effects of the inferred haplotypes were analyzed in the same way as was applied to analysis of polymorphisms. The reported haplotype percents are estimated percents based on allele frequencies and linkage

disequilibrium. The *p* values are based on a comparison of a given haplotype with all other haplotypes combined.

Functional analyses and secondary structure predictions of MUC4 protein

Functional characterization and annotation of MUC4 were performed by aligning the sequence with functional motifs/signatures in PROSITE protein domain database (27). To predict the secondary structure of MUC4 sequence, the Chou-Fasman method was used (28). An improved method was applied to increase accuracy of the predictions by locating nucleation regions with refined wavelet transform technology and by calculating propensity factors with larger data set (29). The program gives propensity of each residue to be a part of an α-helix, a β-strand or a loop. We considered propensities P_α greater than 1.03 as significant for helix, and propensities P_β greater than 1.05 as strand. Predicted regions with less than four contiguous residues were not considered secondary structure units. For a region with both helix and strand tendencies, the secondary structure conformed with higher propensity; $P_\alpha > P_\beta$ or $P_\beta > P_\alpha$, is predicted. To plot hydrophobicity and surface probability, Kyte-Doolittle method (30) and Emini surface accessibility prediction (SAP) (31) were used, respectively. We slide a window along MUC4 sequence to assign ”hydrophobicity” or “surface probability” value to each amino acid. The values were summed in the window, and the results were plotted.

Results

MUC4 gene polymorphisms and endometriosis

Six SNPs in the extracellular domain of *MUC4* gene with frequency more than 20% in Chinese Han Beijing were selected from International HapMap Project databank (http://hapmap.ncbi.nlm.nih.gov) (Supplementary Table 2). Genotype analyses (Fig. 1) indicated rs882605 as a unique SNP with higher frequency of TG genotype in patients than in controls (P = 0.04; OR = 1.97, 95% IC: 1.17 – 3.32) (Table 1), while allele type analyses of these SNPs showed no statistical significance. Of note, the major (also the wild type) allele was used as the reference for the allelic analyses. For the genotype analyses, the homozygous major allele genotype was used as the referent group. To confirm the genetic impact of SNPs on endometriosis, top-two high-risk alleles at rs882605 and rs1104760 were selected for haplotype analyses. Significantly, the frequency of haplotype T-T was found higher in patients than in controls (P = 0.0353) (Table 2) (Supplementary Fig. 1), suggesting the association of *MUC4* SNPs and endometriosis development. The genotyping results were confirmed in duplicate, and the concordance of duplicates was 97.6%.

Association of *MUC4* gene polymorphisms and stages

We next asked whether *MUC4* genetic variations could possibly associate with clinical stages, patients were divided into two groups: the mild stage group with patients at Stages 1 or 2 and the advanced group with patients at Stages 3 or 4. Strikingly, genotype analyses revealed strong association of CC type at rs2688513 (P = 0.04) and GG type at rs2246901 (P = 0.03) with more

advanced endometriosis at Stages 3 or 4 (Table 3). Dominant effects were found for other genetic variations at rs1104760 (CC + CT *vs.* TT) and rs2258447 (AA + AG *vs.* GG) during endometriosis progression. Allele type analyses suggested C allele at rs1104760, C allele at rs2688513, G allele at rs2246901 and A allele at rs2258447 as risk factors that correlated with more severe endometriosis.

MUC4 gene polymorphisms and infertility

Since endometriosis has been suspected as one potent factor leading to infertility in women (3), we also studied the possible linkage between *MUC4* SNPs and infertility. Though no significant difference was found in genotype association study, our data indicated T allele at rs882605 as a protective factor that associated with reduced frequency of infertility in endometriosis patients (Table 4). Two other alleles, C at rs2688513 and G at rs2246901, showed similar protective effect, but the data did not reach statistical significance. Of note, the major (also the wild type) allele was used as the reference for the allelic analyses. For the genotype analyses, the homozygous major allele genotype was used as the referent group. Haplotype analyses thus sought to ascertain the impact of genetic combination of these top-three protective alleles. Table 5 indicated that patients with haplotype T-C-G did show lower frequency of infertility, although the results did not attain statistical difference (P = 0.099). By contrast, haplotype G-T-T showed strong association with infertility in patients (P = 0.012) (Table 5) (Supplementary Fig. 2) and could be used as a risk indicator for patients at higher risk to develop severe complications such as infertility.

MUC4 gene polymorphisms and amino acid substitutions

Because these endometriosis-associated SNPs can cause amino acid substitutions (supplementary Table 2), the biofunctions of MUC4 might be altered by changes of hydrophilicity and protein folding. Figure 2 illustrates the functional domains in MUC4 protein sequence and secondary structures that contain these SNPs. Our data showed that genetic variations of rs882605 and rs2688513 cause amino acid substitutions in long-loop regions (>10 residues) between secondary structure units (α-helices or β-strands) with high hydrophilicity and moderate surface probability (Fig. 2). Amino acid composition analyses also revealed that these two loops (300-319 and 4134-4158) contain several negatively charged Asp (D) residues and reverse turn elements, Gly (G) or Pro (P), suggesting the importance of these loops to protein folding and functional regulation (32, 33). In addition, rs2246901 locates in a VWFD domain responsible for protein-protein interaction and cell adhesion/migration (34, 35). Our findings support functional roles of *MUC4* SNPs in regulating cellular mobility and invasion of endometrial cells during endometriosis development and progression.

Discussion

Previous studies showed polymorphism of cytokines and adhesion molecule which were associated with the pathogenesis of endometriosis (36, 37). To the best of our knowledge, however, no study investigated the possible association of *MUC4* and endometriosis. The purpose of this study was to evaluate whether genetic variation in *MUC4* associates with endometriosis in the Taiwanese population. Our data proved the association of *MUC4* polymorphisms with advanced stages of endometriosis and the related infertility. Since the extracellular domain of MUC4 is critical for HER2 interaction and cell invasiveness, these defined SNPs located in putative functional domains of MUC4 may play important roles during endometriosis development and progression.

The development of endometriosis and certain types of ovarian cancer share several similar clinical features. For example, endometriosis could progressively invade to pelvic viscera, resulting in adhesion and recur after medical treatment or operation. Due to the functions involved in acquisition of adhesion ligands or receptors and loss of anti-adhesion, proteins like MUC4 and MUC1, the two major mucins present in endometrial epithelium, thus become suspect for endometriosis development (16, 17). In addition to gene overexpression, genetic variation in *MUC1* has also been reported as a risk factor contributing to cell mobility and severeness of cancer (38-41). However, the influence of MUC4 genetic variation on cell behavior remains unclear. In this study, Pro4135Ser (rs2688513) and Ala4693Ser (2246901) substitutions in the putative functional domains of MUC4 were found associated with advanced stages of endometriosis. Since MUC4 is an emerging target for ovarian cancer (18, 19, 42, 43), our study provides a new direction to address

the roles of MUC4 in the development of gynecological disorders.

To study the genetic effects of mucin proteins by SNPs, the major interest focuses on the variable number of tandem repeat (VNTR) polymorphisms, which result in different-sized gene transcripts. For examples, *MUC1* variations in the VNTR domain have been found to play roles in regulating *H. pylori* binding to gastric cells (44). Other studies also conclude that VNTR polymorphisms can influence T-antigen presentation and the local immune responses, which consequently have potential impacts on gastric cancer development (45, 46). With regard to *MUC4*, a high degree of polymorphism in the VNTR domain was observed in human tissues including endometrial epithelium (47, 48). However, the different-sized *MUC4* transcripts did not show association with embryo implantation or cancer development. By contrast, MUC4 can promote cell proliferation and anti-apoptotic effects in cancer cells via interacting with HER2 on the cell surface (18, 19, 22), suggesting the potency of functional domains in the extracellular domain of MUC4. In this study, two SNPs (rs2688513 and rs2246901) that locate in a putative functional loop and the VWFD protein binding domain, respectively, were found associated with advanced stages of endometriosis. Further study may clarify whether these amino acid substitutions could change the interaction with HER2 and/or play crucial roles in regulating cellular activity of endometriosis spread.

Endometriosis could cause pelvic adhesion and tubal occlusion that may lead to infertility. However, among patients with endometriosis related infertility, 50-60% of them were diagnosed as minimal or mild endometriosis (3). Impaired folliculogenesis, bad oocyte quality, and impaired

implantation of embryo are therefore considered the possible mechanisms for endometriosis related infertility. Changes of cytokines and growth factors in endometrium, follicular fluid, and peritoneal fluid have been suggested as the key players for inducing above phenomena (5). Recently, several studies showed that MUC 4 could promote cell migration, change endometrial environment and create weak points in epithelium, thus facilitating the failure of embryo implantation (48, 49). Carryway et al study also showed that embryo implantation was associated with down-regulation of MUC4 expression in an animal model (50). In this study, women with T allele in rs882605 had lower risk of endometriosis-related infertility; whereas rs2688513 and rs2246901 SNPs did not show any association with the reproductive ability in patients. The rs882605 SNP locates in a putative functional loop within the VNTR domain of MUC4 that may control T-cell antigen presentation and the local immune responses. Our findings may support the view that the regulation of local immunity, rather than uncontrolled cell proliferation, in endometrium may play a more important role in the development of endometriosis-related infertility.

Our study showed that T/G genotype at rs882605 is unique in patients with endometriosis, as compared to T/T or G/G genotypes. So far, our data could not provide sufficient information to explain why T/T genotype dose not show higher risk of endometriosis than T/G. One reason could be due to the relatively small study group, while other possibility could exist. For examples, the SNPs analyzed are in tight linkage disequilibrium with other unknown allele variants which impart an opposite effect. In this case, only individuals with the T/G heterozygous genotype can be observed. Because this is a hospital-based study with a modest sample size, enrollment of a larger

cohort based on a population approach could help to elucidate the functional role of MUC4 in endometriosis and the related infertility.

Conclusions

This study observed an association of *MUC4* polymorphism with endometriosis development and endometriosis-related infertility in Taiwanese population. However, the true mechanism of how MUC4 modulate the pathogenesis of endometriosis and infertility was not clearly understood. In addition, the risk SNPs for stages and infertility differ, suggesting dissimilar molecular mechanisms for these clinical features. More detailed studies are needed to investigate the biochemical pathways regulated by MUC4 during the development of endometriosis.

Competing interests

The authors declare that they have no conflict of interest.

Authors' contributions

CYYC collected samples, carried out clinical association study, and drafted the manuscript. HWC participated in sample collection and clinical association study. CMC carried out sample preparation and SNP analyses. CYL participated in SNP analyses. CPC participated in clinical association study. CHL carried out statistical analyses. WYL carried out sample pretreatment and RNA extraction. JJCS carried out experimental design and drafted the manuscript. FJT carried out experimental design and revised the manuscript.

Acknowledgements

The authors appreciate technical assistance from Ms. Carmen Chan at China Medical University Hospital. This study was funded by a grant (DMR-98-073) from China Medical University Hospital in Taichung, Taiwan.

References

1. Ballard K, Lowton K, Wright J: **What's the delay? A qualitative study of women's experiences of reaching a diagnosis of endometriosis.** *Fertil Steril* 2006, **86**:1296-1301.

2. Greene R, Stratton P, Cleary SD, Ballweg ML, Sinaii N: **Diagnostic experience among 4,334 women reporting surgically diagnosed endometriosis.** *Fertil Steril* 2009, **91**:32-9.

3. Holoch KJ, Lessey BA: **Endometriosis and infertility.** *Clin Obstet Gynecol* 2010, **53**:429-438.

4. Vercellini P, Somigliana E, Vigano P, Abbiati A, Barbara G, Crosignani PG: **Endometriosis: current therapies and new pharmacological developments.** *Drugs* 2009, **69**:649-675.

5. Gupta S, Goldberg JM, Aziz N, Goldberg E, Krajcir N, Agarwal A: **Pathogenic mechanisms in endometriosis-associated infertility.** *Fertil Steril* 2008, **90**:247-257.

6. Kyama CM, Debrock S, Mwenda JM, D'Hooghe TM: **Potential involvement of the immune system in the development of endometriosis.** *Reprod Biol Endocrinol* 2003, **1**:123.

7. Hadfield RM, Mardon HJ, Barlow DH, Kennedy SH: **Endometriosis in monozygotic twins.** *Fertil Steril* 1997, **68**:941-942.

8. Matalliotakis IM, Cakmak H, Krasonikolakis GD, Dermitzaki D, Fragouli Y, Vlastos G *et al.*: **Endometriosis related to family history of malignancies in the Yale series.** *Surg Oncol* 2010, **19**:33-37.

9. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V *et al.*: **Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26**. *Am J Hum Genet* 2005, **77**:365–376.

10. Kashima K, Ishimaru T, Okamura H, Suginami H, Ikuma K, Murakami T *et al.*: **Familial risk among Japanese patients with endometriosis.** *Int J Gynaecol Obstet* 2004, **84**: 61–64.

11. van Gorp T, Amant F, Neven P, Vergote I, Moerman P: **Endometriosis and the development of malignant tumours of the pelvis.** *Best Pract Res Clin Obstet Gynaecol* 2004, **18**:349-371.

12. Kobayashi H, Kajiwara H, Kanayama S, Yamada Y, Furukawa N, Noguchi T *et al.*: **Molecular pathogenesis of endometriosis-associated clear cell carcinoma of the ovary.** *Oncol Rep* 2009, **22**:233-240.

13. Nezhat F, Datta MS, Hanson V, Pejovic T, Nezhat C: **The relationship of endometriosis and ovarian malignancy: a review.** *Fertil Steril* 2008, **90**:1559-1570.

14. Orezzoli JP, Russell AH, Oliva E, Del Carmen MG, Eichhorn J, Fuller AF: **Prognostic implication of endometriosis in clear cell carcinoma of the ovary.** *Gynecol Oncol* 2008, **110**:336-344.

15. Yonezawa S, Higashi M, Yamada N, Yokoyama S, Goto M: **Significance of mucin expression in pancreatobiliary neoplasms.** *J Hepatobiliary Pancreat Sci* 2010, **17**:108-124.

16. Gollub EG, Goswami S, Kouba D, Marom Z: **Regulation of mucin gene expression in secretory epithelial cells.** *Biochem Biophys Res Commun* 1993, **197**:667-673.

17. Audie JP, Tetaert D, Pigny P, Buisine MP, Janin A, Aubert JP *et al.*: **Mucin gene expression in the human endocervix.** *Hum Reprod* 1995, **10**:98-102.

18. Bafna S, Kaur S, Batra SK: **Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells.** *Oncogene* 2010, **29**:2893-2904.

19. Singh AP, Chaturvedi P, Batra SK: **Emerging roles of MUC4 in cancer: a novel target for diagnosis and therapy.** *Cancer Res* 2007, **67**:433-436.

20. van Elssen CH, Frings PW, Bot FJ, van de Vijver KK, Huls MB, Meek B *et al.*: **Expression of aberrantly glycosylated Mucin-1 in ovarian cancer.** *Histopathology* 2010, **57**:597-606.

21. Chaturvedi P, Singh AP, Batra SK: **Structure, evolution, and biology of the MUC4 mucin.** *FASEB J* 2008, **22**:966-981.

22. Chaturvedi P, Singh AP, Chakraborty S, Chauhan SC, Bafna S, Meza JL *et al.*: **MUC4 mucin interacts with and stabilizes the HER2 oncoprotein in human pancreatic cancer cells.** *Cancer Res* 2008, **68**:2065-2070.

23. Schenken RS: **Modern concepts of endometriosis. Classification and its consequences for therapy.** *J Reprod Med* 1998, 43:269-275.

24. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68**:978–989.

25. Hedrick PW: **Gametic disequilibrium measures: proceed with caution.** *Genetics* 1987, **117**:331–341.

26. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**:263–265.

27. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M *et al.*: **PROSITE: a documented database using patterns and profiles as motif descriptors.** *Brief Bioinform* 2002, **3**:265-274.

28. Chou PY, Fasman GD: **Prediction of protein conformation.** *Biochemistry* 1974, **13**:222-245.

29. Chen H, Gu F, Huang Z: **Improved Chou-Fasman method for protein secondary structure prediction.** *BMC Bioinformatics* 2006, **7 Suppl 4**:S14.

30. Kyte J, Doolittle RF: **A simple method for displaying the hydropathic character of a protein.** *J Mol Biol* 1982, **157**:105-132.

31. Emini EA, Hughes JV, Perlow DS, Boger J: **Induction of hepatitis A virusneutralizing antibody by a virus-specific synthetic peptide.** *J Virol* 1985, **55**:836-839.

32. Fetrow JS: **Omega loops: nonregular secondary structures significant in protein function and stability.** *FASEB J* 1995, **9**:708-717.

33. Pal M, Dasgupta S: **The nature of the turn in omega loops of proteins.** *Proteins* 2003, **51**:591-606.

34. Jorieux S, Fressinaud E, Goudemand J, Gaucher C, Meyer D, Mazurier C: **Conformational changes in the D' domain of von Willebrand factor induced by CYS 25 and CYS 95 mutations lead to factor VIII binding defect and multimeric impairment.** *Blood* 2000, **95**:3139-3145.

35. Voorberg J, Fontijn R, van Mourik JA, Pannekoek H: **Domains involved in multimer assembly of von willebrand factor (vWF): multimerization is independent of dimerization.** *EMBO J* 1990, **9**:797-803.

36. Falconer H, D'Hooghe T, Fried G: **Endometriosis and genetic polymorphisms.** *Obstet Gynecol Surv* 2007, **62**:616-628.

37. Roya R, Baludu GS, Reddy BS: **Possible aggravating impact of gene polymorphism in women with endometriosis.** *Indian J Med Res* 2009, **129**:395-400.

38. Mitsuta K, Yokoyama A, Kondo K, Nakajima M, Arita K, Kohno N: **Polymorphism of the MUC1 mucin gene is associated with susceptibility to lung adenocarcinoma and poor prognosis.** *Oncol Rep* 2005, **14**:185-189.

39. Silva F, Carvalho F, Peixoto A, Teixeira A, Almeida R, Reis C *et al.*: **MUC1 polymorphism confers increased risk for intestinal metaplasia in a Colombian population with chronic gastritis.** *Eur J Hum Genet* 2003, **11**:380-384.

40. Silva F, Carvalho F, Peixoto A, Seixas M, Almeida R, Carneiro F *et al.*: **MUC1 gene polymorphism in the gastric carcinogenesis pathway.** *Eur J Hum Genet* 2001, **9**:548-552.

41. Engelmann K, Baldus SE, Hanisch FG: **Identification and topology of variant sequences within individual repeat domains of the human epithelial tumor mucin MUC1.** *J Biol Chem* 2001, **276**:27764-27769.

42. Shih IeM, Davidson B: **Pathogenesis of ovarian cancer: clues from selected overexpressed genes.** *Future Oncol* 2009, **5**:1641-1657.

43. Davidson B, Baekelandt M, Shih IeM: **MUC4 is upregulated in ovarian carcinoma effusions and differentiates carcinoma cells from mesothelial cells.** *Diagn Cytopathol* 2007, **35**:756-760.

44. Costa NR, Mendes N, Marcos NT, Reis CA, Caffrey T, Hollingsworth MA *et al.*: **Relevance of MUC1 mucin variable number of tandem repeats polymorphism in H pylori adhesion to gastric epithelial cells.** *World J Gastroenterol* 2008, **14**:1411-1414.

45. Santos-Silva F, Fonseca A, Caffrey T, Carvalho F, Mesquita P, Reis C *et al.*: **Thomsen-Friedenreich antigen expression in gastric carcinomas is associated with MUC1**

mucin VNTR polymorphism. *Glycobiology* 2005, **15**:511-517.

46. Fowler JC, Teixeira AS, Vinall LE, Swallow DM: **Hypervariability of the membrane-associated mucin and cancer marker MUC1.** *Hum Genet* 2003, **113**:473-479.

47. Choudhury A, Moniaux N, Winpenny JP, Hollingsworth MA, Aubert JP, Batra SK: **Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.** *J Biochem* 2000, **128**:233-243.

48. Koscinski I, Viville S, Porchet N, Bernigaud A, Escande F, Defossez A *et al.*: **MUC4 gene polymorphism and expression in women with implantation failure.** *Hum Reprod* 2006, **21**:2238-2245.

49. Alameda F, Mejias-Luque R, Garrido M, de Bolos C: **Mucin genes (MUC2, MUC4, MUC5AC, and MUC6) detection in normal and pathological endometrial tissues.** *Int J Gynecol Pathol* 2007, **26**:61-65.

50. Carraway KL, Perez A, Idris N, Jepson S, Arango M, Komatsu M *et al.*: **Muc4/sialomucin complex, the intramembrane ErbB2 ligand, in cancer and epithelia: to protect and to survive.** *Prog Nucleic Acid Res Mol Biol* 2002, **71**:149-185.

Figure legends

Figure 1 – Allelic discrimination plots of the six tested SNPs in *MUC4* gene

The DNA samples from patients and controls were genotyped by using the *Taqman* SNP genotyping assay system. The major (also the wild type) alleles were detected by FAM-labeled probes (blue color) and the minor alleles were detected by VIC-labeled probes. The genotyping results of the six SNPs in MUC4 gene were presented as allelic discrimination plots. Of note, the intensity of FAM signals tended to be similar among samples in our assays, thus the dots for a wild type genotype overlapped each other. "X" indicates the subject that failed to be genotyped.

Figure 2 – Functional domains in MUC4 protein sequence and the predicted secondary structures

Six functional domains/signatures (boxes) were annotated by aligning MUC4 protein sequence in PROSITE database. The boundaries of each signature were listed. Among six SNPs tested in this study (stars), rs882605 and rs2688513 (black) were found located in two different long-loop regions, 300-319 and 4134-4158, respectively (bold letters: amino acid substitution sites). The SNP rs2246901 (gray) was found located in a VWFD domain. The *MUC4* reference sequence in NCBI databank: NP_060876.4.

Table 1. Genotype and allele distributions of the six SNPs in *MUC4* gene in Taiwanese endometriosis patients and controls

	Genotype / allele	No. (%) of patients	HWE	No. (%) of control	HWE	p-value	OR (95% CI)
rs882605	TT	9 (7.0)	0.62	10 (7.2)	0.07	**0.04**	1.13 (0.43-3.00)
	TG	54 (42.2)		40 (28.8)			**1.97 (1.17-3.32)***
	GG	65 (50.8)		89 (64.0)			1.00 reference
	T	72 (28.1)		60 (21.6)		0.08	1.42 (0.96-2.11)
	G	184 (71.9)		218 (78.4)			1.00 reference
rs1104760	CC	7 (5.4)	0.94	9 (6.3)	0.09	0.30	0.96 (0.33-2.82)
	CT	47 (36.2)		39 (27.5)			1.49 (0.89-2.51)
	TT	76 (58.5)		94 (66.2)			1.00 reference
	C	61 (23.5)		57 (20.1)		0.34	1.22 (0.81-1.84)
	T	199 (76.5)		227 (79.9)			1.00 reference
rs2246901	GG	8 (6.0)	0.96	10 (7.1)	0.15	0.56	0.91 (0.33-2.53)
	TG	49 (36.6)		43 (30.5)			1.30 (0.78-2.17)
	TT	77 (57.5)		88 (62.4)			1.00 reference
	G	65 (24.3)		63 (22.3)		0.60	1.11 (0.75-1.65)
	T	203 (75.7)		219 (77.7)			1.00 reference
rs2258447	AA	9 (6.4)	0.37	10 (7.0)	0.07	0.70	1.02 (0.38-2.77)
	AG	44 (32.1)		40 (28.2)			1.25 (0.74-2.11)
	GG	81 (61.4)		92 (64.8)			1.00 reference
	A	62 (22.5)		60 (21.1)		0.57	1.12 (0.75-1.68)
	G	206 (77.5)		224 (78.9)			1.00 reference
rs2291652	CC	11 (8.7)	0.92	11 (8.4)	0.26	0.52	1.17 (0.46-2.96)
	CT	52 (40.9)		45 (34.4)			1.35 (0.81-2.28)
	TT	64 (50.4)		75 (57.3)			1.00 reference
	C	74 (29.1)		67 (25.6)		0.36	1.20 (0.81-1.76)

	T	180 (70.9)		195	(74.4)			1.00	reference
rs2688513	CC	8 (6.0)	0.62	10	(7.0)	0.09	0.66	0.91	(0.33-2.53)
	TC	45 (33.8)		41	(28.9)			1.25	(0.74-2.10)
	TT	80 (60.2)		91	(64.1)			1.00	reference
	C	61 (22.9)		61	(21.5)		0.68	1.09	(0.73-1.63)
	T	205 (77.1)		223	(78.5)			1.00	reference

Allelic frequencies were determined by chi-square test using 2×2 contingency tables.

Genotype frequencies were determined by chi-square test using 2×3 contingency tables.

*: *p* values less than 0.05 were considered statistically significant.

SNP: single-nucleotide polymorphism; OR: odds ratio; 95% CI: 95% confidence interval

HWE: *p* values of deviation from Hardy–Weinberg equilibrium

Table 2. Haplotype frequencies of *MUC4* polymorphisms in endometriosis patients and controls

rs1104760/ rs882605	*case (%)*	*control (%)*	*p-value*
TG	73.2%	78.8%	0.1202
CT	22.9%	20.1%	0.4134
TT	3.9%	1.1%	**0.0353***

The reported haplotype percents are estimated percents based on allele frequencies and the linkage disequilibrium.

The *p* values are based on a comparison of a given haplotype with all other haplotypes combined.

*: *p* values less than 0.05 were considered statistically significant.

Table 3. Genotype and allele distributions of SNPs in *MUC4* gene in endometriosis patients at different clinical stages

SNP	Genotype / allele	No. (%) at mild stage[a]	No. (%) at severe stage[b]	*p* value
rs882605	TT	0 (0.0)	7 (8.0)	0.25
	TG	2 (25.0)	41 (46.6)	
	GG	6 (75.0)	40 (45.5)	
	TT+TG	2 (25.0)	48 (54.5)	0.11
	GG	6 (75.0)	40 (45.5)	
	T	2 (12.5)	55 (31.3)	0.12
	G	14 (87.5)	121 (68.8)	
rs1104760	CC	0 (0.0)	5 (5.8)	0.05
	CT	0 (0.0)	36 (41.9)	
	TT	7 (100.0)	45 (52.3)	
	CC+CT	0 (0.0)	41 (47.7)	**0.01***
	TT	7 (100.0)	45 (52.3)	
	C	0 (0.0)	46 (26.7)	**0.03***
	T	14 (100.0)	126 (73.3)	
rs2688513	CC	0 (0.0)	6 (6.9)	**0.04***
	CT	0 (0.0)	34 (39.1)	
	TT	8 (100.0)	47 (54.0)	
	CC+CT	0 (0.0)	40 (46.0)	**0.01***
	TT	8 (100.0)	47 (54.0)	
	C	0 (0.0)	46 (26.4)	**0.02***
	T	16 (100.0)	128 (73.6)	
rs2246901	GG	0 (0.0)	6 (6.8)	**0.03***
	TG	0 (0.0)	36 (40.9)	
	TT	8 (100.0)	46 (52.3)	
	GG+TG	0 (0.0)	42 (47.7)	**0.01***
	TT	8 (100.0)	46 (52.3)	
	G	0 (0.0)	48 (27.3)	**0.02***
	T	16 (100.0)	128 (72.7)	
rs2258447	AA	0 (0.0)	7 (7.9)	0.08
	AG	0 (0.0)	32 (36.0)	
	GG	7 (100.0)	50 (56.2)	
	AA+AG	0 (0.0)	39 (43.8)	**0.02***
	GG	7 (100.0)	50 (56.2)	

	A	0	(0.0)	46	(25.8)	**0.03***
	G	14	(100.0)	132	(74.2)	
rs2291652	CC	0	(0.0)	7	(8.3)	0.61
	CT	2	(33.3)	36	(42.9)	
	TT	4	(66.7)	41	(48.8)	
	CC+CT	2	(33.3)	43	(51.2)	0.40
	TT	4	(66.7)	41	(48.8)	
	C	2	(16.7)	50	(29.8)	0.33
	T	10	(83.3)	118	(70.2)	

[a]mild stage: patients at clinical Stage1 or Stage 2

[b]severe stage: patients at clinical Stage3 or Stage 4

Allelic frequencies were determined by Fisher's exact test using 2×2 contingency tables.

Genotype frequencies were determined by Fisher's exact test using 2×3 contingency tables.

*: *p* values less than 0.05 were considered statistically significant.

SNP: single-nucleotide polymorphism; OR: odds ratio; 95% CI: 95% confidence interval

Table 4. Genotype and allele distributions of the six SNPs in *MUC4* gene in endometriosis patients with different reproductive ability

SNP	Genotype / allele	No. (%) infertility	No. (%) non-infertility	p-value	OR (95% CI)
rs882605	TT	0 (0.0)	8 (8.7)	0.07	- -
	TG	5 (26.3)	42 (45.7)		0.36 (0.12-1.08)
	GG	14 (73.7)	42 (45.7)		1.00 reference
	T	5 (13.2)	58 (31.5)	**0.03***	**0.33 (0.12-0.89)**
	G	33 (86.8)	126 (68.5)		1.00 reference
rs1104760	CC	0 (0.0)	6 (6.3)	0.41	- -
	CT	4 (25.0)	37 (38.5)		0.48 (0.14-1.60)
	TT	12 (75.0)	53 (55.2)		1.00 reference
	C	4 (12.5)	49 (25.5)	0.12	0.42 (0.14-1.25)
	T	28 (87.5)	143 (74.5)		1.00 reference
rs2688513	CC	0 (0.0)	7 (7.4)	0.32	- -
	TC	5 (25.0)	34 (36.2)		0.52 (0.17-1.56)
	TT	15 (75.0)	53 (56.4)		1.00 reference
	C	5 (12.5)	48 (25.5)	0.09	0.42 (0.15-1.12)
	T	35 (87.5)	140 (74.5)		1.00 reference
rs2246901	GG	0 (0.0)	7 (7.4)	0.25	- -
	TG	5 (25.0)	37 (38.9)		0.46 (0.15-1.38)
	TT	15 (75.0)	51 (53.7)		1.00 reference
	G	5 (12.5)	51 (26.8)	0.06	0.39 (0.14-1.05)
	T	35 (87.5)	139 (73.2)		1.00 reference
rs2258447	AA	1 (5.0)	7 (7.3)	0.41	0.53 (0.05-5.53)
	AG	4 (20.0)	33 (34.4)		0.45 (0.14-1.48)
	GG	15 (75.0)	56 (58.3)		1.00 reference
	A	6 (15.0)	47 (24.5)	0.22	0.54 (0.22-1.38)
	G	34 (85.0)	145 (75.5)		1.00 reference
rs2291652	CC	0 (0.0)	7 (7.8)	0.29	- -
	CT	7 (35.0)	40 (44.4)		0.58 (0.21-1.60)
	TT	13 (65.0)	43 (47.8)		1.00 reference
	C	7 (17.5)	54 (30.0)	0.12	0.49 (0.21-1.19)
	T	33 (82.5)	126 (70.0)		1.00 reference

Allelic frequencies were determined by Fisher's exact test using 2×2 contingency tables.

Genotype frequencies were determined by Fisher's exact test using 2×3 contingency tables.

*: *p* values less than 0.05 were considered statistically significant.

SNP: single-nucleotide polymorphism; OR: odds ratio; 95% CI: 95% confidence interval

Table 5. Haplotype frequencies of *MUC4* polymorphisms in endometriosis patients with infertility

rs882605/ rs2246901/ rs2688513	*infertility (%)*	*non-infertility (%)*	*P value*
GTT	87.2%	67.4%	**0.012***
TGC	12.5%	24.4%	0.099
TTT	0.2%	4.4%	0.206
GGT	0.0%	2.2%	0.344

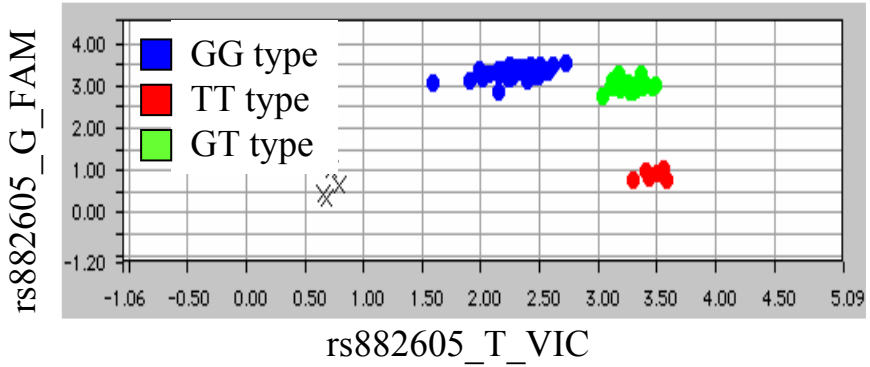
The reported haplotype percents are estimated percents based on allele frequencies and the linkage disequilibrium.

The *p* values are based on a comparison of a given haplotype with all other haplotypes combined.

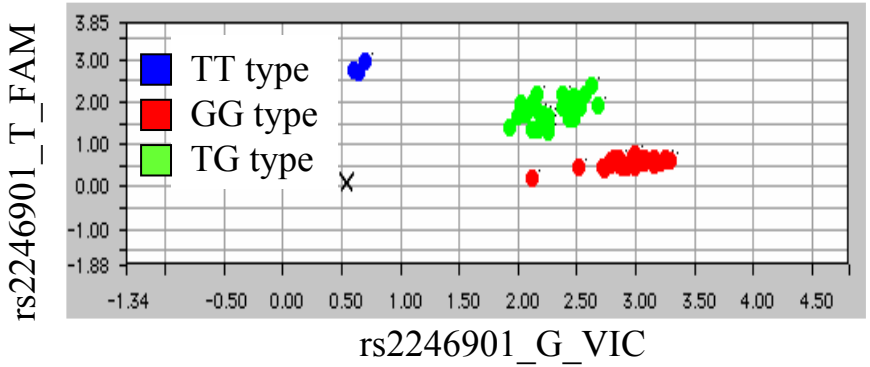
*: *p* values less than 0.05 were considered statistically significant.

Figure 1

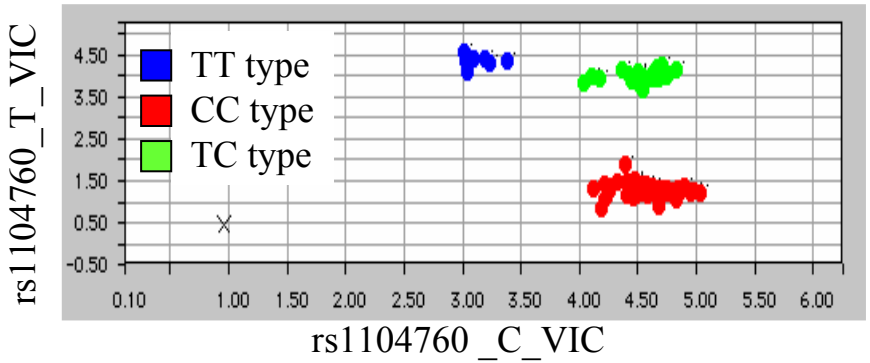
rs882605



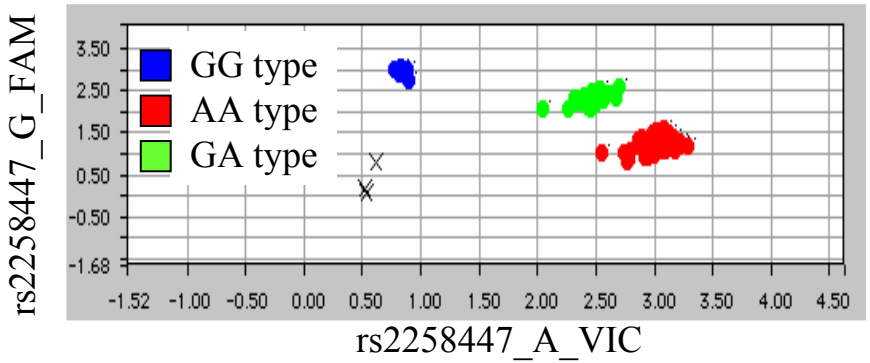
rs2246901



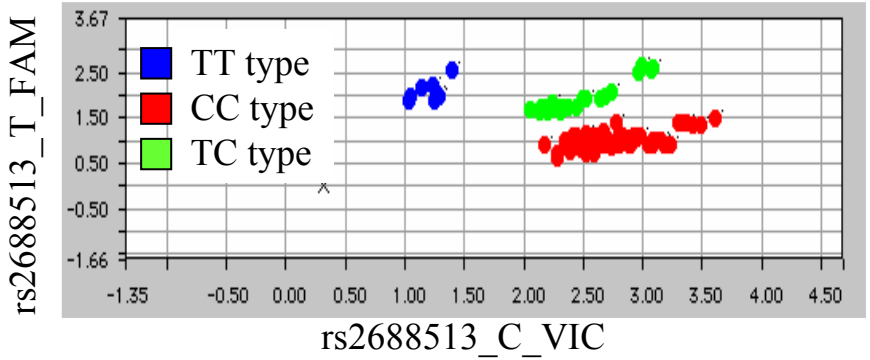
rs1104760



rs2258447



rs2688513



rs2291652

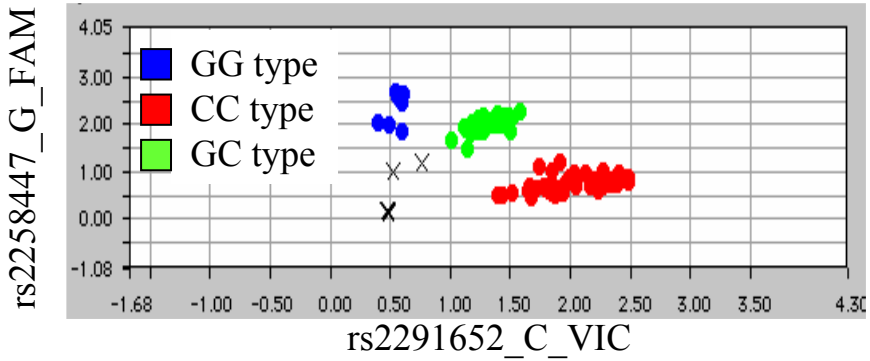
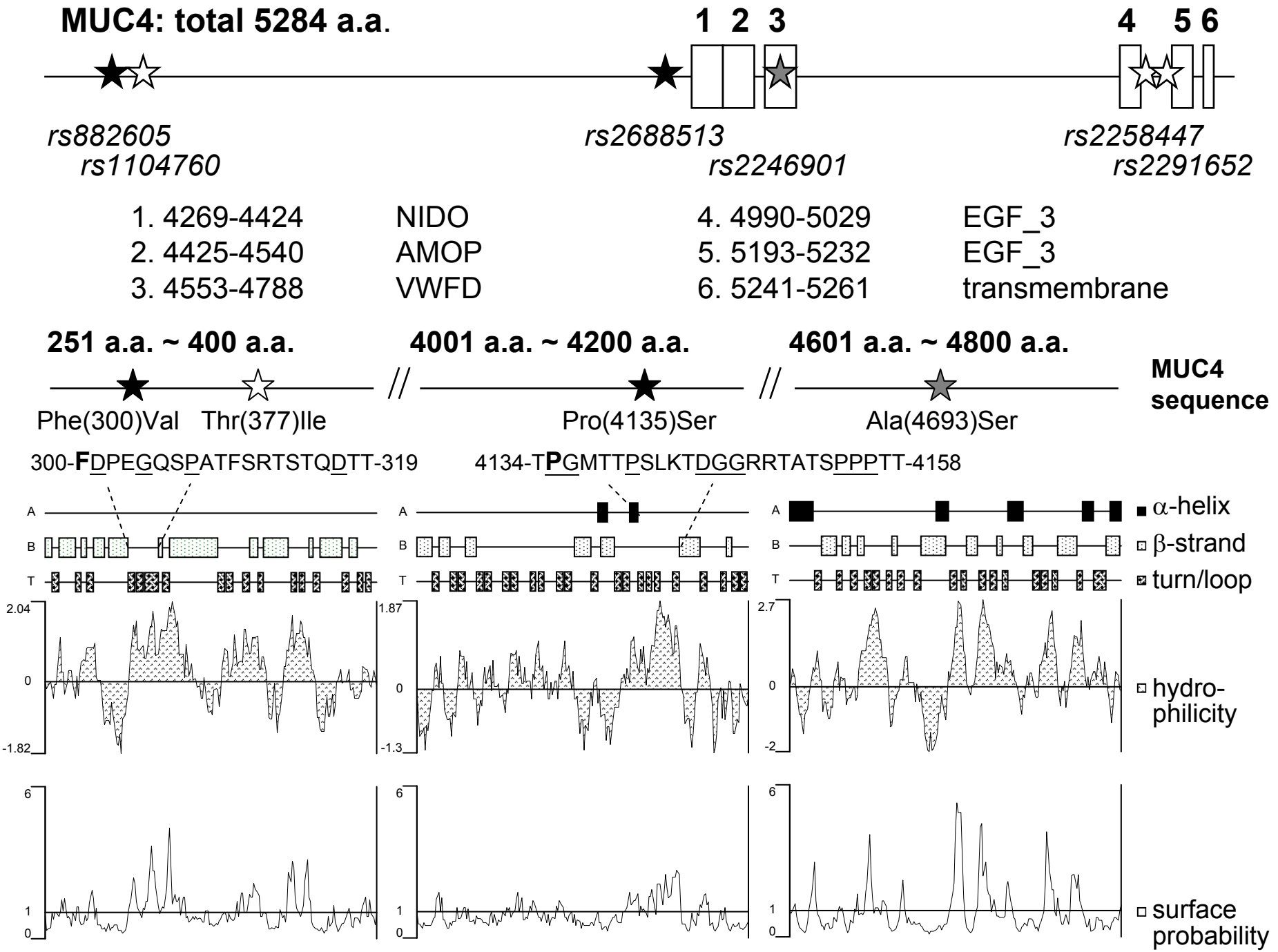


Figure 2


MUC4: total 5284 a.a.
1 2 3
4 5 6
rs882605
rs1104760
rs2688513
rs2246901
rs2258447
rs2291652
1. 4269-4424 NIDO
2. 4425-4540 AMOP
3. 4553-4788 VWFD
4. 4990-5029 EGF_3
5. 5193-5232 EGF_3
6. 5241-5261 transmembrane
251 a.a. ~ 400 a.a.
4001 a.a. ~ 4200 a.a.
4601 a.a. ~ 4800 a.a.
MUC4 sequence
Phe(300)Val
Thr(377)Ile
Pro(4135)Ser
Ala(4693)Ser
300-FDPEGQSPATFSRTSTQDTT-319
4134-TPGMTTPSLKTDGGRRTATSPPPTT-4158
A
B
T
α-helix
β-strand
turn/loop
2.04
0
-1.82
1.87
0
-1.3
2.7
0
-2
hydro-philicity
6
1
0
surface probability

Additional files provided with this submission:

Additional file 1: Supplementary Figures-revised.doc, 118K
http://www.biomedcentral.com/imedia/5960338014893193/supp1.doc
Additional file 2: Supplementary Tables-revised.doc, 62K
http://www.biomedcentral.com/imedia/2112439812489320/supp2.doc