**Summary answer:**

Higher TGFBI levels were detected in plasma samples of endometriosis patients compared to controls. Furthermore, TGFBI alone showed substantially better diagnostic performance than CA-125 in detecting minimal-to-mild endometriosis, with an AUC value of 0.74, sensitivity of 83%, and specificity of 61%. The combination of TGFBI and CA-125 best distinguished endometriosis patients from controls and reached an AUC value of 0.83, specificity of 90%, and sensitivity of 53%. These results suggest that TGFBI is a potential non-invasive biomarker of the early stages of endometriosis and that TGFBI together with CA-125 represent potential non-invasive biomarkers for all stages of endometriosis. To confirm the usefulness of these proposed biomarkers, further studies on larger cohorts are needed.

**What is known already:**

Endometriosis is a common chronic gynecological disease that significantly affects patient quality of life by causing pain and infertility. The gold standard for diagnosis is still visual inspection of pelvic organs via laparoscopy. Non-invasive biomarkers for endometriosis still need to be discovered to reduce diagnostic delays and enable earlier treatment. The potential non-invasive biomarkers for endometriosis evaluated in this study (COMP and TGFBI) were previously identified by proteomic analysis of peritoneal fluid samples.

**Study design, size, duration:**

This was a case–control study divided into a discovery (n=56 patients) and a validation phase (n=237 patients). All patients were treated between 2008 and 2019 at the Department of Obstetrics and Gynaecology, University Medical Centre Ljubljana (Ljubljana, Slovenia).

**Participants/materials, setting, methods:**

Patients were stratified based on the laparoscopic findings. The discovery phase included 32 endometriosis patients (cases) and 28 patients with confirmed absence of endometriosis (controls). The validation phase included 166 endometriosis and 71 control patients. In the discovery phase, COMP and TGFBI levels were measured in plasma samples, whereas CA-125 levels were measured in serum samples. ELISA was used for both discovery and validation experiments. Statistical and receiver operating characteristic (ROC) curve analyses were performed using Graph Pad Prism 9.3 and MetaboAnalyst 5.0 software, respectively. The classification models were built using the linear support vector machine (SVM) method with the SVM built-in feature ranking method.

**Main results and the role of chance:**

The discovery phase revealed significantly increased levels of TGFBI, but not COMP, in plasma samples of endometriosis patients compared to controls. In this smaller cohort, univariate ROC analysis showed fair diagnostic potential of TGFBI, with an AUC value of 0.77, specificity of 84%, and sensitivity of 58%. The classification model built using linear SVM and combining TGFBI and CA-125 showed an AUC value of 0.9, specificity of 89%, and sensitivity of 72% in distinguishing endometriosis patients from controls. The validation phase results confirmed the diagnostic potential of the model combining TGFBI and CA-125, with an AUC value of 0.83, specificity of 90%, and sensitivity of 53%. In addition, TGFBI exhibited good diagnostic potential for early-stage endometriosis (rASRM I-II), with an AUC value of 0.74, specificity of 83%, and sensitivity of 60%. TGFBI exhibited better diagnostic potential compared to CA-125 (which had an AUC value of 0.63, specificity of 67%, and sensitivity of 60%).

**Limitations, reasons for caution:**

The diagnostic models were built and validated from a single endometriosis center, and thus further validation and technical verification in a multicenter study with a larger cohort is needed.

**Wider implications of the findings:**

This study revealedfor the first time increased TGFBI levels in plasma samples of endometriosis patients, particularly those with minimal to mild endometriosis, compared to controls. This is the first step in considering TGFBI as a potential non-invasive biomarker for early-stage endometriosis. It also opens a path for new basic research to investigate the importance of TGFBI in the pathophysiology of endometriosis. Further studies are needed to confirm the diagnostic potential of a model based on TGFBI and CA-125 for the non-invasive diagnosis of endometriosis.

**Study funding/competing interest(s):**

The preparation of this manuscript was supported by grant J3-1755 from the Slovenian Research Agency to T.L.R and EU H2020-MSCA-RISE project TRENDO (grant 101008193). All authors declare that they have no conflicts of interest.

**Trial registration number:** NCT04591548.

**Key words:** endometriosis, biomarker, peritoneal endometriosis, early diagnosis, ELISA, validation

**Introduction**

Endometriosis is a common gynecological benign disease with a complex pathophysiology that is characterized by endometrial-like tissue outside the uterine cavity (Saunders and Horne, 2021). The disease significantly compromises the quality of life of women and is a major cause of infertility (Saunders and Horne, 2021, Zondervan et al., 2020). The gold standard for diagnosis is surgical visual inspection of pelvic organs. Currently, advanced minimally invasive laparoscopy is still an invasive surgical procedure with general anesthesia, endotracheal intubation, and potential perioperative and postoperative complications. Laparoscopy is especially needed to confirm superficial peritoneal endometriosis that cannot be diagnosed using imaging techniques (Nisenblat et al., 2016). Because of nonspecific symptoms and surgery as a standard diagnostic procedure, 6–7 years (on average) can pass before women are diagnosed and properly treated (Nnoaham et al., 2011). Therefore, biomarker research was defined as a research priority in 2011 by the World Endometriosis Research Foundation (Rogers et al., 2017, Rogers et al., 2013).

In the last decades, numerous molecules identified in biological fluids have been considered non-invasive biomarkers of endometriosis, including glycoproteins, cytokines, hormones, growth factors, and markers of oxidative stress, apoptosis, cell adhesion, and angiogenesis (Anastasiu et al., 2020, Hudson et al., 2020, Janša et al., 2021, Rižner, 2014, Tian et al., 2020). Recently, high-throughput technologies (e.g., proteomics, genomics, and metabolomics) have enabled the detection of proteins, genes, polymorphisms, miRNA molecules, metabolites, and lipids associated with endometriosis (Goulielmos et al., 2020). Among the proposed biomarkers, the most investigated is glycoprotein cancer antigen 125 (CA-125) (Anastasiu et al., 2020). Significantly higher serum CA-125 levels are commonly reported in patients with advanced stages of endometriosis. However, CA-125 measurements alone lack the specificity and sensitivity to detect endometriosis and replace current diagnostic techniques (Mol et al., 1998, Nisenblat et al., 2016). Nevertheless, several studies showed improved performance of CA-125 when combined with other blood biomarkers (Nisenblat et al., 2016).

Our research group has identified several biomarker candidates in peritoneal fluid and blood among single proteins and also by using metabolomic and proteomic approaches (Janša et al., 2021, Kocbek et al., 2015, Vouk et al., 2012, Vouk et al., 2016). We hypothesized that the intraperitoneal space with peritoneal fluid is a “natural habitat” of endometriosis and therefore that studies investigating peritoneal fluid might help identify blood biomarkers for the non-invasive diagnosis of endometriosis (Janša et al., 2021, Rižner, 2015). The surface of the peritoneal cavity is large and allows passive dialysis of substances between peritoneal fluid and blood plasma (Bedaiwy and Falcone, 2003, Koninckx et al., 1998, Rižner, 2015, Young et al., 2013). However, peritoneal fluid sampling is more invasive than peripheral blood sampling, and any attempts at identifying clinically useful biomarkers must aim towards being as non-invasive as possible.

We have recently published a prospective case–control study of peritoneal fluid analysis that included a discovery and a validation phase (Janša et al., 2021). In the discovery phase, we used a proteomic approach with high-content antibody protein microarrays targeting 1360 proteins with 1830 antibodies (Sciomics GmbH, Heidelberg, Germany). We included 12 women with primary infertility, who were divided into a group of six women with laparoscopically and histologically confirmed endometriosis and a control group of six women with unexplained primary infertility. Peritoneal fluid samples were collected during laparoscopy. Between the endometriosis and control group, we found differential abundances of 16 different proteins, all of which were > 1.5-fold upregulated in the endometriosis group. We selected angiotensinogen, transforming growth factor-β-induced protein ig-h3 (TGFBI), cartilage oligomeric matrix protein (or thrombospondin-5; COMP), and angiopoietin-4 for validation. To the best of our knowledge, these proteins have not been previously studied in peritoneal fluid or blood from endometriosis patients. Thus, we analyzed the levels of these proteins in a larger group of endometriosis (n=32) and control (n=24) patients using commercially available enzyme-linked immunosorbent assays (ELISA). We found significant differences in the levels of COMP and TGFBI and non-significant differences in angiotensinogen. A classification model based on a linear support vector machine (SVM) revealed very good diagnostic potential with an area under the curve (AUC) of > 0.83, sensitivity of 0.81, and specificity of 1.00.

The aim of the current study was to evaluate COMP and TGFBI alone and in combination with CA-125 as potential blood biomarkers of endometriosis. First, we measured the levels of COMP, TGFBI, and CA-125 in blood samples of the same cohort of patients as in our peritoneal fluid study. Next, we assessed the levels of significantly increased proteins (TGFBI and CA-125) in a larger independent validation cohort of patients and built classification models that included concentrations of both proteins.

**Materials and methods**

**Study design and patient selection**

The study was designed as a case–control study and was conducted with the approvals of The Medical Ethics Committee of the Republic of Slovenia (No. 0120-049/2016-4 (discovery phase); No. 0120-127/2016-2 and No. 0120-541/2019/7 (validation phase)). Informed consent was obtained from all participants included in the study.

The study was divided into discovery and validation phases (Figure 1). The discovery phase comprised the same patient cohort as in our previous study (Janša et al., 2021): patients with primary infertility (n=56), who had either endometriosis (cases; n=32) or unexplained primary infertility (controls; n=24). The validation phase included a new cohort of 237 patients, which were divided into endometriosis cases (n=166) or controls (n=71). All the patients underwent laparoscopy due to clinical indications (infertility and/or symptoms indicative of endometriosis), and the diagnosis was confirmed histologically. All the patients included in the discovery phase (Table 1) had a body mass index (BMI) in the normal range, a regular menstrual cycle (21–35 days), and normal results of partner semen analyses. Further inclusion criteria included no previous pelvic surgery, no known pelvic inflammatory disease, and no pathology (controls) other than endometriosis (cases) (as observed by ultrasound examination). The exclusion criteria included hormonal therapy in the last year, irregular menstrual cycles, autoimmune diseases, malignant or suspected malignant diseases, previous pelvic inflammatory disease, leiomyoma uteri, and polycystic ovaries. None of the patients had undergone previous pelvic surgery. None of the patients included in the validation phase (Table 2) had known pelvic inflammatory or malignant disease. In the validation phase cohort, more than 80% of patients had regular menstrual cycles (89%), normal BMI (83%), and had not undergone previous gynecological surgeries (87%).

# Sample and data collection

# All the patients who met the inclusion criteria were additionally evaluated. They filled out a questionnaire on their health history, stress levels, medication use, diet, lifestyle habits, and types of pain (dysmenorrhea, dyspareunia, or chronic pain) using a validated visual analogue scale (Wewers and Lowe, 1990). Stratification was carried out based on the laparoscopic and histological results: the case and control groups included patients with and without endometriosis, respectively.

# Blood samples were collected 1 day before surgery according to a strict standard operating procedure (Rizner and Adamski, 2019). Briefly, two tubes of blood (each containing 4 ml) were collected from each patient. To obtain plasma, blood was taken into BD Vacutainer tubes containing ethylenediaminetetraacetic acid (EDTA) (#368861, Becton Dickinson and Company, NJ, USA). To obtain serum, BD Vacutainer tubes with a gel separator and clot activator (#369032, Becton Dickinson and Company, NJ, USA) were used. Within 1 h after collection, the samples were centrifuged at 2500 x g (plasma) and 3000 x g (serum) for 10 min at 4 °C. Plasma and serum were aspirated, aliquoted (120 μL), and stored at −80 °C until analysis.

# ELISA and statistical analyses

# Commercially available ELISA kits were purchased for TGFBI (MyBioSource, San Diego, CA, USA; Catalogue No. #MBS177286; Lot No. #7481763813 (discovery phase), Lot No. #748171151118 (validation phase)) and COMP (Merck Millipore, Saint Louis, MO, USA; Catalogue No. #RAB1764-1KT; Lot No. #0524I2396). Serum samples were analyzed using an electrochemiluminescent immunoassay for CA-125 on an immunoassay analyzer (Cobas e411, Roche Diagnostics GmbH, Manheim, Germany).

# Significant differences in protein levels between groups were determined with the following steps. First, the ROUT method with Q set to 1% was used to detect outliers, which were not included in the analysis. The dataset without outliers was tested for normality using Shaphiro-Wilk tests. For normally and non-normally distributed data, the unpaired t-test and Mann Whitney test were used, respectively. Statistical analysis was performed using GraphPad Prism 9.3 (GraphPad Software, San Diego, CA, USA). The level of significance was set at p < 0.05. The receiver operating characteristic (ROC) curve analysis was performed using the Biomarker Analysis module in the MetaboAnalyst 5.0 software (Pang et al., 2021). In brief, the validation dataset had missing values for CA-125 (n=13, 5.48%), which were replaced by the column mean value of each group (mean substitution) (de Goeij et al., 2013). For individual proteins, univariate ROC curve analysis was performed, and AUC and 95% confidence intervals (CI) were calculated (using the 500 bootstrapping method). The optimal cut-off was set using the point closest to the left-top corner (Hoo et al., 2017). For a given cut-off, sensitivity and specificity were calculated. Protein combinations were analyzed using multivariate ROC curve exploratory analysis, and the ROC curves were generated using Monte–Carlo cross validation (Xu and Liang, 2001). The classification models were built using the linear SVM method with the SVM built-in feature ranking method (Chang and Lin, 2008, Li et al., 2002).