JHI JOURNAL OF HEALTH INFORMATICS

System for Storing, Recovering, and Comparing Gene Expression Information from Patients with Endometriosis

Sistema para Armazenamento, Recuperação e Comparação de Informações sobre Expressão Gênica em Pacientes com Endometriose

Sistema de Almacenamiento, Recuperación y Comparación de Informaciones sobre Expresión Génica en Pacientes con Endometriosis

Luiz Aparecido Virginio Junior[1], Juliana Meola[2], Daniel Blasioli Dentillo[3], Júlio César Rosa e Silva[4], Rui Alberto Ferriani[5], Silvana Giuliatti[6]

Keywords:
Endometriosis; Computational Biology; Gene Expression

ABSTRACT

Objective: To develop a web-based system to store, recover, compare, and associate information obtained using high-throughput molecular genetic techniques regarding differentially expressed genes in patients with endometriosis. **Methods:** A web-based tool was developed using the Java programming and XHTML markup languages. Professionals and researchers in the field contributed to the database by uploading research data from scientific articles in which screening techniques had been performed to determine differential gene expression in women with and without endometriosis. **Results:** a web-based system with the main functionalities of article registration and gene searching was developed. These functions allow the user to identify coincidences between the results reported in the registered articles regarding differential gene expression. **Conclusion:** The developed tool allows for the collection of principal data regarding dysregulated gene expression in endometriosis and highlights candidate genes for future studies to better understand the etiopathogenesis of this disease.

Descritores:
Endometriose; Biologia Computacional; Expressão gênica

RESUMO

Objetivo: O presente trabalho apresenta o primeiro sistema em ambiente web capaz de armazenar, recuperar, comparar e relacionar informações sobre expressão gênica em larga escala em pacientes com endometriose. **Métodos:** A ferramenta foi desenvolvida em ambiente *web* utilizando-se a linguagem de programação Java e a linguagem de marcação XHTML. O banco de dados foi alimentado por profissionais e pesquisadores da área, utilizando dados de pesquisas relatadas em artigos científicos, nas quais foram realizadas técnicas de *screening* para expressão gênica diferencial em pacientes com e sem endometriose. **Resultados:** Entre as principais funcionalidades do sistema destacam-se o cadastro de artigos e busca de genes. Tais funções permitem que o usuário identifique coincidências entre os resultados de expressão gênica registrados nos artigos cadastrados no sistema. **Conclusão:** A ferramenta desenvolvida permite reunir os principais dados referentes às alterações de expressão gênica na endometriose, contribuindo para uma melhor compreensão da origem da doença.

Descriptores:
Endometriosis; Biología Computacional; Expresión Génica

RESUMEN

Objetivo: En este trabajo se presenta el primer sistema capaz de almacenar, recuperar, comparar y relacionar en *web,* informaciones sobre expresión génica en gran escala en pacientes con endometriosis. **Métodos:** La herramienta se ha desarrollado en *web* utilizando el lenguaje de programación Java y el lenguaje XHTML. La base de datos fue generada por profesionales e investigadores, con datos de investigaciones publicadas en artículos científicos, en donde se efectuaron las técnicas para la detección de expresión diferencial de genes en pacientes con presencia y ausencia de endometriosis. **Resultados:** Entre las principales características del sistema se incluyen registros de artículos y búsqueda de genes. Estas funciones permiten que el usuario identifique similitudes entre los resultados de expresión de genes grabados en los artículos registrados en el sistema. **Conclusión:** Esta herramienta ha permitido reunir datos de variaciones de expresión génica en la endometriosis, lo que contribuye a una mejor comprensión del origen de esta enfermedad.

[1] *Masters Student in Electrical Engineering at the Department of Computer Engineering and Industrial Automation, Faculty of Electrical and Computer Engineering, University of Campinas (UNICAMP), Campinas (SP), Brazil.*
[2] *Researcher at the Department of Gynecology and Obstetrics of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto (SP), Brazil.*
[3] *Postdoctoral fellow at the Department of Gynecology and Obstetrics of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto (SP), Brazil.*
[4] *PhD and Professor at the Department of Gynecology and Obstetrics of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto (SP), Brazil.*
[5] *Full Professor at the Department of Gynecology and Obstetrics of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto (SP), Brazil.*
[6] *PhD and Professor at the Department of Genetics of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto (SP), Brazil.*

Autor Cooresponde nte: **Luiz Aparecido Virginio Junior**
e-mail: **luiz00jr@gmail.com**

Artigo recebido: **30/10/2014**
Aprovado: **25/03/2015**

INTRODUCTION

Endometriosis is a disease characterized by the presence of tissue that is morphologically and biologically similar to the endometrium in areas outside the uterine cavity (called ectopic tissue or ectopic/endometriotic lesion)[1]. The lesions are most commonly found in pelvic organs such as the ovaries and the peritoneum and less frequently in the abdominal wall and urinary and intestinal tracts[2].

Endometriosis treatment can be clinical (hormonal) and/or surgical. However, recurrence may occur in up to 60% of cases[1]. Endometriosis is currently considered a public health problem because of its impacts on physical and psychological health; its socioeconomic impact resulting diagnosis, treatment, and monitoring costs; and consequent absenteeism[3].

Despite considerable research, the etiopathogenesis of endometriosis remains unclear, thus resulting in great controversy because the disease cannot be explained by a single mechanism. In addition, many molecules and metabolic pathways are altered in patients with this disease, which suggests multiple causes[4].

In an attempt to clarify the mechanisms involved in the origin and development of endometriosis, many studies have used various high-throughput molecular genetics techniques such as subtractive hybridization microarray, and serial analysis of gene expression (SAGE) to analyze the gene expression profiles associated with the disease[5]. These studies have proposed comparisons of gene expression between:

a. eutopic endometrial tissues (eutopic) from women with and without endometriosis,

b. ectopic (endometriosis) and eutopic endometrial tissues from women without endometriosis, and

c. ectopic (endometriosis) and eutopic endometrial tissues from women with endometriosis; this comparison can be performed using samples from the same patient (paired samples) or from different patients (non-paired samples)[6].

In addition to using different methods and comparing different tissues, these studies have incorporated non-homogeneous samples for the menstrual cycle (proliferative or secretory) and/or disease stages (initial or advanced) at which the samples were obtained.

The results of these studies may diverge because of the variety of parameters under analysis. It is therefore necessary to identify methods that allow compilation and association of the relevant information, thus facilitating the collection of convergent data. These methods would improve the search for dysregulated genes associated with endometriosis and the understanding of the molecular mechanisms that initiate endometriosis and would therefore provide a basis for future research.

This study presents the development of a web-based tool that stores, recovers, compares, and associates data from the scientific literature that describes comparative studies of the differential gene expression profiles associated with endometriosis. The aim was to identify genes associated with the etiopathogenesis and pathways potentially involved in the origin of endometriosis in order to clarify the physiopathology of this very enigmatic disease.

METHODS

Obtaining data from articles

Professionals and researchers in the field fed the database with data from research reported in scientific articles that described the use of screening techniques to analyze differential gene expression in patients with and without endometriosis. Only published articles assigned a *digital object identifier (DOI)* were selected.

Relevant data for database storage

Data were stored according to the registered scientific articles and for each article, only data regarding the article identification, patient clinical data used in the study, type and characterization of the collected samples, type of methodology used, and differential gene expression analysis results were selected.

The article identification data included the title, publication date, authors, and DOI. The patient clinical data used included the following: a) the disease stage (initial or advanced), b) menstrual cycle phase (proliferative or secretory) when the samples were collected, and c) disease symptoms (pain and/or infertility).

The following aspects regarding the type and characterization of the collected samples were analyzed: a) comparisons of endometrial tissues (eutopic endometrial tissues in patients with endometriosis, eutopic endometrial tissues in patients without endometriosis, or ectopic endometrial tissues) and b) the locations of the endometriotic lesions (e.g., ovary, peritoneum). The following genetic screening techniques were used: subtractive library, microarray, and SAGE. The gene expression analysis results corresponded to the genes that were differentially expressed according to the results obtained from the selected articles.

Software development process

To develop the database, the MySQL data management system, version 5.5 (http://www.mysql.com/) was used. This free tool uses the Structured Query Language (SQL) as its interface. The following were used to implement the system: Java programming language (http://www.oracle.com/), Extensible Hypertext Markup Language (XHTML) (http://www.w3.org/), and JavaServer Faces (JSF) (https://javaserverfaces.java.net/) and PrimeFaces, version 3.4 (http://www.primefaces.org/) frameworks along with the GlassFish Server, version 3.1.2 (https://glassfish.java.net/) and Netbeans tool, version 7.3 (https://netbeans.org/).

To facilitate the persistent data management in Java, the Hibernate framework, version 3.0 (http://www.hibernate.org/) was used, which is capable of object-relational mapping; in other words, it allows the mapping of database attributes to a Java object model.

In addition, the Model–View–Controller (MVC) project standard was used, the aim of which is to isolate the logical components of the user interface application by dividing the application into 3 parts: model, view, and controller. The model is responsible for data management, whereas the view presents the data represented by the model in a user-appropriate format. The controller is responsible for

handling the user's requests, managing the model, and deciding which view should be presented[7].

The Spring Security framework, version 3.0.3 (http://www.springsource.org), which provides authentication mechanisms and access control, was used for system safety management.

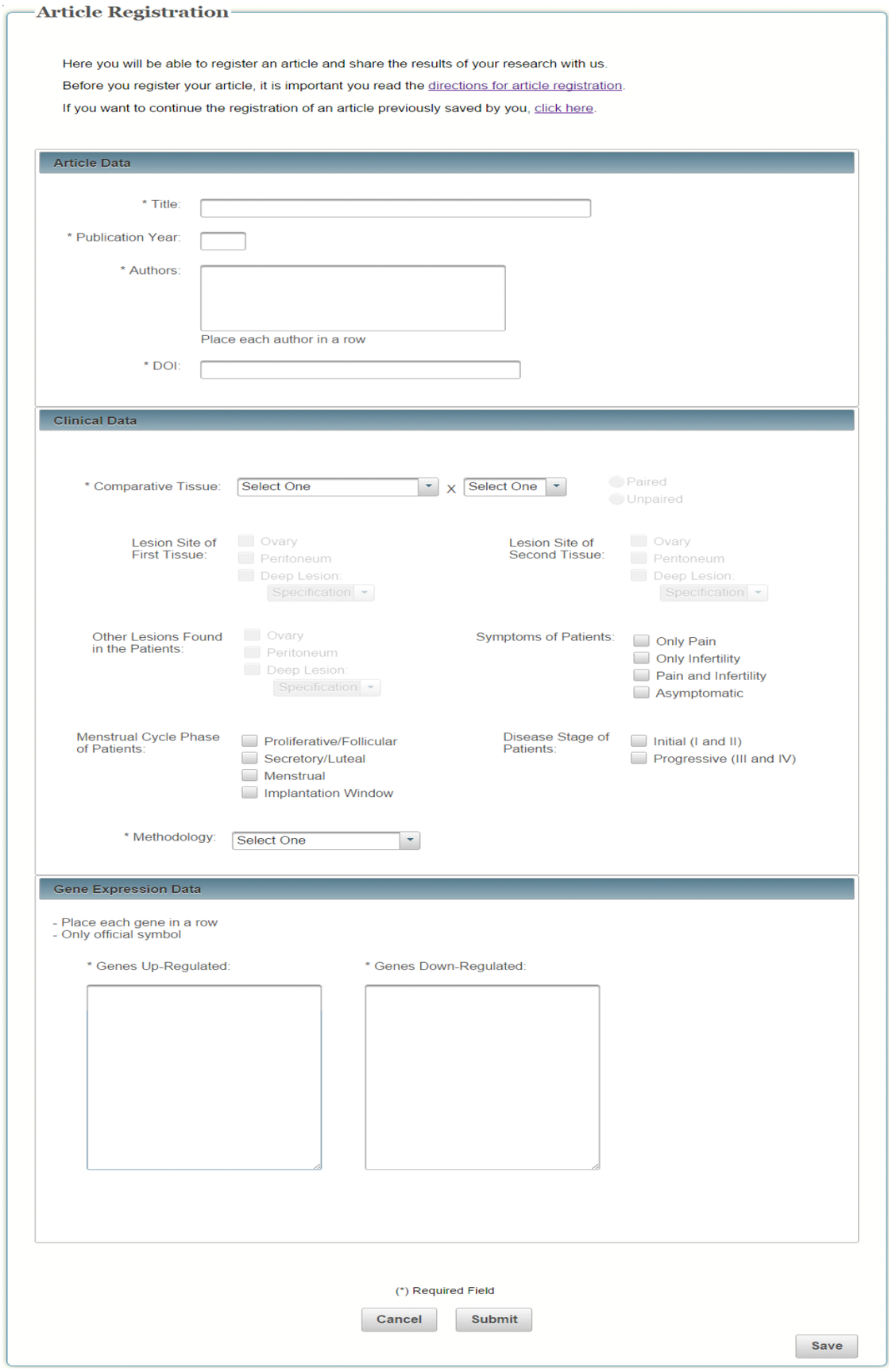

Figure 1 - Article registration page

RESULTS AND DISCUSSION

Article registration and gene searching are the main functionalities of this system. A user who is not a system member can access the article search ("Search for Articles") and gene search functions ("Search for Genes"); however, this user will not have permission to register articles. To access this function, the user will have to register with the system.

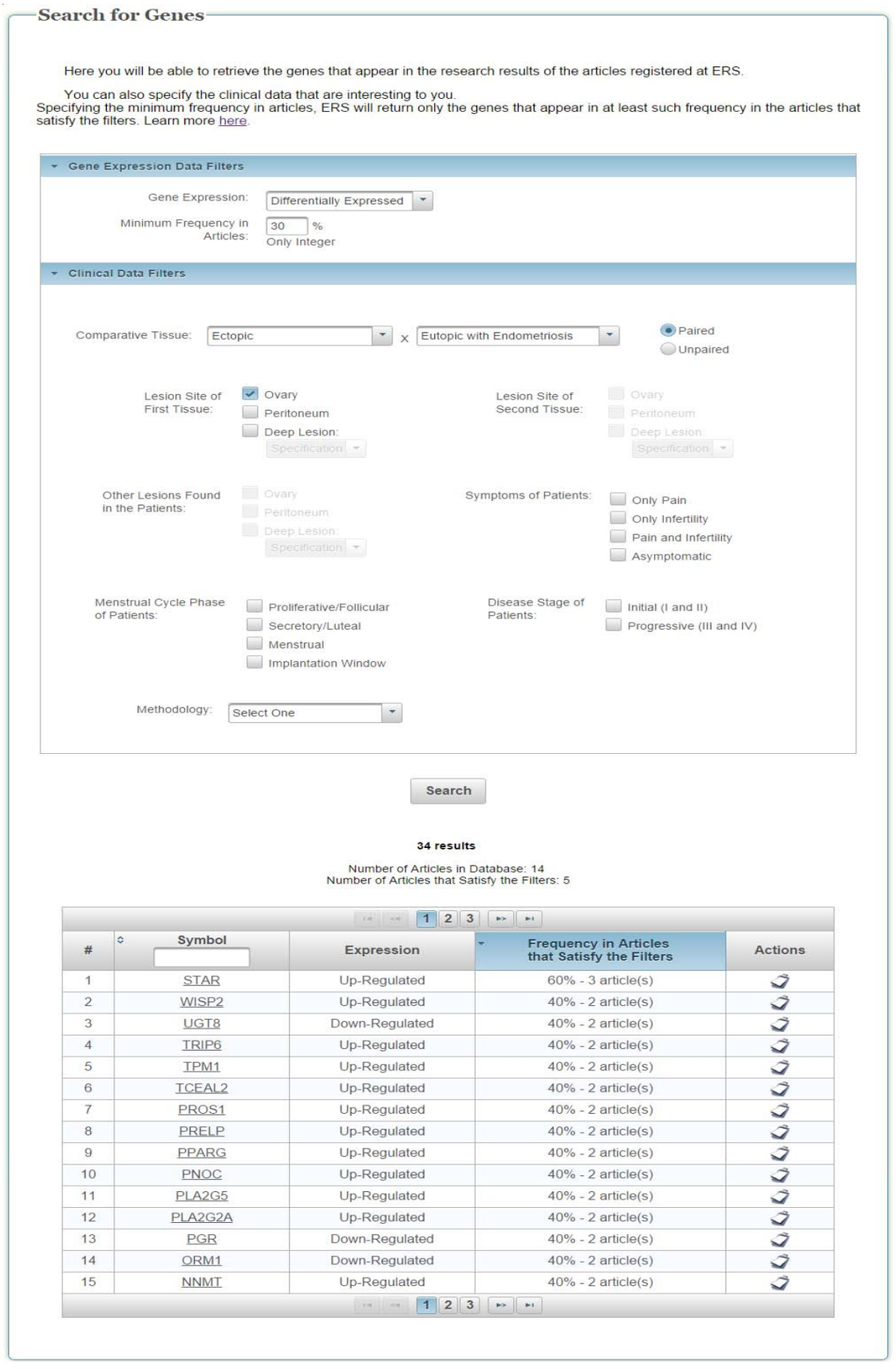

Figure 2 - Example of a gene search

The system offers complementary material along with instructions on how to register articles ("The System" panel—"Registering your Article"). In this way, the user can obtain information regarding the standards adopted in the system for registering compared tissues, endometriotic lesions, and differentially expressed genes.

To register an article, the user should log into the system, access the article registration page, and complete the form with the required information (article identification data, clinical data, and gene expression results). Figure 1 presents the form for article registration.

After an article is registered in the system, it is computed in the article and gene searches. Through a gene search, the user will be able to identify coincidences between the gene expression results reported in the registered articles. For this purpose, the user should access the gene search page and specify the search filters. The "Gene Expression" filter is used to specify whether only the results for the most expressed genes ("Up-Regulated") or the least expressed genes ("Down-Regulated") are returned. If the user selects the "Differentially Expressed" option, the system will return both the most and least expressed genes. The "Minimum Frequency in Articles" filter specifies the minimum percentage of articles in which the genes should appear as differentially expressed. For example, if the "Minimum Frequency in Articles" field is 50, only the genes that appear as differentially expressed in at least 50% of the articles will be returned. Moreover, the user can specify the clinical data in the articles of interest in which the involved genes should be searched.

For example, Figure 2 shows a search for differentially expressed genes ("Gene Expression" field = "Differentially Expressed") that appear in at least 30% of the articles ("Minimum Frequency in Articles" field = "30 %") reporting comparisons between paired ectopic and eutopic tissue samples from women with endometriosis ("Comparative Tissue" field = "Ectopic × Eutopic with Endometriosis – Paired"). In addition, the user selected the first tissue samples collected from ovarian lesions ("Lesion Site of First Tissue" field = "Ovary").

Figure 2 also shows the results of a gene search performed according to the specified filters. The results returned by the system show that, of the 14 articles registered in the system ("Number of Articles in Database: 14"), 5 met the clinical criteria specified by the search filter ("Number of Articles that Satisfy the Filters: 5"). The results table lists 34 genes, of which the STAR gene was the most frequently reported up-regulated gene (column "Expression" = "Up Regulated") in the articles, with a frequency of 60% (column "Frequency in Articles that Satisfy the Filters" = "60% - 3 articles"). Therefore, this specific search shows that the STAR gene appears as the most up-regulated gene in 3 of the 5 articles that meet the filter criteria.

CONCLUSION

The results from the article registration and search for endometriosis-associated genes achieved the proposed goals of this study. The system developed in this study, designated as the Endometriosis Research System (ERS), is available on the Internet at the following address: http://bioinfo1.fmrp.usp.br:8080/endomet.

Currently, this system is in local use by researchers involved in this project and is, to our knowledge, the first and only web-based system to provide these characteristics. However, we aim to divulge this system and facilitate the expansion and dissemination of its use, thus allowing the entire scientific community with an interest in endometriosis to access the information stored in our database and also register the results of their endometriosis gene expression analyses. Researchers will also be able to use and validate the results returned by the system in their laboratories, thus contributing to the advancement of endometriosis research.

It is worth noting that the gradual increase in the number of articles in our database will permit the identification of the most frequently reported differentially expressed genes in endometriosis-related articles, which is of importance because a high frequency would indicate the involvement of these genes in the origin and development of the disease.

ACKNOWLEDGEMENTS

We would like to thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for granting the scientific initiation scholarship that supported the development of this study.

REFERÊNCIAS

1. Arya P, Shaw R. Endometriosis: Current thinking. Curr Obstet Gynaecol. 2005;15(3):191-8.
2. Pritts EA, Taylor RN. An evidence-based evaluation of endometriosis-associated infertility. Endocrinol Metab Clin North Am. 2003;32(3):653-67.
3. Signorile PG, Baldi A. Endometriosis: new concepts in the pathogenesis. Int J Biochem Cell Biol. 2010;42(6):778-80.
4. Calhaz-Jorge C, Chaveiro E, Nunes J, Costa AP. Implications of the diagnosis of endometriosis on the success of infertility treatment. Clin Exp Obstet Gynecol. 2003;31(1):25-30.
5. Meola J, Rosa e Silva JC, Dentillo DB, da Silva WA Jr, Veiga-Castelli LC, Bernardes LA, et al. Differentially expressed genes in eutopic and ectopic endometrium of women with endometriosis. Fertil Steril. 2010;93(6):1750-73.
6. Dentillo DB, Meola J, Rosa e Silva JC, Giuliatti S, Silva Junior WA Jr, Ferriani RA, et al. Deregulation of LOXL1 and HTRA1 gene expression in endometriosis. Reprod Sci. 2010;17(11):1016-23.
7. Gonçalves E. Desenvolvendo aplicações web com jsp, servlets, javaserver faces, hibernate, ejb 3 persistence e ajax. Rio Janeiro: Ciência Moderna; 2007.