**Insilico gene expression and protein structure a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics**

Surya Prabha.U, Sakila.L and Balaji.M

1Post Graduate and Research Department of Zoology, Ethiraj College for Women, Chennai, Tamilnadu, India.
2Associate Professor, Post Graduate and Research Department of Zoology, Ethiraj College for Women, Chennai, Tamilnadu, India.
3Senior Bioinformatician- Akshaya Computational Medicine Pvt, Ltd., Chennai, Tamilnadu, India.

**Abstract**

Ovarian cancer is a cancer that begins in an ovary. It results in abnormal cells that have the ability to invade or spread to other parts of the body. Polycystic ovary syndrome (PCOS) is a set of symptoms due to a hormone imbalance in women. PCOS is due to a combination of genetic and environmental factors. The aim of our present research investigation is to focus on gene expression and molecular structural relationships between Polycystic Ovarian Syndrome and Ovarian Cancer using advanced *Insilico* protocols. We identify the protein targets YAP1 (Yes Associated Protein 1) and FILIPIL (Filamin A Interacting Protein 1 Like) which are potential drug target for Polycystic Ovarian Syndrome and Ovarian Cancer. The three dimensional structure comparison studies clearly show that the two proteins target have similar sequence and structure. The identified protein target and *Insilico* gene expression studies would be useful in the field of clinical oncology.

Keywords: Gene Expression and Protein Comparison.

**INTRODUCTION**

Globally, as of 2010, about 160,000 people died from ovarian cancer, up from 113,000 in 1990 [1]. As of 2014, more than 220,000 diagnoses of epithelial ovarian cancer were made yearly [2]. In 2010, in the United States, an estimated 21,880 new cases were diagnosed and 13,850 women died of ovarian cancer. Around 1800 of the new diagnoses were sex-cord or stromal tumors [3]. In the United Kingdom as of 2014, approximately 7,000-7,100 yearly diagnoses were made and 4,200 deaths occurred [4]. It is the 5th most common cancer in UK women. Ovarian cancer is most commonly diagnosed after menopause, between the ages of 60 and 64. 90% of ovarian cancer occurs in women over the age of 45 and 80% in women over 50. The prevalence of PCOS depends on the choice of diagnostic criteria. The World Health Organization estimates that it affects 116 million women worldwide as of 2010 (3.4% of women)[5]. One community-based prevalence study using the Rotterdam criteria found that about 18% of women had PCOS, and that 70% of them were previously undiagnosed [6].

One study in the United Kingdom concluded that the risk of PCOS development was higher in lesbian women than in heterosexuals [7]. However, two subsequent studies of women with PCOS have not replicated this finding. Ultrasoundographic findings of polycystic ovaries are found in 8-25% of normal women. 14% women on oral contraceptives are found to have polycystic ovaries. Ovarian cysts are also a common side effect of intrauterine devices (IUDs)[8].

**AIMS AND OBJECTIVES**

- To find out the mutational genes involved in PCOs and Ovarian Cancer.
- To find out the highly expressed gene candidates using CAI values techniques.
- To perform molecular mechanics and visualization using advanced *Insilico* tools and database.

To whom correspondence should be addressed:
Surya Prabha U
Email: sprabha154@gmail.com
To compare the 3D protein structures of Ovarian Cancer and PCOS using protein structure comparison software.

**METHODS**

**STEP 1 - GENE SELECTION:** Literature studies were done on Ovarian Cancer and Polycystic Ovarian Syndrome using Pubmed, Pubmed Central and OMIM.

**STEP 2 - GENE EXPRESSION ANALYSIS:** The selected gene candidates are applied into Java Codon Adaptation Tool software in order to identify the highly expressed genes using Codon Adaptation Index methods.

**STEP 3 - SEQUENCE COMPARISON:** The highly expressed gene coded protein sequences are compared using T-COFFEE server in order to identify the sequence homology of Ovarian and Polycystic Ovarian Syndrome.

**STEP 4 - PROTEIN HOMOLOGY MODELING:** The highly expressed gene candidate coded protein sequences of Ovarian Cancer and Polycystic Ovarian Syndrome are modeled using PHYRE 2 model server.

**STEP 5 - PROTEIN 3D STRUCTURE COMPARISON:** The 3D structures of the proteins involved in Ovarian Cancer and Polycystic Ovarian Syndrome are compared using advanced 3D structure comparison server iPBA.

**GENE EXPRESSION ANALYSIS**

**Table 1. Codon adaptation index (cai) analysis**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Gene Name</th>
<th>Chromosome Location</th>
<th>Gene ID</th>
<th>OMIM ID</th>
<th>CAI Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>FSHR</td>
<td>2</td>
<td>2492</td>
<td>136435</td>
<td>0.948</td>
</tr>
<tr>
<td>2.</td>
<td>LHCGR</td>
<td>2</td>
<td>3973</td>
<td>152790</td>
<td>0.955</td>
</tr>
<tr>
<td>3.</td>
<td>INSR</td>
<td>19</td>
<td>3643</td>
<td>147670</td>
<td>0.953</td>
</tr>
<tr>
<td>4.</td>
<td>YAP1</td>
<td>11</td>
<td>10413</td>
<td>606608</td>
<td>0.959</td>
</tr>
<tr>
<td>5.</td>
<td>HMGA2</td>
<td>12</td>
<td>8091</td>
<td>600698</td>
<td>0.957</td>
</tr>
<tr>
<td>6.</td>
<td>AKT1</td>
<td>14</td>
<td>207</td>
<td>164730</td>
<td>0.956</td>
</tr>
<tr>
<td>7.</td>
<td>AR</td>
<td>X</td>
<td>367</td>
<td>313700</td>
<td>0.957</td>
</tr>
<tr>
<td>8.</td>
<td>ADIPOQ</td>
<td>3</td>
<td>9370</td>
<td>605441</td>
<td>0.956</td>
</tr>
<tr>
<td>9.</td>
<td>CRP</td>
<td>1</td>
<td>1401</td>
<td>123260</td>
<td>0.954</td>
</tr>
<tr>
<td>10.</td>
<td>SERPINE1</td>
<td>7</td>
<td>5054</td>
<td>173360</td>
<td>0.958</td>
</tr>
</tbody>
</table>

**Table 2. Codon adaptation index (Cai) analysis**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Gene Name</th>
<th>Chromosome Location</th>
<th>Gene ID</th>
<th>OMIM ID</th>
<th>CAI values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>BRAC1</td>
<td>17</td>
<td>672</td>
<td>113705</td>
<td>0.956</td>
</tr>
<tr>
<td>2.</td>
<td>BRAC2</td>
<td>13</td>
<td>675</td>
<td>600185</td>
<td>0.956</td>
</tr>
<tr>
<td>3.</td>
<td>BABAM1</td>
<td>19</td>
<td>29086</td>
<td>612766</td>
<td>0.957</td>
</tr>
<tr>
<td>4.</td>
<td>TIPARP</td>
<td>3</td>
<td>25976</td>
<td>612480</td>
<td>0.947</td>
</tr>
<tr>
<td>5.</td>
<td>MYC</td>
<td>8</td>
<td>4609</td>
<td>190080</td>
<td>0.956</td>
</tr>
<tr>
<td>6.</td>
<td>MUC16</td>
<td>19</td>
<td>94025</td>
<td>606154</td>
<td>0.957</td>
</tr>
<tr>
<td>7.</td>
<td>SKAP1</td>
<td>17</td>
<td>8631</td>
<td>604969</td>
<td>0.956</td>
</tr>
<tr>
<td>8.</td>
<td>GPR68</td>
<td>14</td>
<td>8111</td>
<td>601404</td>
<td>0.952</td>
</tr>
<tr>
<td>9.</td>
<td>FILIP1L</td>
<td>3</td>
<td>11259</td>
<td>612993</td>
<td>0.960</td>
</tr>
<tr>
<td>10.</td>
<td>DPH1</td>
<td>17</td>
<td>1801</td>
<td>603527</td>
<td>0.956</td>
</tr>
</tbody>
</table>

**Table 3. Codon adaptation index (cai) analysis**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Gene Name</th>
<th>Protein Name</th>
<th>Chromosome Location</th>
<th>CAI values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>YAP1_POLYCYSTIC_OVARIAN SYNDROME</td>
<td>Yes Associated Protein 1</td>
<td>11</td>
<td>0.959</td>
</tr>
<tr>
<td>2.</td>
<td>FILIP1L_OVARIAN CANCER</td>
<td>Filamin A Interacting Protein 1 Like</td>
<td>3</td>
<td>0.960</td>
</tr>
</tbody>
</table>
Surya Prabha U et al., Insilico gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics

SEQUENCE ALIGNMENT

GENE SEQUENCES: YAP1-POLYCYSTIC OVARIAN SYNDROME-HUMAN
MDPGQQPPQAPQGQQGPQPQPQGQGQGPQGPAQTAAQAPPAGQHQVHVRGSDSETDLLELFN
AVMNPKTAVPQTVKMLRKPDSFKPPEPKSHSRQASTDAAGATLTPGQVRPHSSPASLQLGAVSPG
TLPTGVSGPAATPTAQLHRQSSFEPDVPPLAGWEMAKTSQSGRYFNLHIDQTWWQDPRKAML5QNM
NTAPTSPVQQLMNMSASAMNQRSISOAPKVQPPPLAQPQGQGVGMSNQQQQMQRLQQLMKEDELRLKQ
QELLRLQELQLRQLPTLEQDGTVNPVSSPGMSEQELQTTMNSSDPFNLNSGTYHSRDESTDGSLSM
SSYVPRPTPDDFNVDMDTGTINQSTLPSQNNFPDYLEAIGTNVDLGTLGDEMGNINEELMPSL
QEAQSDDINMSVLAATKLDKESFLTWL

GENE SEQUENCES: FILIP1L-OVARIAN CANCER-HUMAN
MRSRGSDTEGAQKKFPRHTGHSFQGPKNKHRQQDKDSPQSDLCPKAEKPHSNGHQAEQDL6RD
DLLFLSIEGELQARDEVIGLKEKMDLALLEAQGFTTPKLEALQRDAFQAKSTPQEDIYEKPM
NELDIKVEKHEKESYRILGQLLVAEKSRRQTILEELEERKKEKEYMEKSDFICLLQECERWSLALLPR
LECNGMILAHCNLLGCSSDSPASAFQVAGITGTRHQAQLVFL

SEQUENCE COMPARISON USING T-COFFEE

Fig 1. T-Coffee server sequence comparison

The above picture represents the comparison of gene sequences in YAPI and FILIP1L
(*) indicates conserved regions, (.) indicates point mutation and (:) indicates single amino acid changes in FILIP1L and YAPI protein sequences.

PROTEIN HOMOLOGY MODELLING

Fig 2. 3D structure of yes associated protein 1-polycystic ovarian syndrome

The above picture represents the Protein Secondary Structure View and Solid ribbon model; Red Colour indicates helix, Green Colour indicates Turn and white Colour indicates Coils regions.
Surya Prabha U et al., *Insilico* gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics

**Fig 3.** 3D Structure of yes associated protein 1 - Polycystic Ovarian Syndrome

![Image](image1.png)

*The above picture represents the basic amino acids (space fill model - yellow)*

**Fig 4.** 3D structure of yes associated protein 1 - Polycystic Ovarian Syndrome

![Image](image2.png)

*The above picture represents the acidic amino acids (space fill model - yellow)*

**Fig 5.** 3D structure of yes associated protein 1 - Polycystic Ovarian Syndrome

![Image](image3.png)

*The above picture represents the hydrophilic amino acids (space fill model - yellow)*
Surya Prabha U et al., *Insilico* gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics

**Fig 6. 3D structure of yes associated protein 1 - Polycystic Ovarian Syndrome**

*The above picture represents the hydrophobic amino acids (space fill model - yellow)*

**Fig 7. 3D Structure of yes associated protein 1 - Polycystic Ovarian Syndrome**

*The above picture represents the side chain amino acids (space fill model - yellow)*

**Fig 8. 3D Structure of yes associated protein 1 - Polycystic Ovarian Syndrome**

*The above picture represents the backbone amino acids (space fill model - yellow)*
**Surya Prabha U et al., Insilico gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics**

**Fig 9. 3D Structure of filamin a interacting protein 1 like-ovarian cancer**

![3D Structure of filamin a interacting protein 1 like-ovarian cancer](image)

Protein Secondary Structure View and Solid ribbon model;
Red Colour indicates helix, Green Colour indicates Turn and white Colour indicates Coils regions.

**Fig 10. 3D Structure of filamin a interacting protein 1 like-ovarian cancer**

The above picture represents the basic amino acids (space fill model- yellow)

**Fig 11. 3D Structure of filamin a interacting protein 1 like-ovarian cancer**

The above picture represents the acidic amino acids (space fill model- yellow)
Surya Prabha U et al., *Insilico* gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics

**Fig 12. 3D structure of filamin a interacting protein 1 like -ovarian cancer**

The above picture represents the hydrophilic amino acids (space fill model- yellow)

**Fig 13. 3D Structure of filamin a interacting protein 1 like -ovarian cancer**

The above picture represents the hydrophobic amino acids (space fill model- yellow)

**Fig 14. 3D Structure of filamin a interacting protein 1 like -ovarian cancer**

The above picture represents the side chain amino acids (space fill model- yellow)
**Surya Prabha U et al., Insilico gene expression and protein structure – a comparative study on polycystic ovarian syndrome and ovarian cancer in human beings using bioinformatics**

**Fig 15. 3D structure of filamin a interacting protein 1 like -ovarian cancer**

The above picture represents the backbone amino acids (space fill model- yellow)

**PROTEIN STRUCTURE COMPARISON**

**Fig 16. Protein structure comparison via ipba**

The above picture represents the Protein structure – Cartoon model view
Green: Yap1-Polycystic Ovarian Syndrome -Human
Red: Filip1l-Ovarian Cancer-Human

**Fig 17. Protein structure comparison via ipba**

The above picture represents the Protein Structure – Trace Model View
GREEN: YAP1-Polycystic Ovarian Syndrome-Human
RED: FILIP1L-Ovarian Cancer-Human
DISCUSSION
The genes involved in Ovarian Cancer and Polycystic Ovarian Syndrome are selected and the complete Insilico gene expression studies were done. The results of gene expression study shows that the highly expressed gene candidates are Yes Associated Protein 1 (YAP1) and Filamin A Interacting Protein 1 Like (FILIP1L).

T-Coffee (Tree-based Consistency Objective Function For alignment Evaluation) is a multiple sequence alignment software using a progressive approach. It generates a library of pairwise alignments to guide the multiple sequence alignment. It can also combine multiple sequences alignments obtained previously and in the latest versions can use structural information from PDB files (3D-Coffee). It has advanced features to evaluate the quality of the alignments and some capacity for identifying occurrence of motifs (Mocca). It produces alignment in the aln format (Clustal) by default, but can also produce PIR, MSF, and FASTA format. The most common input formats are supported (FASTA, PIR).

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns.

Hence according to Notredame C et al., 2000 we have efficiently used the T-COFFEE server. The protein sequences of the identified gene were retrieved from human ensemble genome browser database in order to predict the protein modeling and comparison studies. Before protein modeling, we align the protein sequences using advanced T-COFFEE server. Fig 1 shows that the two cancers share a sequence homology of (55%).

Apart from the WW domain, the modular structure of YAP1 contains a proline-rich region at the very amino terminus, which is followed by a TID (TEAD transcription factor interacting domain). Next, following a single WW domain, which is present in the YAP1-1 isoform, and two WW domains, which are present in the YAP1-2 isoform, there is the SH3-BM (Src Homology 3 binding motif). Following the SH3-BM is a TAD (transcription activation domain) and a PDZ domain-binding motif (PDZ-BM).

YAP1 is a transcriptional co-activator and its proliferative and oncogenic activity is driven by its association with the TEAD family of transcription factors, which up-regulate genes that promote cell growth and inhibit apoptosis. Several other functional partners of YAP1 were identified, including RUNX, SMADs, p73, ErbB4, TP53BP, LAT51/2, PTPN14, AMOTs, and ZO1/2. YAP1 and its close paralog, TAZ (WWTR1), are the main effectors of the Hippo tumor suppressor pathway. When the pathway is activated, YAP1 and TAZ are phosphorylated on a serine residue and sequestered in the cytoplasm by 14-3-3 proteins. When the Hippo pathway is not activated, YAP1/TAZ enter the nucleus and regulate gene expression.

Ovarian cancer is the most lethal gynecologic malignancy with a five-year survival rate below 25% for patients with stages III and IV disease. Identifying key mediators of ovarian cancer invasion and metastasis is critical to the development of more effective therapeutic interventions. We previously identified Filamin A interacting protein 1-like (FILIP1L) as an important mediator of cell proliferation and migration. In addition, targeted expression of FILIP1L in tumors inhibited tumor growth in vivo. Responsive element binding protein (CREB) was shown to bind to the CREB/ATF site in the CpG island of the FILIP1L promoter. The identified gene candidates correlate with the results of wet lab studies.

iPBA is a tool for comparison of protein structures based on similarity in the local backbone conformation. The local backbone conformation is defined as pentapeptide dihedrals, using Protein Blocks (PBs). The protein structures represented as PB sequences, are aligned by dynamic programming scored by a PB substitution matrix. Structurally similar stretches are weighed using an anchor based alignment approach. Hence we have chosen to use protein structure comparison studies of highly expressed genes Table 3.
Fig 2 and 9 represents the secondary structural properties and antigen binding sites of both YAP1 and FILIPIL proteins. Finally, the modeled protein structures are compared using advanced 3 Dimensional structures comparison server (Ipba) in order to find out the molecular structural similarities. Fig 16 explains how the YAP1 and FILIPIL structures bind mutually. These results are basic foundations for structure based drug designing for cancer research. Further research can be concentrated in this field to enhance the quality life in patients those who are suffering from Polycystic Ovarian Syndrome and Ovarian Cancer. This also gives the way in understanding the gene mutation and comparison of highly expressed genes in both Polycystic Ovarian Syndrome and Ovarian Cancer where drug docking can be done, which helps the future generation.

CONCLUSION

In the present research investigation we focused on gene expression and molecular structural relationships between Polycystic Ovarian Syndrome and Ovarian Cancer using advanced Insilico protocols. We conclude that the identified protein targets YAP1 (Yes Associated Protein 1) and FILIPIL (Filamin A Interacting Protein 1 Like) are potential drug target for breast cancer, Polycystic Ovarian Syndrome and Ovarian Cancer. The results of the three dimensional structure comparison studies clearly show that the two proteins target have similar sequence and structure. From the Insilico results, we conclude that the same drug can be used for Polycystic Ovarian Syndrome and Ovarian Cancer.

REFERENCES