

# Bioinformatics Tools: Alternative Approach for Poly Cystic Ovary Syndrome (PCOS) Detection

Riya<sup>1</sup>, Parul Thapar<sup>2\*</sup><sup>1</sup>Student, B.Sc. (hons.) Biotechnology, School of Pharmaceutical and Health care Sciences, CT University, Ludhiana, India<sup>2</sup>Assistant Professor, Biotechnology, School of Pharmaceutical and Health care Sciences, CT University, Ludhiana, IndiaDOI: [10.36348/sjls.2020.v05i09.006](https://doi.org/10.36348/sjls.2020.v05i09.006)

| Received: 13.09.2020 | Accepted: 21.09.2020 | Published: 30.09.2020

\*Corresponding author: Dr. Parul Thapar

## Abstract

Poly cystic ovary syndrome (PCOS) is a complex endocrine disorder prevailing among women of reproductive age, specifically high in teenage girls and has become the most accepted cause of menstrual irregularities and infertility among them. In a study by Parthap *et al.*, 2018, at Department of Gynaecology, Vinayaka Missions Kirupanada Variyer Medical College, Salem, India, it was found that, there were 100 patients who had confirmed diagnosis of PCOS. It has also been observed that medication create side effects in young women. Therefore, it was thought to explore an alternative approach for disease detection, prevention or treatment for PCOS with the help of bioinformatics. Bioinformatics is an expanding field of science involving biology, computer science and mathematics. It is growing in every field of life science including molecular sciences, biotechnology, medicine, agriculture and more. Genetic information stored in the bioinformatics tools can be used to develop personal medicine. In another study, it is said that although the genetics and mechanism of PCOS are not yet understood, the computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease. In the present study, the gene and genome sequences responsible for causing PCOS have been identified using bioinformatics tools like BLAST, PDB, NCBI. This will help to prevent the disease by genetic manipulations. Finally, the primers have been designed using primer designing tools in NCBI which can be further used for the treatment of the disease by manipulating the identified gene through polymerase chain reaction.

**Keywords:** Bioinformatics tools, BLAST, Poly cystic ovary syndrome (PCOS), Protein Data Bank (PDB), CLUSTAL W.

**Copyright @ 2020:** This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

## INTRODUCTION

Poly cystic ovary syndrome (PCOS) is a complex endocrine disorder prevailing among women of reproductive age, specifically high in teenage girls and has become the most prominent cause of menstrual irregularities and infertility among them. This disorder is associated with modernisation of living in middle-income and high-income population. Young adolescent girls experience full range of symptoms from irregular menses, amenorrhea, menorrhagia, hirsutism, acne, skin pigmentation, alopecia and ovarian cysts. Other symptoms like anxiety, depression and thyroid problems may exist. Obesity or propensity to weight gain is a common feature, though it is not uncommon in non-obese women. There are no specific diagnostic tests designed for PCOS, but certain other tests like pelvic exam, blood tests, ultrasound, checking for the signs of excess hair growth, insulin resistance and acne can be performed. The medications recommended by

the doctors to regulate menstrual cycles are the combination of birth control pills and progestin therapy. Examples are Clomiphene, Letrozole, Metformin and Gonadotropins. To reduce excessive hair growth, the medicines used are Oral Control Pills (OCPs), Spironolactone, Metformin (Mt) and Eflornithine. But these medications create certain side effects in young women [1-4]. Metformin is linked with fatal and nonfatal lactic acidosis and OCPs have certain side effects like weight gain, cardiovascular and thromboembolic events [5]. Therefore, it was thought to explore an alternative approach of disease detection, prevention or treatment for PCOS with the help of bioinformatics.

Bioinformatics is an expanding field of science involving biology, computer science and mathematics. It is growing in every field of life science including molecular sciences, biotechnology, medicine,

agriculture and other areas. Various tools and databases used in the field of bioinformatics are National Centre for Biotechnology Information (NCBI), Gen bank, European Nucleotide Archive, Ensemble, Basic Logic Alignment Search Tool (BLAST) which mainly contain information of genome of different organisms present in the universe. Personal medicine is the customised way of healthcare for treating patients using genetic information. These bioinformatics tools and databases are used to demonstrate genetic information [6].

In a study, the pathogenesis by the candidate genes causing PCOS was conducted using bioinformatics analysis. In another study, it is said that although the genetics and mechanism of PCOS are not yet understood, the computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease [1-4].

In the present study, genome sequences and the gene responsible for causing PCOS have been identified so that this particular disease can be prevented by genetic manipulation and there cannot be requirement of taking medicines to treat the disease. The objectives of the study are: 1. Searching for genes causing PCOS. 2. Retrieving the sequence for the identified genes. 3. Designing the primers for detecting the disease.

## MATERIALS AND METHODS

### Searching of Genes Causing PCOS:

From the human genome, two main genes were searched that code for the proteins within the ovary. The source used was, "The human protein atlas" by the method described in human proteome [7].

### Retrieving Sequences for the Genes

The two different genes were searched and sequences were retrieved for both the genes using bioinformatics NCBI database. The NCBI database includes Basic Logic Alignment Search Tool (BLAST) which is available for retrieving the sequences.

- For the first gene, the home page of NCBI was opened, in "all databases" tab; chosen "nucleotide", and put the searched gene in the search bar and then enter. A new web page appeared, chosen the desired gene from all (*Homo sapiens*) and then i.e. "Homo sapiens kelch domain containing 8A, transcript variant 2, mRNA" sequence appeared. By clicking on it, the graphics and FASTA sequence for desired gene was obtained.
- For the second gene, the home page of NCBI was opened, in all databases, chosen "gene", and put the searched gene in the search bar and then enter. A new web page appeared, chosen the desired sequence from "all" and the information of that particular gene sequence appeared. At the end of the webpage, from the related sequences of the genome, the desired nucleotide (*homo sapiens*) was selected. Then, the complete sequence of the

human DNA sequence from clone RP5-87J13 on chromosome Xq22.1-22.3 appeared. By clicking on it, the graphics and FASTA sequence for desired gene was obtained.

For both the retrieved sequences, the number of the sequences was obtained using "highlight sequence" feature and the number of base pairs was calculated by using CLUSTAL W tool.

### Designing of Primers for Detecting the Disease

The designed primers for the sequence may help in the determination of the disease and it can be treated by manipulating the desired identified gene.

- From the first identified sequence, after the FASTA sequence was retrieved, again visited the NCBI database to design the primers. Then, clicked on "pick primers" in the NCBI software for primer designing. The length of the forward and reverse primer was filled as 100 to 400 and 500 to 800 respectively and then clicked on "get primers". Therefore, the primers were designed for the identified gene.
- From the second identified sequence, after the FASTA sequence was retrieved, again visited the NCBI database to design the primers. Then, clicked on "pick primers" in the NCBI software for primer designing. The length of the forward and reverse primers were filled as 100 to 400 and 500 to 900 respectively and then clicked on "get primers". Therefore, the primers were designed for the second identified gene.

## RESULTS AND DISCUSSION

### Searching of Genes Causing PCOS

The two different genes were searched from the human protein atlas. The first gene identified was KLHDC8A. This gene encodes a kelch domain sequence which can be the cause of cancer. Sometimes the gene encoding kelch domain sequence is regulated due to the formation of tumours which activates the tumour into Onco cells and causes cancer. Since this gene is also enriched in ovary tissue therefore it is the most prominent cause of PCOS [5]. The second gene identified was MUM1L1 i.e. melanoma associated antigen (mutated) 1-like 1. This gene encodes a protein which contains a mutated melanoma-associated antigen-1 domain on its membrane. The proteins which contain mutated antigens are expressed at high levels in certain types of cancers. The tissues enhanced are ovary and thyroid gland. It is not detected into immune cells and predicted location is intracellular. It targets the melanoma cells if present in the ovaries [5].

In a study conducted by Panda *et al*, 2016 similar results have been isolated from Uni Prot and PDB tools for follicle stimulating hormone receptor with accession number: AAB26480, Uni Prot id: P23945 and PDB id: 1XUN. Also genes like

Melanocortin 4 receptor with accession number: AAO92061, Uni Prot id: P32245, PDB id: 2IQP have been identified with the same bioinformatics tools [6].

**Retrieving Sequences for the Identified Genes**

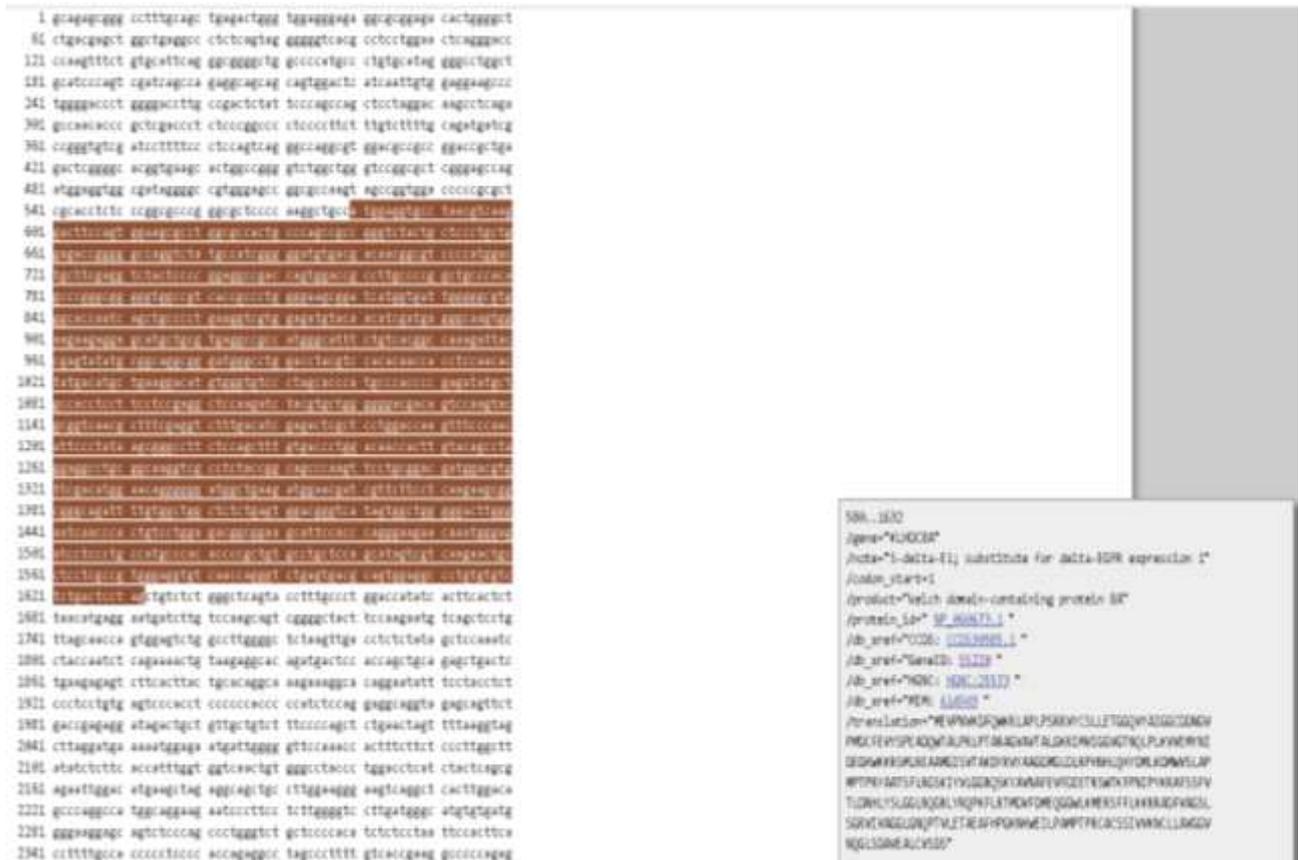
The sequences were retrieved through NCBI database using bioinformatics tools like BLAST for both the genes KLHDC8A and MUM1L1 respectively. The number of sequences retrieved for the gene KLHDC8A was 1,632 by using NCBI software using “highlight sequence” feature and the number of base pairs obtained using CLUSTAL W tool was 2,970. The obtained data is represented in the Table-1. The number of sequences retrieved are shown in the Figure-1 and

the base pairs in the Figure-2(a). The number of sequences retrieved for the gene MUM1L1 was 61,921 by using NCBI software using “highlight sequence” feature and the number of base pairs was 61,957 by using CLUSTAL W tool. The obtained data is represented in the Table-1. The number of sequences retrieved are shown in the Figure 3 and the base pairs in the Figure-2(b).

Similar results have been found in a study by Panda *et al*, 2016. The sequence for the follicle stimulating hormone receptor gene encoding the PCOS has been identified in the accession number: AAB26480 [6].

**Table-1: Features of the Retrieved Sequences**

S. No	Genes Identified	No. of sequences	No. of base pairs
1.	KLHDC8A	1,632	2,970
2.	MUM1L1	61,921	61,957



**Fig-1: Number of sequences for gene KLHDC8A**

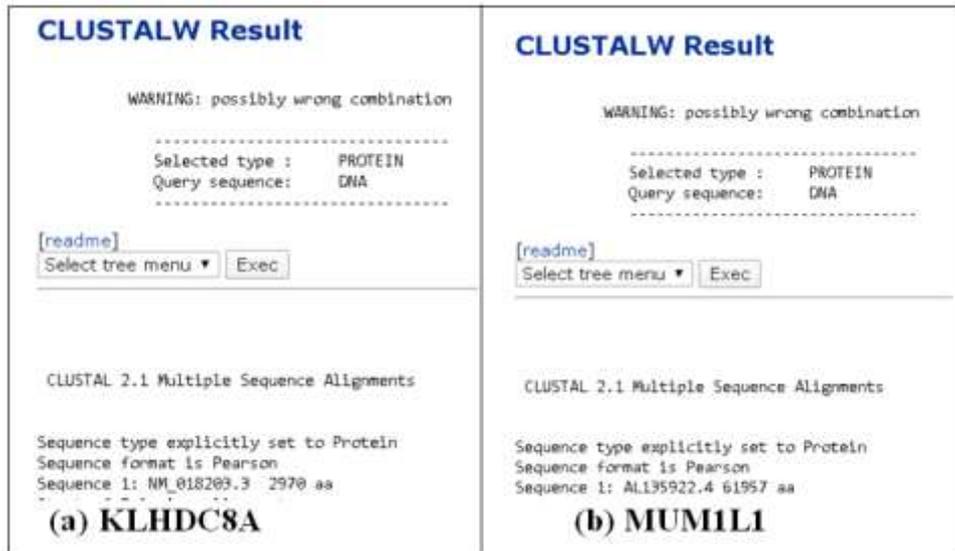


Fig-2: Number of base pairs of identified genes

```

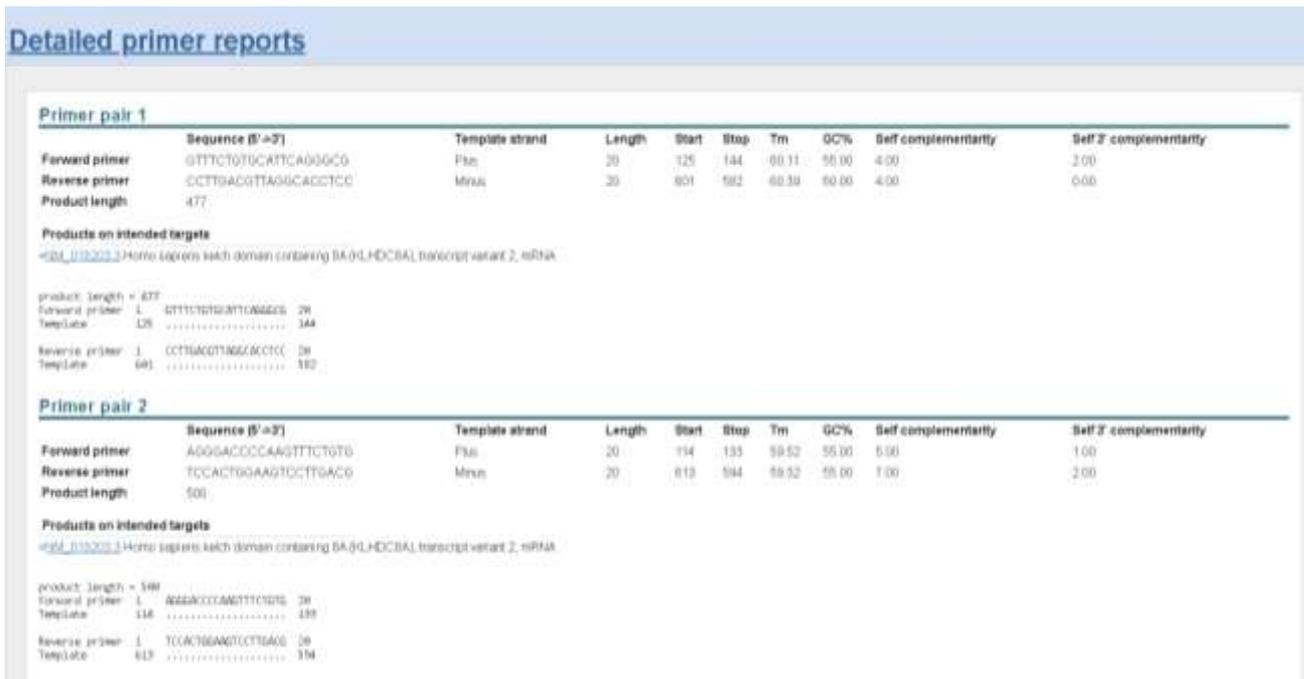
61021 gggactactg cattttctta ctaagccttg agcttttaaa tgtctaaca taccctgtaa
61081 tactttatga gcttcaggtc ttaagggata ttgcttttga taaggaaaaa tggtagggctc
61141 ttttagcctg atttggactg ggtgggcatt ttttgccctt ccgaattgtc ctccaatgc
61201 ccagacttca aggttgattc cctcctcaaa taggggacaa caaatgggta actgtttccc
61261 catattcatg tagatcatag cttcagcttt ggctaatatg tcccctccca ataaggggtg
61321 gggactttta ggcataagaa aggcattgta aaagagcaaa gtctcccagt tacaatgag
61381 gaggtagggag aaatacctgg ttacaggctg tcccaggatt ccttgatgg taacgggctc
61441 tagggacagc tgtctaggat agggagatac actgagaaa acacgccagt gtccaggagg
61501 aagtaattt cctggccctc aatggttaaa catatccggg gctcagttag ggtgatgaca
61561 taggctggcg cttgtcctgg tcaccctcag tctgtttgtt ggatcatctg tttgggctt
61621 ctggcccaga gaacctttgt cctttggggc agtgtgcctt ccagtattg ccttggcata
61681 gtggacatgg gcgaggaggc agcttgtttc ttgttggaca atctttttta aggtgtcttt
61741 gcaaacctca ctgtaacaa gccctaccgg gtgattggcc tgctccattt tctgtcctc
61801 ctgaaccacc aaggtttgtt tgtctgaagg ccatgactaa ggctgcagcc tttctctgat
61861 ctgccttttc ctttttggcc tgttctctt ggtccctatt atagaacacc gaggttgcca
61921 ggttaataa tgccctcaga ttttattcag ggtccag
    
```

Fig-3: Number of sequences for gene MUM1L1

### Designing of Primers for detecting the disease

The primers for the retrieved sequences were designed through NCBI database. For each gene, KLHDC8A and MUM1L1, 10 different primers were designed. The designed primers are shown in the Figure 4(a) & 4(b). These primers can be used in future for the treatment of the PCOS disease by manipulating the identified gene through polymerase chain reaction followed by genetic engineering.

In a study conducted by Artimani *et al*, 2016, similar results have been identified for primer designing of follicle stimulating hormone receptor gene using bioinformatics tool and software. Accession number for the gene is 000145 and forward and reverse primers were GAGAGCAAGGTGACAGAGATTC and TTGATGTAGAGCAGGTTGTTGG respectively [7].



(a) KLHDC8A



(b) MUM1L1  
 Fig-4: The designed Primers

**SUMMARY AND CONCLUSION**

Poly Cystic ovary syndrome (PCOS) is a common cause of an ovulatory infertility in upto 5-10% of women of reproductive age. It is believed that both genetic and environmental factors play important roles in the occurrence and development of PCOS. It is an endocrine disorder. The medications recommended by the doctors to regulate menstrual cycles are the combination of birth control pills and progestin therapy, but these pills have certain side effects like weight gain, cardiovascular and thromboembolic events. Therefore, as an alternative, computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease. The various tools and databases used in the field of bioinformatics are National Centre for Biotechnological Information (NCBI), Gen bank, European Nucleotide Archive, Ensemble, Basic Logic Alignment Search Tool (BLAST). The present study shows, genes responsible for causing the disease (PCOS)- KLHDC8A and MUM1L1 have been identified. The sequences for the identified genes have been retrieved using the

computational software like CLUSTAL W. Through NCBI and BLAST, 10 pairs of the primers for each identified gene have been designed which can be further used for the detection of PCOS by polymerase chain reaction and treatment by genetic manipulation. This can help in the treatment of the disease without any side effects which are caused by taking medications.

**REFERENCES**

1. Prathap, A., Subhalakshmi, T. P., & Varghese, P. J. (2018). A cross-sectional study on the proportion of anxiety and depression and determinants of quality of life in polycystic ovarian disease. *Indian journal of psychological medicine, 40(3)*, 257-262.
2. Choudhary, A., Jain, S., & Chaudhari, P. (2017). Prevalence and symptomatology of polycystic ovarian syndrome in Indian women: is there a rising incidence. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology, 6(11)*, 4971-4975.

3. Bharathi, R. V., Swetha, S., Neerajaa, J., Madhavica, J. V., Janani, D. M., Rekha, S. N., ... & Usha, B. (2017). An epidemiological survey: Effect of predisposing factors for PCOS in Indian urban and rural population. *Middle East Fertility Society Journal*, 22(4), 313-316.
4. Radhakrishnan, R., & Verghese, A. (2018). A study on anxiety and depression among patients with polycystic ovary syndrome. *Journal of Drug Delivery and Therapeutics*, 8(5-s), 338-340.
5. Domecq, J. P., Prutsky, G., Mullan, R. J., Sundaresh, V., Wang, A. T., Erwin, P. J., ... & Murad, M. H. (2013). Adverse effects of the common treatments for polycystic ovary syndrome: a systematic review and meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, 98(12), 4646-4654.
6. Thapar, P. (2018). Bioinformatics- tools and applications. In the proceedings of the 12th INDIACom International Conference, 5044-5047.
7. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Olsson, I. (2015). Tissue-based map of the human proteome. *Science*, 347(6220).
8. Panda, P. K., Rane, R., Ravichandran, R., Singh, S., & Panchal, H. (2016). Genetics of PCOS: a systematic bioinformatics approach to unveil the proteins responsible for PCOS. *Genomics data*, 8, 52-60.
9. Artimani, T., Saidijam, M., Aflatoonian, R., Ashrafi, M., Amiri, I., Yavangi, M., ... & Mehdizadeh, M. (2016). Downregulation of adiponectin system in granulosa cells and low levels of HMW adiponectin in PCOS. *Journal of assisted reproduction and genetics*, 33(1), 101-110.