

Identification of Hub Genes and Pathways in Endometrial Cancer via Bioinformatics Approach

**Md. Foyzur Rahman^{1,*}, Md. Rezanur Rahman¹, Tania Islam²,
Toyfiqz Zaman¹, Md. Tofazzal Hossain³, Md. Rezaul Karim¹ and
Md. Ayub Ali⁴**

¹Department of Biochemistry and Biotechnology, School of Biomedical Science, Khwaja Yunus Ali University, Enayetpur, Sirajganj, Bangladesh

²Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh

³Department of Biochemistry and Molecular Biology, University of Rajshahi, Rajshahi, Bangladesh

⁴Department of Statistics, Faculty of Science, University of Rajshahi, Rajshahi, Bangladesh

*Correspondence should be addressed to Md. Foyzur Rahman
(Email: rahman.foyzur5239@gmail.com)

[Received December 10, 2020; Revised February 3, 2021; Accepted February 25, 2021]

Abstract

Endometrial cancer (EC) is one of the most common cancers of post-menopausal women. The key molecules and mechanisms that underlie to EC are still poorly known. Therefore, we aimed to identify potential biomarkers and networks to clarify the potential mechanisms behind EC. We employed an integrative bioinformatics analysis of three mRNA datasets (GSE17025, GSE115810, and GSE36389) of EC matched with normal tissue from public repository to identify differentially expressed genes (DEGs) in EC tumor. Then, we identified gene ontology and pathways to interpret their critical biological significance. A total of 63 aberrantly expressed genes (DEGs) involved in various intricate molecular pathways were identified. We detected eight key hub proteins (EZR, NR4A1, SFRP4, SOCS2, VAV3, WT1, TCF4, and MYO6) and transcription factors (FOXC1, GATA2, RUNX1T1, ZBTB20, E2F1, FOXL1, NFKB1, YY1, and HINFP) based on topological analysis of molecular networks. Besides, several drugs interacting with protein HTR2B were identified as candidate drugs influencing the key

pathways. In conclusion, we have identified a set of candidate molecular biomarkers and pathways in EC which may enhance the understanding of the underlying molecular mechanisms of EC that might help future studies.

Keywords: Endometrial Cancer, Transcription Factor, Molecular Pathways, Biomarkers, Candidate Drugs.

AMS Classification: 92B10.

1. Introduction

Endometrial cancer (EC) is a major female health issue and the fourth most common cancer occurring in post-menopausal women worldwide (Clarke et al., 2018). The American Cancer Society has estimated about 61,880 new cases and about 12,160 deaths for the cancer of the endometrium in the United States for the ongoing year (Siegel et al., 2019). EC, that arises from the lining of the uterus can be classified into two types: type- 1, estrogen-mediated, lower grade and accounts for 70-80% of the whole cases and type- 2, non-endometrioid, higher grade and mostly occur in older women (Clarke et al., 2018; Tran and Gehrig, 2017). Several factors such as hyperplasia, obesity, estrogen level, diabetes mellitus, life style etc. cause EC in women (Rosen et al., 2019; Ju et al., 2015; Han et al., 2015). In addition, patients having breast cancer can also develop EC because they share some common genes and pathways (Rahman et al., 2019). EC can be treated better if it is diagnosed at an early stage comparing to the advanced stage (Leslie et al., 2012). For the proper prognosis and treatment of diseases, bioinformatics analyses using various sophisticated tools playing vital roles in the recent years (Shen, 2013; Vogelstein et al., 2013). Thus, using microarray technology and bioinformatics analysis, researchers have revealed biological processes of EC (Colas et al., 2012); however, candidate biomarkers and molecular pathways were not confirmed strongly. Hence, in the present study, we tried to reveal the key hub proteins and pathways related to the pathogenesis of EC.

2. Materials and Methodology

2.1. Microarray data, data processing and DEGs screening

We mined array-based datasets for EC from the NCBI-GEO database (Barrett et al., 2012). We downloaded three Affymetrix based microarray datasets (GEO accession numbers GSE17025, GSE115810, GSE36389) for our study. The

GSE17025 dataset contained 103 samples, including 91 samples of pathologically reviewed stage I EC and 12 samples of atrophic endometrium as control. GSE115810 dataset contained 27 samples, including 3 control samples and 24 metastatic EC samples. For microarray analysis, the GSE36389 dataset used 20 samples which included 7 controls and 13 metastatic samples. After obtaining, log₂ transformation process was used to normalize all the datasets. Then, we screened the genes of each dataset using the Limma R package to identify differentially expressed genes (DEGs). A p-value of less than 0.01 and an absolute log₂ fold change (FC) > 1 were set as the threshold criteria throughout the whole process.

2.2. GO and KEGG pathway enrichment analysis of DEGs

We used DAVID 6.8 database (Sherman and Lempicki, 2009), a bioinformatics data resource consisting of an integrated biological knowledge base and analytic tools, to perform GO functional and KEGG pathway analysis of the integrated DEGs. Adjusted p-value < 0.05 was considered to indicate a statistically significant difference. The goal of this step was to explain gene functions and find relevant pathways.

2.3. PPI network construction

A protein-protein interaction network of the integrated DEGs of EC was constructed by STRING (Szklarczyk et al., 2016). Then, we used Network Analyst (Zhou et al., 2019) tool to visualize the PPI network. P-value < 0.05 was considered to indicate a statistically significant difference.

2.4. TF-gene interactions analysis

The TF-gene interactions network was created using Network Analyst (Zhou et al., 2019) to identify combinatorial and synergic regulatory relationships between TFs and the DEGs. The degree ≥ 20 was set as threshold value to obtain the TFs.

2.5. Protein-drug interactions analysis

We queried the Drug Bank database (Version 5.1.4) to obtain protein-drug interactions by using Network Analyst [(Zhou et al., 2019) to reveal candidate drugs of EC.

3. Results

3.1. Identification of DEGs in EC

We used Limma package of R software to identify the DEGs (3009 in GSE17025; 414 in GSE115810; and 212 in GSE36389) in the EC samples. Then we performed Venn diagram using jvenn tool (Bardou et al., 2014) to identify overlapping DEGs among the datasets. We found sixty-three genes common to all the datasets (Figure 1). Finally, we considered these 63 DEGs for further study.

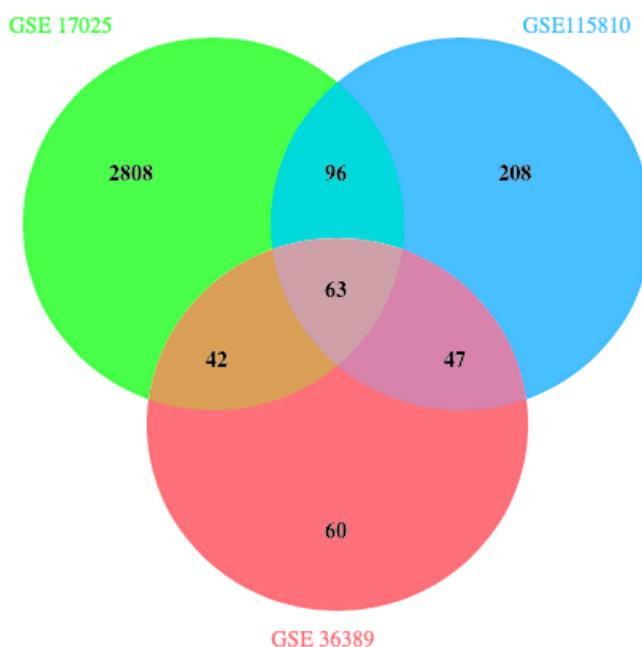


Figure 1: Venn diagram of DEGs. Sixty-three transcripts were found common among the three datasets.

3.2. GO and KEGG pathway analysis of DEGs

We queried DAVID to obtain biological process, molecular function, and cellular component enriched by the DEGs. We sorted out top five GO terms of biological process (BP), molecular function (MF), and cellular component (CC) considering their p-value (Table 1). The significant enriched BP were cell adhesion, negative regulation of transcription from RNA polymerase II promoter, positive regulation of transcription, DNA-templated, branching involved in ureteric bud

morphogenesis, and embryonic eye morphogenesis (Table 1). The suggestive CC were identified as extracellular space, proteinaceous extracellular matrix, extracellular exosome, extracellular region, and extracellular matrix (Table 1). The significantly enriched MFs by the DEGs were integrin binding, heparin binding, transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, extracellular matrix structural constituent, and growth factor activity (Table 1). We also studied pathways using the common DEGs to obtain the significant molecular pathways. The major pathways, that were involved in the pathogenesis of EC, were leukocyte trans-endothelial migration, ECM-receptor interaction, protein digestion and absorption, and focal adhesion (Table 2).

3.3. Identification of hub proteins

We used the common DEGs to construct protein-protein interaction (PPI) network to get insights into the functional interactions of the DEGs (Figure 2). We detected eight key hub proteins (*EZR*, *NR4A1*, *SFRP4*, *SOCS2*, *VAV3*, *WT1*, *TCF4*, and *MYO6*) from the constructed PPI network.

3.4. Identification of transcriptional factors

We integrated the DEGs with a human transcriptional regulatory network to identify significant transcription factors of EC. Thus, we obtained nine TFs (*FOXC1*, *GATA2*, *RUNX1T1*, *ZBTB20*, *E2F1*, *FOXL1*, *NFKB1*, *YY1*, *HINFP*) as the transcriptional regulatory components in EC datasets (Figure 3).

3.5. Identification of candidate drugs

To identify potential therapeutic agents for the treatment of EC, we performed Network Analyst which resulted in a protein-drug interaction network (Figure 4). The central protein *HTR2B* (5-Hydroxytryptamine receptor 2B) has several drug/compounds interactions. From the interaction network, we found some possible therapeutic agents namely amoxapine, asenapine, cariprazine, eletriptan, doxepin, pergolide, ketamine etc.

Table 1: GO terms (Top 5) in three categories enriched by the common DEGs in EC.

Categories	GO ID	GO Terms	P-value	Genes
Biological Process	GO:0007155	cell adhesion	3.35E-07	SEMA5A, F11R, SRPX, COL1A1, COL16A1, CXCL12
	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	2.54E-05	ZBTB20, FOXL2, EZR, OSR2, TLE4, PTCH1, FOSB, ZEB1
	GO:0045893	positive regulation of transcription, DNA-templated	3.27E-04	FOXL2, OSR2, GDF7, HOXA11, PTCH1, COL1A1
	GO:0001658	branching involved in ureteric bud morphogenesis	3.88E-04	HOXA11, PTCH1, DCHS1, WT1
	GO:0048048	embryonic eye morphogenesis	5.01E-04	FOXL2, EFEMP1, FBN1
Cellular Component	GO:0005615	extracellular space	7.69E-05	MUC1, GDF7, LTBP4, CXCL12
	GO:0005578	proteinaceous extracellular matrix	2.24E-04	OGN, LTBP4, EFEMP1, FBN1
	GO:0070062	extracellular exosome	3.43E-04	MUC1, F11R, MYO6, VAV3, PTPRF
	GO:0005576	extracellular region	0.001624	OGN, GDF7, EFEMP2, LTBP4, SFRP4

	GO:0031012	extracellular matrix	0.00267	OGN, LTBP4, EFEMP1, FBN1, COL1A1
Molecular Function	GO:0005178	integrin binding	2.98E-05	LTBP4, FBN1, COL16A1, ADAMTS5, COL5A1
	GO:0008201	heparin binding	2.20E-04	OGN, PTPRF, FBN1, PTCH1, ADAMTS5
	GO:0001077	transcriptional activator activity	0.00128	OGN, PTPRF, FBN1, PTCH1, ADAMTS5
	GO:0005201	extracellular matrix structural constituent	0.001577	EFEMP2, FBN1, COL1A1, COL5A1
	GO:0008083	growth factor activity	0.00231	OGN, GDF7, EFEMP1, CXCL12

Table 2: Significant KEGG pathways and relative genes.

KEGG ID	Pathways	P-value	Genes
hsa04670	Leukocyte trans-endothelial migration	0.0011	F11R, VAV3, EZR, CXCL12, THY1
hsa04512	ECM-receptor interaction	0.0483	COL1A1, COL5A1, SPP1
hsa04974	Protein digestion and absorption	0.0493	MME, COL1A1, COL5A1
hsa04510	Focal adhesion	0.0501	VAV3, COL1A1, COL5A1, SPP1

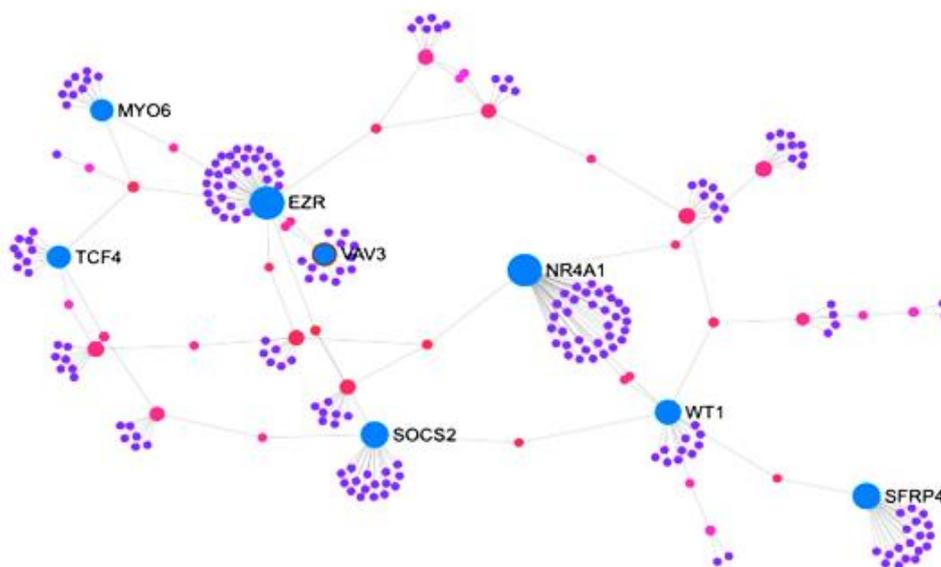


Figure 2: Protein-protein interaction of the common differentially expressed genes (DEGs) with other genes of EC. All nodes (circular points) indicate the DEGs, and the edges indicate the interactions between two genes. The eight large nodes indicate the hub proteins.

Table 3: List of hub proteins and their involvement in EC.

Hub proteins	Name	Roles of the biomolecules	References
EZR	ezrin/ cytovillin	involved in the process of invasion of endometrial cancer cells	Yan et al., 2018
NR4A1	nuclear receptor 4A1	exhibited pro-oncogenic activity in endometrial cells	Mohankumaret al., 2019
SFRP4	secreted frizzled-related protein 4	plays an important role in endometrial carcinogenesis	Pohl et al., 2015
SOCS2	suppressor of cytokine signaling 2	activates downstream signaling in EC	Bakkum-Gamezet al., 2015
VAV3	guanine nucleotide exchange factor VAV3	play an important role in the development and progression of EC	Boesch et al., 2018

WT1	Wilms' tumor gene	induction of angiogenesis in endometrial cancer	Dvorakova et al., 2013
TCF4	transcription factor 4	positive feedback loop of TCF4 plays an unfavorable role in the metastasis of endometrial carcinoma	Xiao et al., 2019
MYO6	Myosin VI	plays vital role in endometrial carcinogenesis	Romero-Pérez et al., 2013

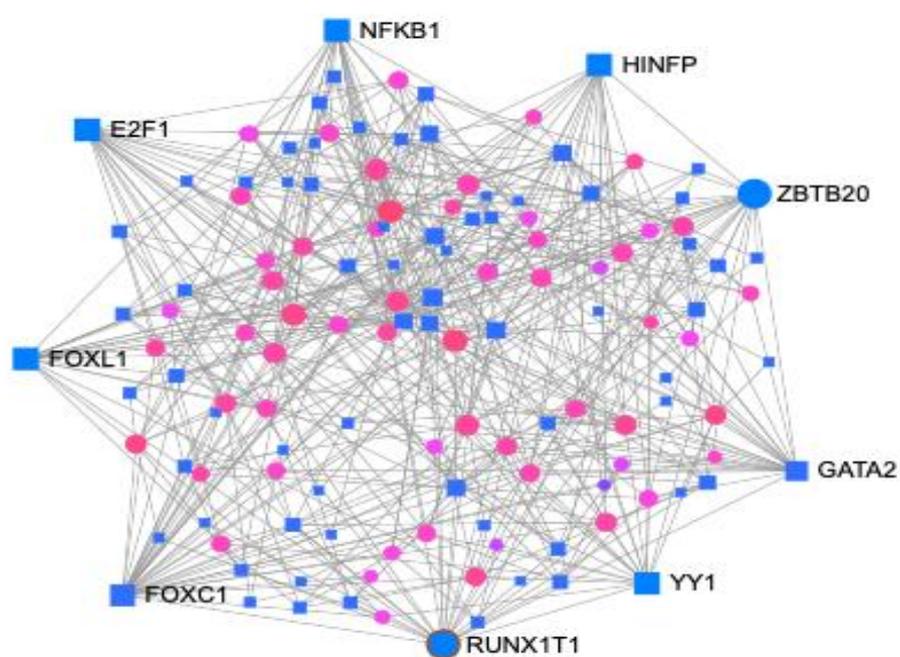


Figure 3: TF-gene interactions network. Nine transcription factors were identified as significant for EC.

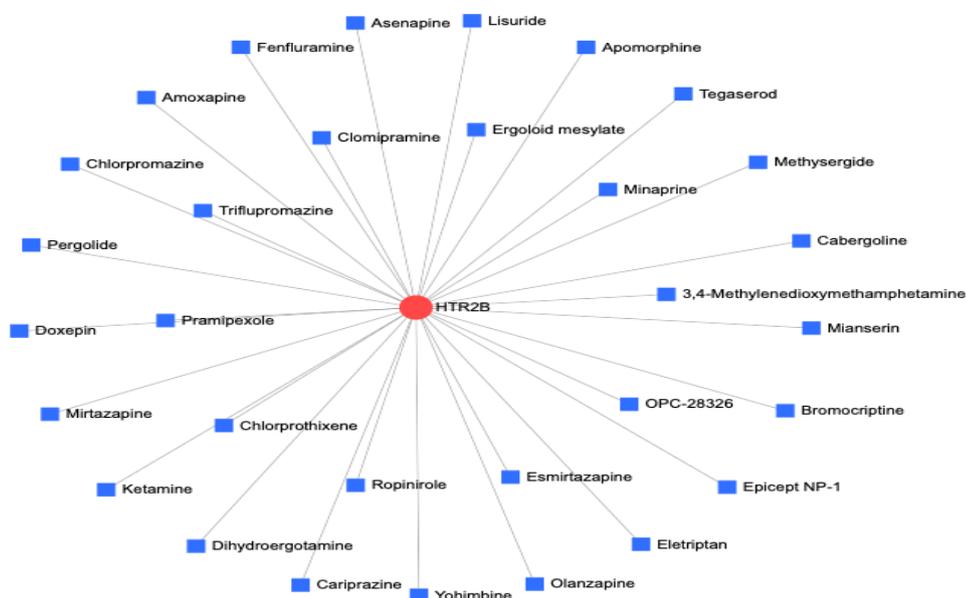


Figure 4: Protein-drug interactions network. Several candidate drugs were identified which could be used for EC treatment.

4. Discussion

Using microarray technology, now-a-days, several deadly diseases namely, ovarian cancer, cervical cancer etc. have been studying to explore the pathogenesis of the diseases for better treatment (Ye et al., 2015; Ai et al., 2013). Hence, in the present study, we also applied bioinformatics approaches to reveal molecular signatures and key pathways in EC, since the etiology of the disease remains poorly known, thus obstructing effective therapeutic opportunities for the disease. In our study, we downloaded three microarray datasets from the NCBI-GEO database and analyzed them to acquire DEGs between EC and normal tissues. We identified 63 common DEGs (Figure 1) which may be considered as EC biomarkers and possible therapeutic targets. DAVID tool analysis revealed significant GO terms (Table 1) and subsequent genes which were enriched in cellular, extracellular, growth and transcriptional activity. Several studies have reported that these biological processes, molecular functions, and cellular components affect EC tumorigenesis (Huang et al., 2013; Suman and Mishra, 2018). The KEGG pathway enrichment analysis revealed some altered signaling pathways which may be involved in the development and progression of EC. In a

study, Rahman, et al. (2019) reported protein digestion and absorption, and ECM-receptor interaction pathways enrichment by the colorectal cancer DEGs. To understand the underlying molecular mechanisms of EC, we constructed a PPI network (Figure 2) using the common DEGs. This resulted in eight key molecules i.e., hub proteins namely *EZR*, *NR4A1*, *SFRP4*, *SOCS2*, *VAV3*, *WT1*, *TCF4*, and *MYO6* which were then selected for study. Among the identified hub genes, *EZR* is involved in the process of invasion of endometrial cancer cells (Yan et al., 2018). The hub protein *NR4A1* exhibited pro-oncogenic activity in endometrial cells (Mohankumar et al., 2019). *SFRP4* and *MYO6* play vital role in endometrial carcinogenesis (Pohl et al., 2015; Romero-Pérez et al., 2013). Another hub protein *SOCS2* activates downstream signaling in EC (Bakkum-Gamez et al., 2015). *VAV3*, *WT1*, and *TCF4* play vital roles in the pathogenesis of EC, induce angiogenesis in EC and are involved in early signaling for *trans*-differentiation towards the molecular phenotype of EC cells, respectively (Boesch et al., 2018; Dvorakova et al., 2013; Xiao et al., 2019). We also studied the TFs-DEGs interaction and constructed a network using webtool from which we identified nine key transcriptional factors (*FOXC1*, *GATA2*, *RUNX1T1*, *ZBTB20*, *E2F1*, *FOXL1*, *NFKB1*, *YY1*, and *HINFP*) as regulator of the DEGs in EC. Furthermore, we studied the protein-drug interactions to identify therapeutic agents or compounds which can be used in the treatment of EC patients. A total 31 drugs, which were connected with a target protein- *HTR2B*, were revealed from the interaction network (Figure 4). The modes of bindings between the drugs and the protein were evaluated and an established association was confirmed by existing databases. In conclusion, our study revealed some signature key proteins which were involved in the signaling pathways at the molecular levels in EC progression and hence may be regarded as suggestive biomarkers and therapeutic targets of EC. The revealed biomarkers and drugs in our present study require more studies to define their clinical values.

5. Conclusion

In the present study, we have analyzed transcriptome profiles in EC using bioinformatics approaches to reveal the significant cellular pathways and candidate prognostic gene signatures, so that the underlying molecular mechanism behind onset of EC in women may be revealed. Molecular pathways that altered significantly include ECM-receptor interaction, protein digestion and absorption, focal adhesion and leukocyte trans- endothelial migration. We identified eight key

prognostic gene signatures: EZR, NR4A1, SFRP4, SOCS2, VAV3, WT1, TCF4, and MYO6 and nine transcription factors: FOXC1, GATA2, RUNX1T1, ZBTB20, E2F1, FOXL1, NFKB1, YY1, and HINFP based on topological analysis of molecular networks. In addition, we found some possible therapeutic agents namely amoxapine, asenapine, cariprazine, eletriptan, doxepin, pergolide, ketamine etc. from the interaction network with the central protein HTR2B. We have detected a potential prognostic hub genes and pathways for EC which may provide a new avenue for development and treatment of the disease.

References

- [1] Ai, Z., Wang, J., Xu, Y., and Teng, Y. (2013). Bioinformatics analysis reveals potential candidate drugs for cervical cancer. *Journal of Obstetrics and Gynaecology Research*, 39(5), 1052-1058.
- [2] Bakkum-Gamez, J. N., Wentzensen, N., Maurer, M. J., Hawthorne, K. M., Voss, J. S., Kroneman, T. N., ...and Cliby, W. A. (2015). Detection of endometrial cancer via molecular analysis of DNA collected with vaginal tampons. *Gynecologic oncology*, 137(1), 14-22.
- [3] Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC bioinformatics*, 15(1), 293.
- [4] Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., ...and Yefanov, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*, 41(1), 991-995.
- [5] Boesch, M., Sopper, S., Marth, C., Fiegl, H., Wiedemair, A., Rössler, J., ...and Zeimet, A. G. (2018). Evaluation of Vav3.1 as prognostic marker in endometrial cancer. *Journal of cancer research and clinical oncology*, 144(10), 2067-2076.
- [6] Clarke, M. A., Long, B. J., Morillo, A. D. M., Arbyn, M., Bakkum-Gamez, J. N., and Wentzensen, N. (2018). Association of endometrial cancer risk with postmenopausal bleeding in women: a systematic review and meta-analysis. *JAMA internal medicine*, 178(9), 1210-1222.
- [7] Colas, E., Pedrola, N., Devis, L., Ertekin, T., Campoy, I., Martínez, E., ...and Cabrera, S. (2012). The EMT signaling pathways in endometrial carcinoma. *Clinical and Translational Oncology*, 14(10), 715-720.
- [8] Dvorakova, E., Chmelarova, M., Laco, J., Palicka, V., and Spacek, J. (2013). Methylation analysis of tumor suppressor genes in endometroid carcinoma

- of endometrium using MS-MLPA. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 157(4), 298-303.
- [9] Han, J., Zhang, L., Guo, H., Wysham, W. Z., Roque, D. R., Willson, A. K., ... and Bae-Jump, V. L. (2015). Glucose promotes cell proliferation, glucose uptake and invasion in endometrial cancer cells via AMPK/mTOR/S6 and MAPK signaling. *Gynecologic oncology*, 138(3), 668-675.
- [10] Huang, V., Zheng, J., Qi, Z., Wang, J., Place, R. F., Yu, J., ... Li, L. C. (2013). Ago1 Interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. *PLoS genetics*, 9(9), e1003821.
- [11] Ju, W., Kim, H. J., Hankinson, S. E., De Vivo, I., and Cho, E. (2015). Prospective study of body fat distribution and the risk of endometrial cancer. *Cancer epidemiology*, 39(4), 567-570.
- [12] Leslie, K. K., Thiel, K. W., Goodheart, M. J., De Geest, K., Jia, Y., and Yang, S. (2012). Endometrial cancer. *Obstetrics and gynecology clinics of North America*, 39(2), 255–268.
- [13] Mohankumar, K., Li, X., Sridharan, S., Karki, K., and Safe, S. (2019). Nuclear receptor 4A1 (NR4A1) antagonists induce ROS-dependent inhibition of mTOR signaling in endometrial cancer. *Gynecologic oncology*.
- [14] Pohl, S., Scott, R., Arfuso, F., Perumal, V., and Dharmarajan, A. (2015). Secreted frizzled-related protein 4 and its implications in cancer and apoptosis. *Tumor Biology*, 36(1), 143-152.
- [15] Rahman, M. F., Rahman, M. R., Islam, T., Zaman, T., Shuvo, M. A. H., Hossain, M. T., ...and Moni, M. A. (2019). A bioinformatics approach to decode core genes and molecular pathways shared by breast cancer and endometrial cancer. *Informatics in Medicine Unlocked*, 17, 100274.
- [16] Rahman, M., Islam, T., Gov, E., Turanli, B., Gulfidan, G., Shahjaman, M., ...and Moni, M. A. (2019). Identification of prognostic biomarker signatures and candidate drugs in colorectal cancer: Insights from systems biology analysis. *Medicina*, 55(1), 20.
- [17] Romero-Pérez, L., Castilla, M. Á., López-García, M. Á., Díaz-Martín, J., Biscuola, M., Ramiro-Fuentes, S., ...and Moreno-Bueno, G. (2013). Molecular events in endometrial carcinosarcomas and the role of high mobility group AT-hook 2 in endometrial carcinogenesis. *Human pathology*, 44(2), 244-254.
- [18] Rosen, M. W., Tasset, J., Kobernik, E. K., Smith, Y. R., Johnston, C., and Quint, E. H. (2019). Risk Factors for Endometrial Cancer or Hyperplasia

- in Adolescents and Women 25 Years Old or Younger. *Journal of pediatric and adolescent gynecology*, 32(5), 546-549.
- [19] Shen, B. (Ed.). (2013). *Bioinformatics for diagnosis, prognosis and treatment of complex diseases* (Vol. 4). Springer Science & Business Media.
- [20] Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), 44-57.
- [21] Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1), 7-34.
- [22] Suman, S., and Mishra, A. (2018). Network analysis revealed aurora kinase dysregulation in five gynecological types of cancer. *Oncology letters*, 15(1), 1125-1132.
- [23] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ...and Jensen, L. J. (2016). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*, gkw937.
- [24] Tran, A. Q., and Gehrig, P. (2017). Recent advances in endometrial cancer. *F1000 Research*, 6, 81. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., and Kinzler, K. W. (2013). Cancer genome landscapes. *science*, 339(6127), 1546-1558.
- [25] Xiao, Y. Y., Lin, L., Li, Y. H., Jiang, H. P., Zhu, L. T., Deng, Y. R., ... and Chen, S. C. (2019). ZEB1 promotes invasion and metastasis of endometrial cancer by interacting with HDGF and inducing its transcription. *American journal of cancer research*, 9(11), 2314.
- [26] Yan, H., Sun, B. M., Zhang, Y. Y., Li, Y. J., Huang, C. X., Feng, F. Z., and Li, C. (2018). Upregulation of miR-183-5p is responsible for the promotion of apoptosis and inhibition of the epithelial-mesenchymal transition, proliferation, invasion and migration of human endometrial cancer cells by downregulating Ezrin. *International journal of molecular medicine*, 42(5), 2469-2480.
- [27] Ye, Q., Lei, L., and Aili, A. X. (2015). Identification of potential targets for ovarian cancer treatment by systematic bioinformatics analysis. *European journal of gynaecological oncology*, 36(3), 283-289.
- [28] Zhou, G., Soufan, O., Ewald, J., Hancock, R. E., Basu, N., and Xia, J. (2019). Network Analyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic acids research*, 47(1), 234-241.