International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 1 | Issue – 6

Computational Prediction of Damage Associated Non Synonymous SNPs of CYP17A1 and CYP19A1 Gene

Richa Bhatnager, Ranjeet Kaur, Twinkle Dahiya, Amita Suneja Dang*
Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India

ABSTRACT

The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. *CYP17A1* and CYP19A1 are the key proteins involved in steroidogenesis. Majority of polymorphisms have yet not been evaluated for their possible damaging effects on protein structure and function of both the genes. In present analysis, we tried to find out the effect of damage associated nsSNPs of both CYP17A1 and CYP19A1 gene by using various computational algorithms. Out of 205 and 246 nsSNPs reported for CYP17A1 and CYP19A1 gene, total 13 and 7 were predicted to affect the structure and function respectively. R125Q, R416H and F453S substitutions of CYP17A1 gene were not reported previously hence provide the new dataset of unexplored SNPs for their validation. By our results we also found that substitution of mainly Arginine or Phenyl alanine within conserved domain of redox binding, heam binging and redox partner site would cause partial or complete loss of function of both CYP17 and CYP19 gene.

Keywords: *CYP450, Steroidogenesis, Aromatase, SNP, Oestrogen, Testosterone*

INTRODUCTION

Cytochrome P450 comprise superfamily of genes that synthesise cell membrane bound hemoproteins which are involved in the metabolism of various molecules and chemicals within the cell (Guengerich, 2008). Cytochrome P450 enzymes account for 70 to 80 percent of enzymes involved in drug metabolism, thus mutation in these genes can affect the metabolism of drugs (Hanukoglu, 1992) and may results in build up of substances which are harmful for the body or prevent other necessary molecules from being produced. In addition to metabolism, some members of CYPs family (CYP17 and 19) are involved in the synthesis of sex steroids. Cytochrome P450 17A1 (CYP17A1; also P450c17), also called 17α-hydroxylase or 17, 20 lyase, is an enzyme of the hydroxylase class that in humans is encoded by the CYP17A1 gene on chromosome 10q24.3. It contains eight exons which encodes a polypeptide of 508 amino acid of 57.4 kDa (Akhtar et al. 2005). It is ubiquitously expressed in many tissues and cell types, including the zona reticularis of the adrenal cortex and zona fasciculata as well as gonadal tissues (Lunn et al.1999; Miller et al. 1997). CYP17A1 protein has two main functions i.e. in steroidogenesis and in drug metabolism. It catalyzes the 17 α -hydroxylation of pregnenolone (Preg) to 17α-OH pregnenolone (17 α–OH Preg), and subsequently through its C17,20 lyase activity, it can further convert 17 α -OH Preg to dehydroepiandrosterone (DHEA), which finally can be oxidized at the 3rd position by 3 β -hydroxysteroid dehydrogenase (3β-HSD) to generate androstenedione. Its activity is regulated by the interaction of its redox partners (cytochrome b5 and Cytochrome P450 reductase). Any mutation in them may abolish its activity. A number of natural variant of CYP17A1 gene are reported in Congenital Adrenal Hyperplasia. Mutations in Cytochrome P450 Reductase (POR) are reported to decrease the activities of CYP17A1 by 60-80% (Pandey et al. 2007). Some mutations in Cytochrome b5 (Cytb5) coding region leading to the formation of premature

stop codons that results in an faulty (incomplete) Cytb5 domain and loss of C17,20 lyase deficiency (Flück et al. 2010).

CytochromeP450 19A1, also called oestrogen synthetase or aromatase is an enzyme complex (P450arom) that catalyses the conversion of the C19 steroids to C18 steroids (Simpson ER, 1994) i.e. the conversion of the C19 androgens, androstenedione and testosterone, to the C18 estrogens, estrone and estradiol respectively (Meigs, 1968). The human CYP19A1 (P450arom) gene is located on the long arm of chromosome 15 (15q21.2). Gene spans a region that consists of a 30 kb coding region and a 93 kb regulatory region, a total of 123 kb in length (Chen, 1988). Only the 30 kb region encodes aromatase, whereas a large 93 kb 5'-flanking region serves as the regulatory unit of the gene which comprises 10 promoter sites (Shozu et al. 1998). The aromatase enzyme can be found in many tissues including gonads (granulosa cells), brain, adipose tissue, placenta, blood vessels, skin, and bone, as well as in tissue of endometriosis, uterine fibroids, breast cancer, and endometrial cancer (Chen et al. 2004). The mutations in CYP19 gene causes androgen excess and decrease of oestrogen which may lead to ovarian inactivity and deficient follicular development (Shozu et al. 2003). Also, CYP19A1 polymorphisms have been associated with numerous complex diseases such as heart disease, diabetes, and cancer (Dungan et al. 2016).

Both CYP17 and CYP 19 are the key candidates in steroidogenic cycle. CYP17 involves in the synthesis of DHEA which is the precursor for androstenedione; intermediate for testosterone and oestrogen synthesis whereas CYP19 catalyse the conversion of androgens to estrogen. Both testosterone and oestrogen are responsible for puberty and reproduction in male and female respectively. In absence of androstenedione, testosterone or oestrogen will not form and resulted in disorders. It is well known that structure of protein is one of the factor responsible for its function thereby any alteration in their structure function have the potential affect its function. The most common genetic mutations are the single-nucleotide polymorphisms (SNPs). Nonsynonymous SNPs are functionally important SNPs; since they occur in a coding region and caused an amino-acid change in the respective protein thereby have the potential to alter structure as well as function of the protein. In human genome, around 500,000 SNPs occurs into the coding region (Collins, 1998). Therefore present study was carried out to study the nsSNPs associated with CYP17 and CYP 19 gene by different computational software to predict their potential effect on the structure and function of the protein so that their establish their association with the diseases.

MATERIALS AND METHODS

SNP retrieval: SNPs of CYP17A1 and CYP19A1 genes and their protein sequences were retrieved in FASTA format from dbSNP database (http:// www.ncbi.nlm.nih.gov/SNP/) and protein database of NCBI. SNPs related to *Homo sapiens* were selected by using filters non synonymous, missense, nonsense, stop gained SNP and human (Bhagwat, 2010). Gene Ids and ENSEMBL ids were retrieved from Genbank database of NCBI (http://www.ncbi.nlm.nih.gov/gene/).

Prediction of the effect of nsSNP in-silico: Nonsynonymous changes can be scrutinized according to their potential to change biochemical property that results because of the amino acid substitution within the protein sequence. There are a number of computational methods available to anticipate the effect of amino acid substitutions on protein structure and function. These methods employ different algorithms and are based on different approaches like structure, sequence, Hidden markov model and machine based programs.

SIFT (http://sift.jcvi.org/): Sorting Intolerant from Tolerant (SIFT) predicts the functional impact of an amino acid substitution based on the alignment of highly similar orthologous and paralogous protein sequences. Its prediction is based on the conservation of amino acid in the protein family which reflects the importance of that amino acid for the normal function and structure of that protein (Ng, 2003). It takes the query sequence and performs multiple sequence alignment to predict the function score in the form of median info and probability score. Positions with probability score less than 0.05 are considered to be deleterious; those greater than or equal to 0.05 are considered to be tolerated. In our study, we submitted rsids retrieved from dbSNP as a query to make prediction.

PROVEAN: Protein variation effect analyzer predicts whether the substitution of amino acid is deleterious or tolerated. The threshold for a mutation to be

deleterious is -2.5; if the score lies below the threshold then the substitution will be deleterious and vice versa. Provean program was used to predict the functional effect of single or multiple amino acid substitutions, insertions or deletion (Choi et al. 2012).The input data was those amino acid substitutions that were analyzed by SIFT program. NCBI ref sequence id of prolidase protein was also used for provean analysis.

Mutation Assessor: (http://mutationassessor.org/r3/)**:** It predicts the functional impact of amino acid substitutions in proteins which is based on evolutionary conservation of the affected amino acid in protein homologs. It employs a combinatorial entropy optimization' (CEO) to search all the sets of evolutionarily related proteins and find key functional residues and then assigns a conservation score to them. Its output contains two annotation i.e. FI score (functional impact score) and functional impact (high, medium, neutral). In our study, we submitted amino acid substitutions predicted by SIFT server as a query.

PANTHER (http://pantherdb.org): PANTHER (Protein ANalysis THrough Evolutionary Relationships) is a classification System was designed to classify genes and their proteins in order to facilitate high throughput analysis. It aligns a query sequence to HMMs of protein families and subfamilies in its collection. In gene list analysis, it analyzes the list of gene, and expression data files with PANTHER. Evolutionary analysis of coding SNPs estimates the likelihood that a particular nonsynonymous coding SNP will cause a functional impact on the protein or not. Score is a negative algorithm of the probability ratio of the wild type and mutant amino acid at a given position and falls between 0(neutral) to approx (-10) deleterious. We submitted protein sequence and prediction of SIFT server as input for PANTHER prediction.

PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/): It predicts the functional impact of single amino acid substitution on protein function by using physical and comparative models generated by the sequence information. Its prediction is based on a number of features such as sequence, structure and phylogenetic comparison to analyze the mutation (Adzhubei et al. 2010). For our analysis, we submitted FASTA sequence along with amino acid substitutions as query.

nsSNP Analyzer (http://snpanalyzer.uthsc.edu/): It predicts the phenotypic effect of nonsynonymous substitution. It uses multiple sequence alignment and protein 3D structure to predict the result. nsSNPAnalyzer uses "Random Forest" network i.e. a machine learning method to classify the nsSNP from native one. Its prediction is purely dependent on Swissprot database and was trained using a curated SNP dataset. nsSNP Analyzer summarizes the structural environment of the mutated residue and similarity between the substituted and native residue from the normalized probability of the substitution in the multiple sequence alignment (Bao et al. 2005). FASTA sequence was submitted to carry out the analysis.

PhDSNP (http://snps.biofold.org/phd-snp/phd-snp.html): It is support vector machine based software which supports the local sequence environment and output of multiple sequence alignment to predict the nature of a particular mutation. It requires input in the form of protein sequence, residue position, and new residue. Output is based on reliability score which predict whether the substitution is disease causing or neutral (Capriotti et al. 2006). In our analysis, protein sequence along with substituted amino acid and substitution position was submitted to carry out analysis.

SNAP (https://rostlab.org/services/snap2web/): SNAP2 predicts the effect of substitution of single amino acid on the function of protein. It is a machine learning method that uses neural networks to make predictions between wild type and mutant residue by considering a number of sequences and properties in account. It subjects the submitted sequences to multiple sequence alignment for evolutionary similarities and uses results as its input. Its input format only requires protein FASTA sequence to make its prediction.

StSNP (http://ilyinlab.org): It is a web server which compares structural nsSNP distributions in many proteins or protein complexes. It provides the ability to analyze and compare human nsSNPs in protein structure, protein complexes, protein-protein interfaces and metabolic processes. Its input supports multiple formats i.e. protein identification number /PDB code/ keyword or full name of the protein/ rsids. We submitted protein name for its prediction

MutPred (http://mutpred.mutdb.org/): It is a web tool which classifies an amino acid substitution as disease-associated or neutral in human along with molecular factors related to this substitution. MutPred is based upon SIFT and a gain or loss of 14 different structural (namely secondary structure, trans-membrane helices, coiled-coil structure, stability, solvent accessibility, B-factor, and intrinsic disorder) and functional (namely DNA-binding residues, catalytic residues, calmodulin-binding targets, as well as the phosphorylation, methylation, ubiquitination and glycosylation sites) properties (Finn et al. 2016; Delorenzi, 2002). The output of MutPred consists of a general score (g), i.e., P (deleterious) the probability that the amino acid substitution is deleterious or disease-associated, and top five characteristic scores (p), where p is the P-value that certain functional and structural characteristics of the protein are impacted. Certain combinations of high values of 'g' (p deleterious) and low values of 'p' (property scores) are referred as hypotheses. • Scores for an aas with $g > 0.5$ and $p < 0.05$, are referred as actionable hypotheses. • Scores for an aas with $g > 0.75$ and $p < 0.05$, are referred as confident hypotheses. • Scores for an aas with $g > 0.75$ and $p < 0.01$, are referred as very confident hypotheses (Sickmeier, 2007). User input involves FASTA sequence and amino acid substitutions.

I-Mutant (http://folding.biofold.org/i-mutant/i-mutant2.0.html): It is a neural-network-based web server for the automatic prediction of protein stability changes upon single-site mutations. I-Mutant 2.0 can be utilized both as a classifier that predicts the signs of the protein stability changes upon a variation and as a regression estimator that predicts the relative change in Gibbs-free energy (ΔG) at a given temperature (Capriotti et al. 2005). We submitted protein's sequence with substituted amino acid and positions for the effect of mutation on protein stability. Temperature and pH range were set as default i.e. 25°C and pH 7.

RESULTS

Retrieval of SNP IDs:

All the nsSNPs were retrieved from dbSNP database using appropriate filters. For CYP17A1, a total of 2032 SNPs were found, out of which 1428 were related to human. After applying filters, a total of 205 nsSNPs were reported as nsSNPs (Figure 1). For CYP19A1, a total of 22200 SNPs were found out of which, 15737 were human related. After filer applications, total 246 nsSNPs were predicted as nsSNPs(Figure 2). rsIDs of all these human related nsSNPs and their FASTA sequence were obtained which were used for the prediction of their effect on the structure and function of both the genes.

Prediction of amino acid substitution by SIFT:

SIFT provides prediction for a list of nsSNPs based on sequence homology and physical property of amino acids. It predicts whether the amino acid substitution at a given position is tolerated or cause disease. A total of 205 rsIDs (CYP17A1)and 246rsIDs (CYP19A1) were submitted for SIFT prediction and it predicted 25nsSNPs and 8 nsSNPs as damaging for CYP17A1(Table1) and CYP19A1 (Table 2) respectively. Amino acid substitutions predicted by SIFT server were further analyzed by different softwares.

Identification of disease-associated nsSNPs by Mutation Accessor

Mutation Accessor predicts the functional impact of nsSNP on the basis of evolutionary conservation of affected amino acid in protein homolog. It is validated on the OMIM, PDB, and SwissProt to make prediction and produces functional impact score. For CYP17, 28 substitutions were submitted to predict their functional effect and it was predicted that 11 nsSNPs were having high impact, 6 were at a medium scale, 5 nsSNP did not had any impact on protein structure and function(Table 1). For CYP19, Out of 12 substitutions submitted for prediction, 7 were found to have medium impact, 1 at low scale that do not impact protein structure and function (Table 2).

Identification of disease-associated nsSNPs by PolyPhen:

Polyphen i.e. polymorphism phenotype is a tool predicts the possible effect of amino acid substitution on function and structure of protein in HumDiv and HumVar. Polyphen classify amino acid substitutions in three different categories i.e. probably damaging, possibly damaging and benign based on PSIC (position-specific independent count) matrix. Prediction having score near to 1 are predicted as probably damaging whereas in between 0.5-1 are grouped as possibly damaging while those having score below 0.5 are predicted as benign. For CYP17, 28 substitutions predicted by SIFT serve were used as input and it predicted 19 substitutions were probably damaging, 3 were possibly damaging and remaining 6

were benign (Table 1). For CYP19, Out of 12 substitution submitted for PolyPhen analysis ,8 were predicted as probably damaging, 1 was possibly damaging and remaining 3 were benign(Table 2).

Prediction of functional effect of nsSNPs by PROVEAN:

PROVEAN predicts the functional effect of substitution on the protein. A threshold of prediction i.e. -2.5 is set, below this threshold SNPs showed negative effect and were predicted to be deleterious. For CYP17, total 28 substitutions were submitted, out of which 17 were predicted as deleterious (Table 1), For CYP19, total 12 substitutions were submitted, out of which 10 were found deleterious (Table 2).

Prediction of disease associated SNP by SNAP2:

SNAP2 predicted the functional impact of nsSNP with its association to cause disease. It produces output in heat map representation where dark red color indicates the affected one and green color indicates the neutral substitution. For CYP17, Out of 28 substitution, it predicted 18 substitutions were functionally damaging whereas remaining 10 were neutral (Table 1). For CYP19, Out of 12 substitutions, it predicted 10 substitutions were associated with diseases whereas remaining 2 were neutral (Table 2).

Disease associated SNP prediction by nsSNP analyzer and PhD SNP

Both nsSNP analyzer and PhD-SNP predict the phenotypic effect of non synonymous substitutions. They also predict whether the substitution is disease associated or not. nsSNP analyzer utilises the information stored in protein 3D structure and subjected it to multiple sequence alignment to make its prediction, while PhD-SNP prediction was based on machine learning methods and utilises protein sequence and profile information to predict the results. Out of 28 nsSNPs of CYP17 and 12 nsSNPs of CYP 19, nsSNP analyzer predicted 18 SNPs had severe phenotypic effects in CYP17 and 8 had in CYP19(Table 1). PhD-SNP predicted 16 SNPs of CYP17 and 8 of CYP19 had severe phenotypic effects and associated with diseases (Table 2).

Prediction of the Stability Change by I-Mutant:

I-Mutant predicts the effect of nsSNP on protein stability due to mutation. Its prediction comes out in two forms i.e. change in DDG and ΔG. Positive G value leads to increased stability whereas negative G values correspond to decreased stability. For CYP17, out of 28 SNPs, 27 were found to decrease the stability of protein stability (Table 1). For CYP19, all 12 nsSNPs were decrease the stability of protein (Table 2).

Prediction of disease associated substitutions by MutPred:

MutPred (Mutation Prediction) is an algorithm which predicts whether an amino acid substitution (AAS) will be disease-associated or neutral. As it predicts the molecular cause of disease/deleterious, it also predicts the effect of non synonymous mutations on the protein binding activity, chaperone binding activity, protein function and interaction with other subunits or transcription factors. A missense mutation with a MutPred score > 0.5 could be considered as 'harmful', while a MutPred score >0.75 should be considered a high confidence 'harmful' prediction. For CYP17, 21 of these nsSNPs the program indicated a ‘harmful’ prediction with g score more than 0.5. 19 nsSNPs showed a probability of being a ‘confidence harmful’, with *g* scores higher than 0.75 (Table 3). For CYP19, 8 nsSNPs predicted ‘harmful’ with g score more than 0.5 and 7 of these were predicted ‘confidence harmful’ with g score more than 0.75(Table 4).

DISCUSSION

Cytochrome P450 is a collection of proteins which contain heam as a cofactor that are involved in metabolism and steroidogenesis. They are, in general, the terminal oxidase enzymes in electron transfer chains, broadly categorized as P450-containing systems. There are around 200,000 members of CYP family which are involved in metabolism (Guengerich, 2008.) CYP17 and CYP19 are the members of CYP family that are involved in production of sex hormones and their intermediate. Therefore any alteration in these proteins affects the synthesis of sex steroids and contributes to the diseases.

CYP17A1 (steroid 17α-hydroxylase/17, 20-lyase) plays a crucial role in the steroid hormone biosynthesis (Lunn et al. 1999). CYP17 is unique because of its ability to catalyze two different types of reactions, the 17 α-hydroxylase and 17,20-lyase reactions, in one active site. 17,20- lyase activity of this enzymatic system is particularly sensitive to alterations in the interactions between P450c17 and its cofactor proteins P450-oxidoreductase and Cytochrome B5 (Miller et al. 1997). So any mutation

in the binding site of redox partners abolishes the enzyme activity and leads to severe consequences.205 rsids were submitted for SIFT prediction and it predicted 28 SNPs which were evaluated further for their effect on structure and function of the protein. During prioritization, we got 16 (R96W, R96Q, R125Q, R347H, R347C, R358Q, R362C, R416H, F93C, F114V, F417C, F453S, S106P, P342T, W406R and D116V) mutations that were predicted damaging by all the tools used in this study. These substitutions involve the changes mainly in two amino acids i.e. Arginine and Phenyl alanine. Interestingly, in all the 28 substitutions it is noted that replacement of Arginine is not tolerated, making the prediction damaging and substitution pathogenic. Out of these 16 substitutions, 13 were the natural variant in CYP17 gene studied in congenital adrenal hyperplasia only. From our analysis, we found 3 novel substitutions (R125Q, R416H, and F453S) which are not reported yet. R125Q substitution entails the replacement of arginine by glutamine. This is a conserved residue present at the fifth position in the canonical WXXXR motif at the amino terminus of c helix. Being a conserved residue, any substitution at these position leads to structural changes in motif and interfere the interaction with other motif to form a complete reaction centre hence results in complete loss of 17α-hydroxylase/17,20-lyase activities . R416H substitution occurs because of a missense mutation at position +1247G/A substitution in the exonic region of CYP17A1. Residues ranging from 414-417 forms the helix which is a part of the redox centre of the protein that needs to interact with cytochrome b(5) complex to carry out electron transport. After substitution of arginine, structure of redox centre get distorted leading to loss of activity of the enzyme. Another mutation at this helix is F417C involves the substitution of phenyl alanine to cysteine. Both phenyl alanine and cysteine belongs to different group i.e. 'C' is an uncharged amino acid having polar side chain where 'F' is a non polar amino acid. Substitution of F by C alters the helix structure due to formation of sulphide bond because of Cysteine. F/C substitution also results in the loss of phosphorylation for the activation of the components of redox centre. Hence both of these mutation results in complete loss of 17α-hydroxylase/17, 20-lyase activities. F453S substitution entails the replacement of Phenyl alanine by Serine and located in the helix ranging from 445-462. This substitution was found to decrease the stability of protein as predicted by MutPred and I-Mutant program. In addition to stability, it increases the relative solubility and methylation of protein probably due to –OH group of serine residue. But this substitution resulted in loss of catalytic property as predicted by Mutpred.

CYP19A1 also known as aromatase catalyzes the formation of aromatic C18 estrogens from C19 androgens i.e. convert testosterone into oestrogen and estrone. Besides aromatase activity it has a number of other activities like steroid hydrolysis, as electron carrier, heam binding and oxido reductase activity (Shozu et al.1993; Shozu et al. 2003). Its deficiency as well as excess due to the point mutation can cause disorders like aromatase excess syndrome (AXES) and aromatase deficiency syndrome (AROD). AXES is an autosomal dominant disorder characterized by increased extra glandular aromatization of steroids that presents with heterosexual precocity in males and iso-sexual precocity in females. On the other hand, AROD is a rare disease in which fetal androgens are not converted into estrogens due to placental aromatase deficiency. During prioritization, 7 (M364T, R159L, C437Y, R365Q, R435C, R375C, E210K) mutation were predicted as damaging by all the tools used in this study. R365Q mutation is at base pair 1094 in exon 9 of the P-450 aromatase gene, results in a glutamine instead of an arginine at position 365. Amino acid residues 354-366 are involved in helix formation with is a part of redox centre of the protein thereby abolish the electron transfer chain and may cause aromatase deficiency leading to the excess of testosterone (C19 androgen). M364T involves the change of methionine (S containing amino acid) into threonine (-OH group containing amino acid). As 'M' gets substituted by 'T', conversion of non polar residue into uncharged polar residue would leads to disruption of protein structure and thus, abolishment of its function. Arginine and cysteine are highly conserved residues among the entire CYP family. Any mutation leading to change in these amino acids would cause disorders. R435C and C437Y involve the substitution within these conserved residues. These substitution were found in exon X having missense mutations 1303C/T and 1310G/A. These missense mutations reside within the heme-binding region of the protein. Cysteine-437 is the conserved cysteine that makes up the fifth coordinating ligand of the heme iron, while Arginine-435 is also a highly conserved residue within the heme-binding region among mammalian P450s. Both of these mutations would subject to complete loss of heam binding activity. R375C, another missense

mutation results in disruptive α-helix because of change of arginine to cysteine. Cysteine molecule here tends to make sulphide bond and may disrupt the helix. E210K substitution involves the drastic change of Glutamic acid (acidic amino acid) into lysine (a basic amino acid). This region involves the change in helix ranging from 210-227. It makes the third coordinating position in heam binding subunit and alters the electron transfer chain, resulting in hampering of protein function. The last damaging mutation predicted by our results is R159L, involves the substitution of basic amino acid into non polar into helical region (155-172). This domain forms the redox binding partner and involves in transport of electrons to cytb5 subunit. As being non polar leucine cannot confer electron transport making the ETC to hamper and abolish the activity of protein.

The analysis of the SNP involved in the determination of variation in phenotypes or in complex diseases is a challenge that requires different approaches to study them. Here, we used different methods to predict the most damaging mutations in the human *CYP17A1* and *CYP19A1*, the key proteins for steroidogenesis. Although some of the polymorphisms found in both genes have been studied in the laboratory, many others have not yet been evaluated with respect to their possible damaging effects on protein structure and function. 13nsSNPs in CYP17A1 gene and 7 nsSNPs in CYP19A1 were predicted as damaging by the algorithms used in the study. Out of 13nsNSPs, 3 nsSNPs of CYP17A1 gene are not reported till date hence can be validated to find their association with the diseases.

References:

1) Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A 2010. A method and server for predicting damaging missense mutations. *Nat Methods*, 7: 248–249.
2) Akhtar MK, Kelly SL, Kaderbhai MA 2005. Cytochrome b(5) modulation of 17{alpha} hydroxylase and 17-20 lyase (CYP17) activities in steroidogenesis. *J Endocrinol*, 187:267-274.
3) Bao L, Zhou M, Cui Y 2005. nsSNPAnalyzer: identifying disease-associated non synonymous single nucleotide polymorphisms. *Nucleic Acids Res*, 33:480–482.
4) Bhagwat M, 2010. Searching NCBI's dbSNP database. *Curr, Protoc, Bioinformatics* (Chapter 1:Unit 1,19).
5) Capriotti E, Calabrese R, Casadio R 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics,* 22: 2729–2734.
6) Capriotti E, Fariselli P, Casadio R 2005. I-Mutant 2,0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res,* 33:W306–W310.
7) Chen SA, Besman MJ, Sparkes RS, Zollman S, Klisak I, Mohandas T, Hall PF, Shively JE 1988. Human aromatase: cDNA cloning,Southern blot analysis, and assignment of the gene to chromosome15. *DNA*, 7:27–38
8) Choi Y, Murphy GE, Miller S, Chan JR 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, 7:e46688.
9) Collins FS and Brooks LD 1998. A DNA polymorphism discovery resource for research on human genetic variations. *Genomic Res*, 8:1229–1231
10) Delorenzi M and Speed T 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625.
11) Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res,* 44: D279–D285.
12) Flück CE, Mallet D, Hofer G, Samara-Boustani D, Leger J, Polak M, Morel Y, Pandey AV 2011. Deletion of P399_E401 in NADPH cytochrome P450 oxidoreductase results in partial mixed oxidase deficiency. *BiochemBiophys Res Commun*, 412: 572- 577.
13) Guengerich FP 2008. Cytochrome p450 and chemical toxicology. *Chem Res Toxicol,* 21:70-83.
14) Hanukoglu L 1992. Steroidogenic enzymes: Structure, function and role in regulation of steroid hormone biosynthesis. *Journal of Steroid Biochemistry and Molecular Biology,* 43:779-804.
15) Jennifer DR, Xuejun Q, Benjamin HD, John CF, Singh A, Melissa H, Elizabeth G, Carol H, Gregory SG, Shah SH, Elizabeth RH, William KE 2016.Case-Only Survival Analysis Reveals Unique Effects of Genotype, Sex, and Coronary Disease Severity on Survivorship. *PLoS One,* 11: e0154856

16) Lunn RM, Bell DM, Mohler JL, Taylor JA 1999. Prostate cancer risk and polymorphism in 17-hydroxylase (CYP17) and steroid reductase (SRD5A2). *Carcinogenesis*, 20: 1727-1731.
17) Makio S, Siby S, Kazuto T, Wei-Tong H, Roger A. Kirk N, Michael B, Serdar EB 2003. Estrogen Excess Associated with Novel Gain-of- Function Mutations Affecting the Aromatase Gene. *N Engl J Med*, 348:1855-1865
18) Meigs RA and Ryan KJ 1968. Cytochrome P-450 and steroid biosynthesis in the humans. *Biochem. Biophys. Acta*, Ã65: 476-482, 1968.
19) Miller WL, Auchus RJ, Geller DH 1997. The regulation of 17, 20 lyase activity. *Steroids*, 62(1): 133-142.
20) Ng PC and Henikoff S 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31: 3812–3814.
21) Pandey AV, Kempná P, Hofer G, Mullis PE, Flück CE 2007. Modulation of human CYP19A1 activity by mutant NADPH P450 oxidoreductase. *Mol Endocrinol*, 21: 2579-2595.
22) Sickmeier M 2007. DisProt: the database of disordered proteins. *Nucleic Acids Res,* 35:D786–D793.

Table 1. Effect of nsSNP of CYP17A1 gene on protein structure and function by different software

rs IDs	Substitution	SIFT	Provean	PolyPhen	SNAP2	PhDSNP	nsSNP	FATHMM	StSNP	Mutation Assessor	I-Mutant
rs762563	C22W	Dam	Neu	Ps.D	Neu	Neu	Neu	Tol	Inc	Low	**Dec**
rs58822002	T152R	Dam	Neu	Benign	Neu	Neu	Neu	Tol	Inc	Neu	**Dec**
rs61754262	M1I	Dam	Neu	Ps.D	Neu	Neu	Dis	Tol	Dec	-	**Dec**
rs61754263	R21K	Tol	Neu	Benign	Neu	Neu	Neu	Tol	Inc	Neu	**Dec**
rs61754265	G77S	Dam	Neu	Pr.D	Neu	Neu	Neu	Tol	Dec	Low	**Dec**
rs61754273	V197D	Dam	Neu	Benign	Neu	Dis	Neu	Tol	Inc	Low	**Dec**
rs61754275	D241Y	Dam	Neu	Ps.D	Effect	Neu	Dis	Tol	Dec	Low	**Dec**
rs61754278	R347H	Dam	Del	Pr.D	Effect	Dis	Dis	Tol	Inc	High	**Dec**
rs61754279	L363F	Tol	Neu	Benign	Neu	Neu	Dis	Tol	Inc	Neu	**Dec**
rs72559703	T11I	Tol	Neu	Benign	Neu	Neu	Neu	-	NA	Neu	**Dec**
rs104894135	S106P	Dam	Del	Pr.D	Effect	Dis	Dis	Tol	NA	Med	**Inc**
rs104894137	P342T	Dam	Del	Pr.D	Neu	Dis	Dis	Tol	NA	High	**Dec**
rs104894138	R96W	Dam	Del	Pr.D	Effect	Dis	Neu	Dam	NA	High	**Dec**
rs104894139	R358Q	Dam	Neu	Pr.D	Neu	Dis	Dis	Tol	NA	Med	**Dec**
rs104894140	F417C	Dam	Del	Pr.D	Effect	Dis	Dis	Dam	NA	High	**Dec**
rs104894142	R362C	Dam	Del	Pr.D	Effect	Neu	Dis	Dam	NA	High	**Dec**
rs104894143	W406R	Dam	Del	Pr.D	Effect	Dis	Dis	Tol	NA	High	**Dec**
rs104894144	Y329D	Dam	Neu	Benign	Effect	Dis	Neu	Tol	NA	Low	**Dec**
rs104894145	P428L	Dam	Del	Pr.D	Effect	Neu	Neu	Tol	NA	Med	**Dec**
rs104894146	F93C	Dam	Del	Pr.D	Effect	Dis	Dis	Dam	NA	High	**Dec**
rs104894147	F114V	Dam	Del	Pr.D	Effect	Dis	Dis	Tol	NA	Med	**Dec**
rs104894148	D116V	Dam	Del	Pr.D	Effect	Neu	Dis	Tol	NA	High	**Dec**
rs104894149	R347C	Dam	Del	Pr.D	Effect	Dis	Dis	Tol	NA	High	**Dec**
rs104894150	Y201N	Dam	Del	Pr.D	Effect	Dis	Neu	Tol	NA	Low	**Dec**
rs104894151	F453S	Dam	Del	Pr.D	Effect	Neu	Dis	Tol	NA	High	**Dec**
rs104894153	R96Q	Dam	Del	Pr.D	Effect	Dis	Dis	Dam	NA	High	**Dec**
rs104894154	R125Q	Dam	Del	Pr.D	Effect	Dis	Dis	Dam	NA	Med	**Dec**
rs104894155	R416H	Dam	Del	Pr.D	Effect	Dis	Dis	Dam	NA	Med	**Dec**

Abbreviations: Dam: damaging, Del: deleterious, Pr.D: probably damaging, Ps.D: possibly damaging, Dis: disease, Tol: tolerated, Dec: decrease, Neu: neutral

Table2. Effect of nsSNP of CYP19A1 gene on protein structure and function by different software

rsIDs	Mutation	SIFT	PROVEAN	PolyPhen	SNAP2	nsSNP	PhDSNP	PANTHER	StSNP	Mutation Assessor	I-Mutant
rs700519	R264C	Tol	Del	Benign	Effect	Dis	Neu	Neu	Dec	low	**Dec**
rs2236722	W39R	Tol	Del	Benign	Effect	Neu	Dis	Neu	Dec	Med	**Dec**
rs2304462	R264H	Tol	Neu	Benign	Neu	Neu	Dis	Neu	Dec	Med	**Dec**
rs17853490	P207	Dam	Del	Pr.d	Neu	Neu	Neu	Neu	Dec	Med	**Dec**
rs28757184	T201M	Tol	Neu	Benign	Effect	Neu	Neu	Neu	Inc	Med	**Dec**
rs56658716	M364T	Dam	Del	Pr.d	Effect	Dis	Neu	Dis	Dec	Med	**Dec**
rs76174961	R159L	Dam	Del	Ps.d	Effect	Dis	Dis	Dis	Na	Med	**Dec**
rs78310315	C437Y	Dam	Del	Pr.d	Effect	Dis	Dis	Dis	Na	Med	**Dec**
rs80051519	R365Q	Dam	Del	Pr.d	Effect	Dis	Dis	Dis	Na	Na	**Dec**
rs121434534	R435C	Dam	Del	Pr.d	Effect	Dis	Dis	Dis	Na	Na	**Dec**
rs121434536	R375C	Dam	Del	Pr.d	Effect	Dis	Dis	Dis	Na	Na	**Dec**
rs121434538	**E210K**	**Dam**	**Del**	**Pr.d**	**Effect**	**Dis**	**Dis**	**Dis**	**Na**	**Na**	**Dec**

Abbreviations: Dam: damaging, Del: deleterious, Pr.D: probably damaging, Ps.D: possibly damaging, Dis: disease, Tol: tolerated, Dec: decrease, Neu: neutral

Table 3. Effect of nsSNP on the function of CYP 17A1 gene predicted by Mutpred

S. No	rsids	Substitutions	Top five mutation
1.	rs61754278	R347H	Loss of catalytic residue at R347 (P = 0.0572) Loss of MoRF binding (P = 0.1236) Loss of helix (P = 0.1299) Gain of disorder (P = 0.1642) Loss of stability (P = 0.1977)
2.	rs104894135	S106P	Gain of sheet (P = 0.0827) Gain of ubiquitination at K110 (P = 0.0913) Gain of catalytic residue at S106 (P = 0.0953) Loss of MoRF binding (P = 0.175) Gain of methylation at K110 (P = 0.1775)
3.	rs104894137	P342T	Gain of catalytic residue at P342 (P = 0.0419) Loss of stability (P = 0.0966) Gain of glycosylation at P342 (P = 0.1039 Loss of disorder (P = 0.1884) Gain of MoRF binding (P = 0.2233)
4.	rs104894138	R96W	Loss of disorder (P = 0.0121) Gain of catalytic residue at M99 (P = 0.0178) Gain of ubiquitination at K91 (P = 0.0671) Loss of MoRF binding (P = 0.0788) loss of mrthylation at R96 p0.099
5.	rs104894139	R358Q	Loss of MoRF binding (P = 0.0907) Loss of methylation at R362 (P = 0.2469) Loss of solvent accessibility (P = 0.3744) Loss of helix (P = 0.3949) gain of catalytic residue atE359 P=0. 445
6.	rs104894140	F417C	Gain of catalytic residue at L418 (P = 0.0086) Loss of stability (P = 0.1398) Loss of disorder (P =

			0.2179) Loss of loop (P = 0.3664) Loss of helix (P = 0.4763)
7.	rs104894142	R362C	Gain of catalytic residue at L363 (P = 0.0138) Loss of methylation at R362 (P = 0.0693) loss of loop p=0.4786 Loss of MoRF binding (P = 0.1028) Loss of helix (P = 0.3949)
8.	rs104894143	W406R	Gain of disorder (P = 0.0134) Loss of catalytic residue at P409 (P = 0.0253) Loss of ubiquitination at K404 (P = 0.1231) Gain of glycosylation at P409 (P = 0.26 loss of stability p0.3048
9.	rs104894146	F93C	Loss of methylation at K89 (P = 0.1065) Loss of MoRF binding (P = 0.1069) Loss of disorder (P = 0.1995) Gain of catalytic residue at G95 (P = 0.2217) Gain of loop (P = 0.2754)
10.	rs104894147	F114V	Gain of MoRF binding (P = 0.0999) Loss of ubiquitination at K110 (P = 0.1343) Gain of sheet (P = 0.1539) Gain of methylation at K110 (P = 0.1844) gain of loop p=0.2045
11.	rs104894148	D116V	Gain of MoRF binding (P = 0.0284) Gain of sheet (P = 0.0827) Loss of disorder (P = 0.1079) Gain of loop (P = 0.2045) Gain of glycosylation at S117 (P = 0.2502)
12.	rs104894149	R347C	Loss of disorder (P = 0.0337) Loss of MoRF binding (P = 0.1004) Gain of helix (P = 0.1736 Loss of catalytic residue at R347 (P = 0.2039) loss of phosphoryalationat s345 p=0.343
13.	rs104894151	F453S	Loss of stability (P = 0.0754) Gain of relative solvent accessibility (P = 0.1259) Gain of methylation at R449 (P = 0.1985 Gain of solvent accessibility (P = 0.28) loss of catalytic residue at a450 p=0.455
14.	rs104894153	R96Q	Loss of catalytic residue at R96 (P = 0.0297) Loss of MoRF binding (P = 0.0472) Loss of methylation at R96 (P = 0.0919) Gain of disorder (P = 0.1257) Gain of ubiquitination at K91 (P = 0.1825)
15.	rs104894154	R125Q	Loss of MoRF binding (P = 0.0632) Loss of helix (P = 0.3949) Gain of catalytic residue at W121 (P = 0.4312) Loss of stability (P = 0.4927) Gain of disorder p=0.5086
16.	rs104894154	R416H	Gain of catalytic residue at L418 (P = 0.043) Gain of disorder (P = 0.077) Loss of helix (P = 0.2662) Loss of stability (P = 0.2674) Loss of sheet (P = 0.5184)

Table 4. Effect of nsSNP on the function of CYP 19A1 gene predicted by Mutpred

S.no	rs ID's	Substitutions	Top 5 features
1.	rs56658716	M364T	Gain of phosphorylation at M364(P =0.0863) Loss of stability(P =0.184) Loss of helix(P =0.3949) Gain of catalytic residue at M364 (P =0.4305) Loss of methylation at R365 (P =0.4383)
2.	rs76174961	R159L	Loss of MoRF binding(P =0.0831) Loss of helix(P =0.1299) Gain of catalytic residue at V161 (P =0.1575) Loss of solvent accessibility(P =0.1922) Gain of loop (P =0.2045G)
3.	rs78310315	C437Y	Gain of MoRF binding(P =0.1213) Loss of ubiquitination at K440(P =0.1691) Loss of methylation at R435(P =0.2042) Gain of helix (P =0.2059) Loss of catalytic residue at Y441 (P =0.2987)
4.	rs80051519	R365Q	Gain of phosphorylation at Y361(P =0.2333) Loss of stability(P =0.2725) Gain of loop(P =0.2754) Loss of catalytic residue at R365 (P =0.3004) Loss of methylation atR365(P =0.3519)
5.	rs121434534	R435C	Gain of catalytic residue at P434 (P = 0.007) Loss of methylation at R435 (P =0.0118) Gain of ubiquitination at K440 (P =0.0581) Loss of MoRF binding(P =0.1839) Gain of phosphorylation at K440 (P =0.1938)
6.	rs121434536	R375C	Gain of methylation at K376 (P =0.0139) Gain of ubiquitination at K376 (P =0.1123) Gain of sheet (P = 0.1945) Loss of loop(P =0.2237) Gain of catalytic residue at M374 (P =0.2753)
7.	rs121434538	E210K	Gain of ubiquitination at E210 (P =0.0177) Gain of methylation at E210 (P =0.0575) Loss of catalytic residue at E210 (P =0.1651) Gain Of MoRF binding(P =0.3043) Gain of loop (P =0.3485)

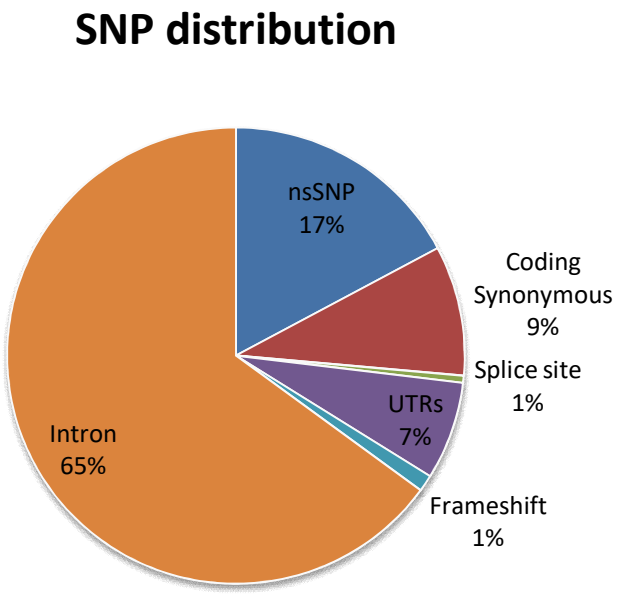

Figure 1. Distribution of SNPs in Human CYP17A1 gene

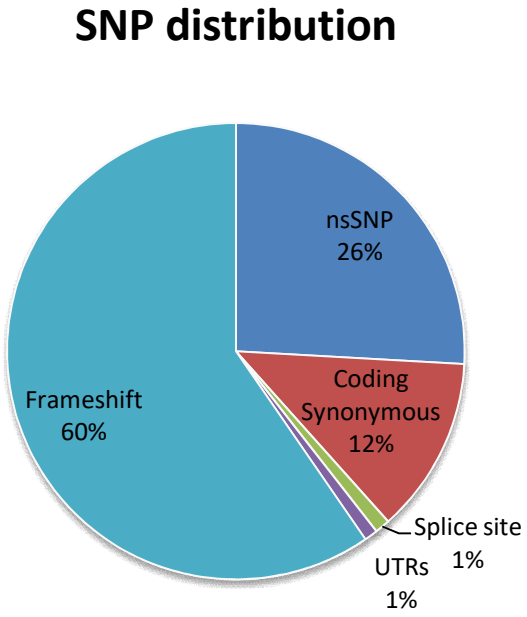

Figure 2. Distribution of SNPs in Human CYP19A1 gene