ORIGINAL RESEARCH

SNP analysis of follistatin gene associated with polycystic ovarian syndrome

Palanisamy Panneerselvam[1]
Kanakarajan Sivakumari[1]
Ponmani Jayaprakash[1]
Ramanathan Srikanth[2]

[1]Department of Zoology, Presidency College, Chennai, Tamil Nadu, India; [2]Department of Biotechnology, Sri Venkateswara College of Engineering, Chennai, Tamil Nadu, India

Abstract: Follistatin has been reported as a candidate gene for polycystic ovarian syndrome (PCOS) based on linkage and association studies. In this study, investigation of polymorphisms in the *FST* gene was done to determine if genetic variation is associated with susceptibility to PCOS. The nucleotide sequence of human follistatin and the protein sequence of human follistatin were retrieved from the NCBI database using Entrez. The follistatin protein of human was retrieved from the Swiss-Prot database. There are 344 amino acids and the molecular weight is 38,007 Da. The ProtParam analysis shows that the isoelectric point is 5.53 and the aliphatic index is 61.25. The hydropathicity is −0.490. The domains in FST protein are as follows: Pfam-B 5005 domain from 1 to 92; EGF-like subdomain from 93 to 116; Kazal 1 domain, occurred in three places, namely, 118–164, 192–239, and 270–316. There are 31 single-nucleotide polymorphisms (SNPs) for this gene. Some are nonsynonymous, some occur in the intron region, and some in an untranslated region. Two nonsynonymous SNPs, namely, rs11745088 and rs1127760, were taken for analysis. In the SNP rs11745088, the change is E152Q. Likewise, in rs1127760, the change is C239S. SIFT (Sorting Intolerant from Tolerant) showed positions of amino acids and the single letter code of amino acids that can be tolerated or deleterious for each position. There were six SNP results and each result had links to it. The dbSNP id, primary database id, and the type of mutation whether silent and if occurring in coding region are given as phenotype alterations. The FASTA format of protein was given to the nsSNP Analyzer tool, and the variation E152Q and C239S were given as inputs in the SNP data field. E152Q change was neutral and C239S causes disease. Using PANTHER for evolutionary analysis of coding SNPs, the protein sequence was given as input and analyzed for the E152Q and C239S SNPs for deleterious effect on protein function. The genetic association database results showed that *FST* gene SNPs are linked to PCOS coming under the disease class of metabolic disorders. The list of intronic and synonymous SNPs, with their nucleotide position, amino acid change information, and dbSNP link, is provided for further analysis.

Keywords: FST, polycystic ovarian syndrome, single-nucleotide polymorphism analysis

Introduction

Polycystic ovarian syndrome (PCOS) is a common endocrine disorder that is found in ≈4% of women of reproductive age[1] and results in reduced fertility and a sevenfold increased risk for type 2 diabetes mellitus.[2] The syndrome is characterized by hyperandrogenism and chronic anovulation. It is also associated with polycystic ovaries, hirsutism, obesity, and insulin resistance. The observation of familial aggregation of PCOS[3–5] showed that it is consistent with a genetic basis for this disorder. The mode of inheritance is uncertain at this point, and the role of shared environmental factors such as diet and lifestyle in the presentation of the disease is unknown. Loci proposed and investigated

Correspondence: Kanakarajan Sivakumari
Department of Zoology, Presidency College, Chennai 600 005, Tamil Nadu, India
Mobile +91 9790934900
Email dr.sivakumari@rediffmail.com

as possible PCOS genes include *CYP11A*, the insulin gene, and a region near the insulin receptor gene.[6] The wide range of PCOS symptoms likely plays a significant role in the inability thus far to identify a specific gene mutation.

Single-nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is changed. Most SNPs, actually about two of every three SNPs, involve the replacement of cytosine (C) with thymine (T). SNPs occur every 100–300 bases along the human genome. SNPs are stable from an evolutionarily standpoint not changing much from generation to generation.

GGCTTC*A***G**AATGGCC

GGCTTC*A***A**AATGGCC

SNPs, which make up about 90% of all human genetic variation, occur every 100–300 bases along the 3-billion-base human genome. SNPs can occur in both coding (gene) and noncoding regions of the genome. Many SNPs have no effect on cell function, but scientists believe others could predispose people to disease or influence their response to a drug. SNP maps will help them to identify the multiple genes associated with such complex diseases as cancer, diabetes, vascular disease, and some forms of mental illness.

Several pathways have been implicated in the etiology of PCOS. These include the metabolic or regulatory pathways of steroid hormone synthesis,[7,8] regulatory pathways of gonadotropin action,[9] the insulin-signaling pathway,[10–12] and pathways regulating body weight.[13] Several genes from these pathways have been tested as candidate genes for PCOS.[7,8,14–19] Although mutation analysis, linkage studies, and case-control association studies have been carried out with these candidate genes, evidence that any of them play a role in PCOS has not been replicated widely and is still inconclusive. These uncertainties are common in 'complex' genetic diseases, where identifying the contributing genes is made difficult by likely genetic heterogeneity, environmental contributions, and multiple etiologies. However, the mode of inheritance of PCOS has not been firmly established. Although some studies support a single dominant gene with high penetrance,[20–22] others do not.[23] A study was conducted by Urbanek et al[24] on a carefully chosen collection of 37 candidate genes for linkage and association with PCOS or hyperandrogenemia in data from 150 families. The authors found that strongest evidence for linkage was with the follistatin gene.

Strong evidence for a link between the follistatin gene and PCOS has been recently found in a well-designed large-scale study. The pathway in which follistatin gene is involved is shown in Figure 1 as in KEGG database: Follistatin binds to activin

and affects its functions, for example, stimulation of follicle-stimulating hormone (FSH) synthesis and secretion. Thus, it may play a role in the functional impairment of the FSH–granulosa cell axis in PCOS. Several genes involved in the biosynthesis of androgen, and action of insulin and gonadotrophin have been examined as candidate genes for PCOS. These genes include those for the cholesterol side-chain cleavage enzyme (*CYP11A*), 17-hydroxylase/17,20-lyase (*CYP17*), insulin, insulin receptor, and *LH*. Two of them, *CYP11A* and the insulin gene, and variable number of tandem repeats, have been proposed as predisposing genetic factors contributing to PCOS. However, neither of them have been widely accepted as a major cause for this syndrome. Recently, in a well-designed large-scale study, Urbanek et al[24] tested for linkage and association between 37 candidate genes including those previously studied and PCOS. These genes were carefully selected from those involved in the action of androgen, gonadotrophin, and insulin, and also the regulation of obesity and energy. They found evidence for linkage only with the *CYP11A* and follistatin genes. However, only the linkage with follistatin gene remained significant after correction for multiple testing. Nevertheless, this study apparently confirmed previous findings which demonstrated the linkage of *CYP11A* gene with PCOS.

Follistatin is a single-chain glycosylated polypeptide that can bind to activin with high affinity and neutralize its biological action of stimulating the secretion of FSH and increasing *FSHβ* mRNA levels and may therefore, arrest folliculogenesis. Indeed, overexpression of follistatin in the transgenic mice resulted in the suppression of both serum concentrations of FSH and ovarian folliculogenesis, similar to the clinical features commonly found in PCOS patients. Follistatin is expressed in numerous tissues including the ovary, pituitary, adrenal cortex, and pancreas. The human cDNAs encoding follistatin have been cloned. There is a single follistatin gene that can generate two mature mRNA transcripts by alternative splicing, thus encoding proteins of 315 (FS-315) and 288 (FS-288) amino acid residues, respectively. The FS-288, a carboxy-truncated variant with increased biological potency, was found to bind strongly to heparin sulphate proteoglycans of the cell membrane, whereas FS-315 had little or no such binding affinity. Furthermore, in the anterior pituitary cells, FS-288 was more potent in suppressing FSH release. This cell-associated protein was also found to accelerate the uptake of activin into pituitary cells, leading to an increase in its degradation by lysosomal enzymes and thus playing a role in the activin clearance system.

In view of this, it was thought that it would be worthwhile to explore the role of follistatin gene and its linkage with PCOS. The objectives of the work are retrieval of follistatin
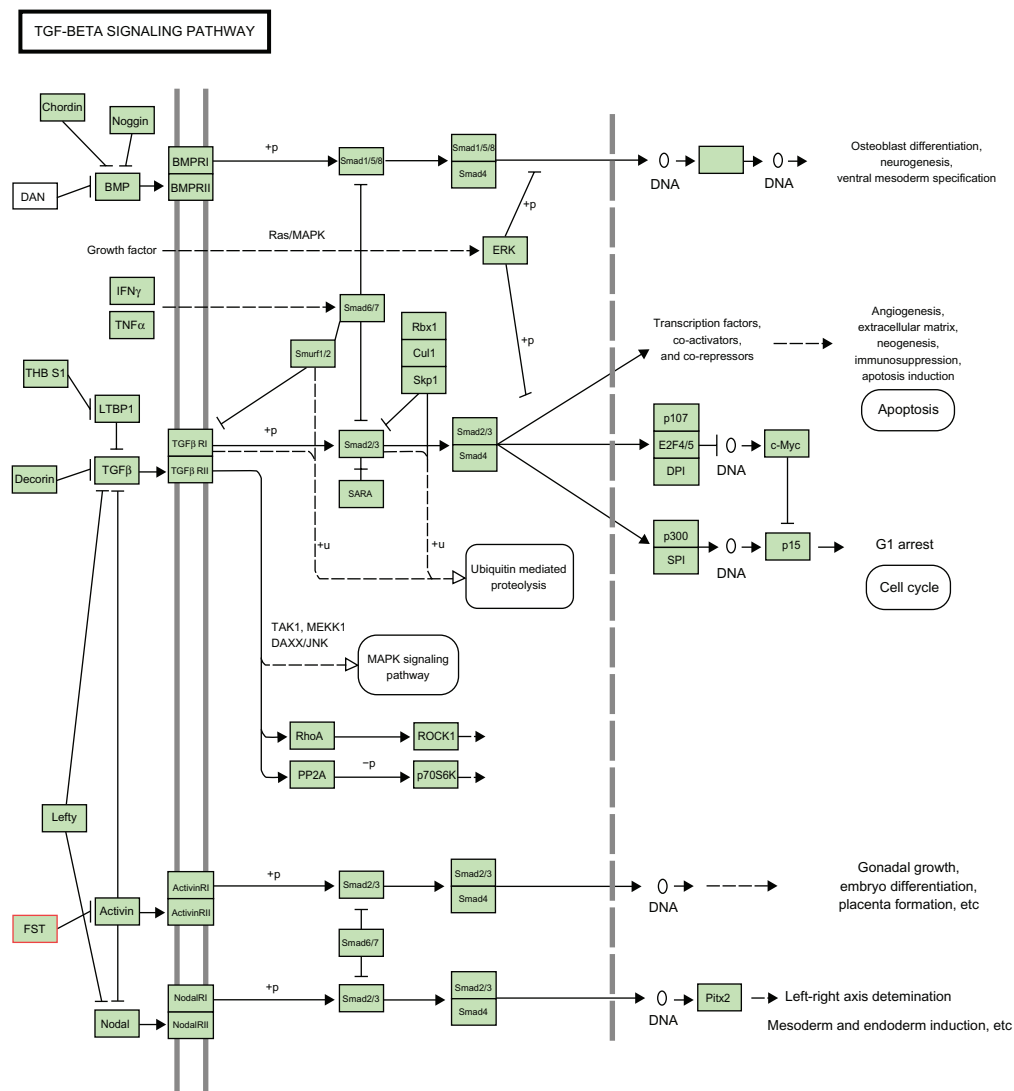
Figure 1 The pathway in which follistatin gene is involved is shown below as in KEGG database.

gene and protein, analysis of follistatin gene and protein, and SNP studies on follistatin to study SNP association with PCOS. Hence, in this work, SNPs in the *FST* gene were studied to determine if genetic variation is associated with susceptibility to PCOS or key phenotypic features of PCOS patients.

Materials and methods

The various methods, databases, tools, and software used for carrying out this project and their details are given below.

Sequence retrieval

1. Nucleotide – NCBI: The nucleotide sequence of human follistatin was retrieved from NCBI database using Entrez.
2. Protein – NCBI: The protein sequence of human follistatin was retrieved from NCBI database using Entrez.

3. Protein – Swiss-Prot: The follistatin protein of human was retrieved from Swiss-Prot database.

Sequence analysis

1. Protparam: The protein sequence of follistatin was given to Protparam search and analyzed.
2. Pfam: The protein sequence of follistatin was given to Pfam search and analyzed for the protein family.

SNP analysis

The follistatin gene and protein was subjected to various SNP analyses.

1. NCBI – dbSNP: The SNPs of human follistatin were retrieved from database. The list of NCBI SNP entries was obtained from genecard link.

2. SIFT (Sorting Intolerant from Tolerant): The follistatin protein sequences were submitted as input for the SIFT tool.

3. Single Amino Acid Polymorphism (SAAP): The SAAP is to link SNP to phenotype alterations. In this tool, the UniProt id P19883 was given in the search box.

4. nsSNP Analyzer: The FASTA format of protein was given to the nsSNP Analyzer tool and the variation E152Q and C239S were given as input in the SNP data field.

5. PANTHER – Evolutionary analysis of coding SNPs: The protein sequence was given as input and analyzed for the E152Q and C239S SNPs. Top of Form Estimates the likelihood of a particular nonsynonymous (amino acid changing) coding SNP to cause a functional impact on the protein. It calculates the substitution position-specific evolutionary conservation (subPSEC) score based on an alignment of evolutionarily related proteins, as described by Thomas et al[25] and Thomas and Kejariwal.[26] Multiple sequence alignment (MSA) position to view the column in the MSA where the substitution occurs was also carried out.

6. SNPs3D: SNPs3D is a Web site which assigns molecular functional effects of nonsynonymous SNPs based on structure and sequence analysis.

Results and discussion
Sequence retrieval

1. Nucleotide – NCBI: The nucleotide sequence of human follistatin was retrieved from NCBI database using Entrez. The accession number is BC004107 and there are totally 1326 base-pairs in the mRNA.

    The sequence in FASTA format is given below:
    >gi|33871153|gb|BC004107.2| Homo sapiens follistatin, mRNA (cDNA clone MGC:10663 IMAGE:3688745), complete cds

2. Protein – NCBI: The protein sequence of human follistatin was retrieved from NCBI database using Entrez. The accession number is AAH04107 and there are totally 344 amino acids in the protein.

    The protein in FASTA format is given below:
    >gi|13278648|gb|AAH04107.1| follistatin [Homo sapiens]

3. Protein – SWISSPROT: The follistatin protein of human was retrieved from SWISSPROT database and the accession number is P19883. There are 344 amino acids and the molecular weight is 38,007 Da. Out of the many genes follistatin was considered as a main candidate gene for PCOS. So, this gene was selected for study.

Follistatin was considered a candidate for the following reasons. It is an activin-binding protein that neutralizes the biological activity of activin in vivo and in vitro and is expressed in multiple tissues, including the ovary, pituitary, adrenal cortex, and pancreas. Activin, a member of the transforming growth factor-ß superfamily, modulates the production of androgens by ovarian thecal cells, the development of ovarian follicles, and the secretion of FSH by the pituitary and insulin by pancreatic ß-cells. Because follistatin inhibits the activity of activin, altered follistatin activity would be expected to affect follicular development, ovarian androgen production, pituitary FSH secretion, and insulin release. All these processes have been shown to be perturbed in PCOS.[27,28]

Sequence analysis

1. Protparam: The protein sequence of follistatin was given to Protparam search and analyzed.

2. Pfam: The protein sequence of follistatin was given to Pfam search and analyzed for the protein family.

    FST_HUMAN (P19883): DE follistatin precursor (FS) (activin-binding protein). The protparam analysis showed that the isoelectric point is 5.53 and the aliphatic index is 61.25. The hydropathicity is −0.490. The domains in FST protein are Pfam-B 5005 domain from 1 to 92, EGF-like subdomain from 93 to 116, Kazal 1 domain occurring in three places, namely, 118–164, 192–239, and 270–316.

    Family: Kazal_1 (PF00050): Alignment also includes a single domain from transporters in the OATP/PGT family (Swiss: P46721).

    Interpro entry IPR002350: This family of Kazal inhibitors belongs to MEROPS inhibitor family I1, clan IA. They inhibit serine peptidases of the S1 family. These proteins contain between 1 and 7 Kazal-type inhibitor repeats. The structure of the Kazal repeat includes a large quantity of extended chain, two short α-helices and a three-stranded antiparallel β sheet. The inhibitor makes 11 contacts with its enzyme substrate: unusually, 8 of these important residues are hypervariable.

    Altering the enzyme-contact residues, and especially that of the active site bond, affects the strength of inhibition and specificity of the inhibitor for particular serine proteases. The presence of this Pfam domain is usually indicative of serine protease inhibitors; however, Kazal-like domains are also seen in the extracellular part of agrins which are not known to be proteinase inhibitors.

Clan

This family is a member of clan Kazal (CL0005), which contains the following two members: Kazal_1 and Kazal_2.

Family: EGF-like_subdom (PF09120): Members of this family are reminiscent of EGF-like modules, as indicated by an identical disulphide connectivity. This subdomain comprises a stretch of residues in an extended conformation held in place by two disulphide linkages to two β strands. In follistatin, it is followed by a Kazal-like sequence: both are required for heparan sulphate binding.

Interpro entry IPR015204: This is an EGF-like domain, as indicated by an identical disulphide connectivity, which resembles the N-terminal domain of follistatin. It comprises a stretch of residues in an extended conformation held in place by two disulphide linkages to two β strands. In follistatin, it is followed by a Kazal-like sequence: both are required for heparan sulphate binding.

AB	Genomic Data				Transcription Data				Allele Frequencies				
	SNP ID	Valid	Chr 5 pos	Sequence	Recs	AA Chg	Type	More	Recs	Allele freq	Pop	Total sample	More
	-	1st	-	--	--	--	2nd	--	--	-	-	-	--
AB	rs11745088[1,2]	C	52815267(+)	GTAAAG/CAG CAG	2	Q/E	ns[1]	🔍	4	(pie)	EU EA WA	406	🔍
AB	rs11746136[1,2]	C,F	52816786(+)	CTACTG/AGA AGT	2	L/L	syn[1] ese[3]	🔍	2	(pie)	NA	182	🔍
--	rs722910[1,2]	C,F,A	52817354(-)	AAAAAT/ATT TTT	2	--	int[1] utr[1]	🔍	1	(pie)	MN	184	🔍
AB	rs3797297[1,2]	C,F	52813413(-)	TCATAC/ACT AGC	2	--	int[1]	🔍	5	(pie)	EA EU WA	1920	🔍
AB	rs3756498[1,2]	C,F	52813967(-)	GAGAAC/TGA GAG	2	--	int[1]	🔍	5	(pie)	EA EU WA	1902	🔍
--	rs3797296[1,2]	C,F	52813478(-)	AGTAAT/CGC CTT	2	--	int[1]	🔍	5	(pie)	EA EU WA	1882	🔍
AB	rs12153205[1,2]	F	52816343(+)	ACTAAT/CGT AGT	2	--	int[1]	🔍	4	(pie)	EU EA WA	418	🔍
AB	rs12152850[1,2]	F	52816110(+)	TATTCC/TGT CTT	2	--	int[1]	🔍	4	(pie)	EU EA WA	416	🔍
AB	rs1423560[1,2]	C,F	52811519(+)	TGGGGG/TTG GCT	2	--	loc[1]	🔍	1	(pie)	MN	184	🔍
AB	rs8434[1,2]	--	52817573(-)	GTCTTC/AAT CTT	2	*/E	ns[1] utr[1]	🔍	4	(pie)	EU EA WA	396	🔍
--	rs1127761[1,2]	--	52815876(+)	AAGTGA/TAT CAA	2	C/*	ns[1]	🔍	0	--	--	--	--
--	rs1127760[1,2]	--	52815874(+)	GAAAGA/TGT ATC	2	C/S	ns[1]	🔍	0	--	--	--	--
--	rs1062809[1,2]	--	52815848(+)	CAGATA/C/G TATTG	4	S/Y S/C	ns[1] ese[3]	🔍	0	--	--	--	--
--	rs1044310[1,2]	--	52815847(+)	GCAGAT/GCT ATT	2	A/S	ns[1]	🔍	1	(pie)	MN	184	🔍
AB	rs15208[1,2]	--	52817600(-)	AGAAGA/GTA TAG	2	S/P	ns[1] utr[1]	🔍	1	(pie)	MN	184	🔍
AB	rs3203788[1,2]	--	52815839(+)	CCTGCT/AGG GCA	2	Q/L	ns[1] ese[3]	🔍	1	(pie)	MN	184	🔍
AB	rs1127759[1,2]	--	52815867(+)	TATGAA/GGG AAA	2	E/E	syn[1]	🔍	0	--	--	--	--

(*Continued*)

--	rs3203789 [1,2]	--	5281584 6(+)	GGCAG**A/G**TC TAT	2	R/R	syn[1]	🔍	1	●	MN	184	🔍
--	rs5867883 [1,2]	--	5281767 0(+)	aaaaa-/**AAAA** aa aaa	2	--	utr[1]	🔍	0	-.-	-.-	-.-	==
--	rs3083659 [1,2]	--	5281767 7(+)	aaaaa-/**AAAA** ga aaa	2	--	utr[1]	🔍	0	-.-	-.-	-.-	==
--	rs5867882 [1,2]	--	5281766 0(+)	TGTGC-/**AAAA** aa aaa	2	--	utr[1]	🔍	0	---	---	---	--
--	rs3477064 6[1,2]	--	5281714 7(+)	TTTTT-/**T**TTCCA	2	--	int[1]	🔍	0	-.-	-.-	-.-	--
--	rs1265267 3[1,2]	--	5281262 7(+)	ATGAC**A/C**CT TTA	2	--	int[1]	🔍	0	-.-	-.-	-.-	--
--	rs1007620 5[1,2]	--	5281284 6(+)	CACAT**C/T**TT TTG	2	--	int[1]	🔍	0	---	---	---	--
--	rs4127030 1[1,2]	--	5281590 0(+)	TTGAG**C/G**AA GGA	2	--	int[1]	🔍	0	-.-	-.-	-.-	==
--	rs2070858 [1,2]	--	5281434 0(+)	TCTCA**C/T**TT CCC	2	--	int[1]	🔍	0	---	---	---	--
--	rs7709554 [1,2]	--	5281288 9(+)	AGAAA**A/C**CC TGG	2	--	int[1] trp[3]	🔍	0	-.-	-.-	-.-	==
--	rs3395157 5[1,2]	--	5281714 0(+)	TTTTT-/**TT**TTTTT	2	--	int[1]	🔍	0	-.-	-.-	-.-	--
--	rs2070859 [1,2]	--	5281437 5(+)	atccc**C/T**GC CGG	2	--	int[1]	🔍	0	-.-	-.-	-.-	==
🅰🅱	rs3806901 [1,2]	--	5281158 0(+)	CCCAG**A/G**CT GCG	2	--	loc[1]	🔍	0	-.-	-.-	-.-	==
--	rs3566484 0[1,2]	--	5281802 2(+)	TTTCC-/**C**TTGTA	2	--	loc[1]	🔍	0	-.-	-.-	-.-	==

Figure 2 Thirty-one NCBI SNPs in *FST* format.

Clan

This family is a member of clan EGF (CL0001), which contains the following eight members: EGF, EGF-like_subdom, EGF_2, EGF_alliinase, EGF_CA, FOLN, Laminin_EGF, and Tme5_EGF_like.

SNP analysis

The follistatin gene and protein were subjected to various SNP analyses.

1. NCBI – dbSNP: The SNPs of human follistatin were retrieved from database. The list of NCBI SNP entries was obtained from genecard link. There are 31 SNPs for this gene and some are nonsynonymous, some occur in intron region, and some in untranslated region (Figure 2). Two nonsynonymous SNPs, namely, rs11745088 and rs1127760, are taken for analysis. In the SNP rs11745088, the change is E152Q. That is E in position

152 is changed to Q. Likewise, in rs1127760 the change is C239S.

Recently, in a well-designed large-scale study, Urbanek et al[24] tested for linkage and association between 37 candidate genes including those previously studied and PCOS. These genes were carefully selected from those involved in the action of androgen, gonadotrophin, and insulin, and also the regulation of obesity and energy. They found evidence for linkage only with the *CYP11A* and follistatin genes. However, only the linkage with follistatin gene remained significant after correction for multiple testing.

Likewise, studies of Jones et al[28] in 173 Caucasian women suggest that polymorphism in the follistatin gene is associated with key androgenic phenotypes of PCOS. In their observation, the SNP rs3797297, located in intron 1 of the *FST* gene, was significantly associated with both

FAI and SHBG. In both instances, post hoc analysis revealed that the main effect was due to subjects who were homozygous for rare A-allele of the polymorphism. The data for rs3797297 did not appear to be consistent with a codominant mode of expression at this locus. In subjects with two copies of the polymorphic allele at this locus, the FAI was well above the reference range, representing more severe hyperandrogenemia than in those women with a copy of the ancestral allele. Thus, a recessive effect appeared more likely. In the case of the SNP rs11745088, which was significantly associated with DHEA-S level, only two heterozygotes were identified for whom DHEA-S was available, with both these individuals having elevated DHEA-S. Although their study was larger than many molecular genetics studies of PCOS published to date, the power of the study remains relatively modest.

2. SIFT: The follistatin protein sequences were submitted as input for the SIFT tool. Amino acids that were not corresponding to its protein were indicated as 'not tolerated' means that the change might cause some deleterious effects. Forty-three sequences were selected to be closely related to our query sequence. At position 25R in the query sequence, 81% of the sequences have an amino acid appearing at this position. M, e, I, y, v, l, p, a, g, T, s, e, H, k, N, Q, D, R are predicted as tolerated and are observed in the alignment (capitalized). w and c are predicted to be deleterious because they have normalized probabilities <0.05 and none of these appear in the alignment (small letters). Amino acids

are color coded: nonpolar, uncharged polar, basic, and acidic.

3. SAAP: SIFT showed positions of amino acids and the single letter code of amino acids that can be tolerated or deleterious for each position. The SAAP is to link SNP to phenotype alterations. In this tool, the UniProt id P19883 was given in the search box. The SAAP is to link SNP to phenotypic changes. There were six SNP results and each result had links to it. The dbSNP id, primary database id, and the type of mutation whether silent and if occurring in coding region are given phenotype alterations.

For 28 of the 37 candidate genes analyzed by Urbanek et al,[24] there is at least one polymorphic marker within 1 centiMorgan (cM) of the candidate gene, and for the remainder, the markers are 1–4 cM from the candidate gene. This proximity improves the power of the study because recombination is likely to be minimal.

4. nsSNP Analyzer: The FASTA format of protein was given to the nsSNP Analyzer tool and the variation E152Q and C239S was given as input in the SNP data field. E152Q change is neutral and the C239S causes disease.

5. PANTHER – Evolutionary analysis of coding SNPs: The protein sequence was given as input and analyzed for the E152Q and C239S SNPs. The probability that a given variant will cause a deleterious effect on protein function is estimated by $P_{deleterious}$ such that a subPSEC score of -3 corresponds to a $P_{deleterious}$ of 0.5.
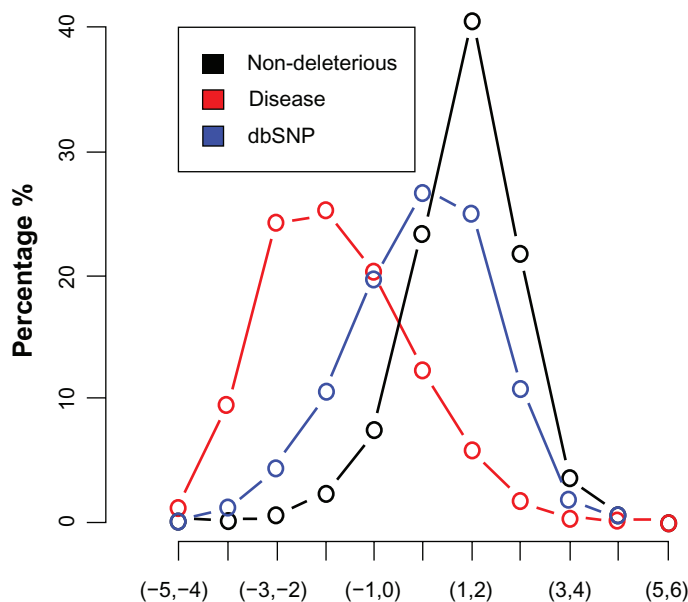

Figure 3 Histogram of SVM profile.

6. SNPS3D: The disease (HGMD) and nondeleterious dataset were used to train and test the model. '0' was used as the threshold of classification. A SNP with a negative support vector machine (SVM) score is classified as damaging SNP. A SNP with a positive SVM score is classified as neutral SNP. A higher score (±) relates a higher confidence in prediction. A score >0.5 is considered to be significant (Figure 3). The results showed that *FST* gene SNPs are linked to PCOS coming under the disease class of metabolic disorders.

Activin, a member of the transforming growth factor-b superfamily, and follistatin are expressed in several tissues, such as ovary, pituitary, adrenal cortex, and pancreas. This similar pattern of tissue distribution suggests that follistatin is likely to be an auto-/paracrine regulator of the hormonal effects of activin within the tissues. In particular, activin promotes ovarian follicular development, enhances LH binding sites and progesterone production, playing a role in preventing premature luteinization of the ovarian follicles, inhibits thecal cell androgen production, and increases pituitary FSH secretion and insulin secretion by pancreatic b-cells. An increase in circulating level or biological activity of follistatin may, therefore, cause follicular development to be arrested, the process of follicular luteinization or atresia to be favored, ovarian androgen production to be increased, FSH circulating levels to be reduced, and insulin release to be impaired. These effects are all characteristic features of PCOS.[29]

Follistatin is a single-chain glycoprotein that primarily acts to regulate the activity of activin, which is responsible for ovarian follicular development, inhibition of theca cell androgen production and increases in both pituitary FSH secretion and pancreatic insulin secretion.[30] Primarily synthesized in the granulosa cells of the ovarian antral follicle, follistatin mRNA increases within the dominant follicle during development and declines during the atretic process.[31] Follistatin acts to suppress aromatase activity in the granulosa cell and also LH-stimulated progesterone release from thecal cells.[32] The overexpression of follistatin in mice has been shown to result in arrested ovarian follicular development and reduced levels of FSH, both key phenotypes of PCOS.[27]

The list of intronic and synonymous SNPs, with their nucleotide position, amino acid change information, and dbSNP link, is provided for further analysis. Researchers were unable to detect any mutation of the activating or inhibiting type in the entire coding region of follistatin gene in 64 patients with PCOS. Therefore, mutations in the coding regions of the follistatin gene may not be a common cause of PCOS in the population studied. However, it is possible that mutations may reside in the regulatory region of the gene, which should be screened once its sequence is known. Furthermore, it would be of great interest to investigate the presence of mutations in PCOS patients in other ethnic populations, especially of European origin, as it was in this population that the linkage between the follistatin gene and PCOS was established.[33]

Strong evidence for a link between the follistatin gene and PCOS has been recently found in a well-designed large-scale study. Odunsi and Kidd[34] are of the opinion that over-expression of follistatin will be expected to lead to increased ovarian androgen production and reduction in circulating FSH levels, which are features of PCOS. The genetic association database results show that *FST* gene SNPs are linked to PCOS coming under the disease class of metabolic disorders. The list of intronic and synonymous SNPs, with their nucleotide position, amino acid change information, and dbSNP link, is provided for further analysis.

The authors thus conclude that there is an association between FST gene SNPs and polycystic ovarian syndrome.

Disclosure

The authors report no conflict of interest in this work.

References

1. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, Azziz R. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab*. 1998;83:3078–3082.
2. Legro RS, Kunselman AR, Dodson WC, Dunaif A. Prevalence and predictors of risk for type 2 diabetes and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999;84:165–169.
3. Cooper HE, Spellacy WN, Prem KA, Cohen WD. Hereditary factors in the Stein-Leventhal syndrome. *Am J Obstet Gynecol*. 1968;100:371–387.
4. Givens JR. Familial polycystic ovarian disease. *Endocrinol Metab Clin North Am*. 1988;17:771–783.
5. Legro RS, Spielman R, Urbanek M, Driscoll D, Strauss JF, Dunaif A. Phenotype and genotype in polycystic ovary syndrome. *Recent Prog Horm Res*. 1998;53:217–256.
6. Legro RS, Strauss JF. Molecular progress in infertility: polycystic ovary syndrome. *Fertil Steril*. 2002;78:569–576.
7. Carey AH, Waterworth D, Patel K, et al. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum Mol Genet*. 1994;3:1873–1876.
8. Gharani N, Waterworth DM, Batty S, et al. Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum Mol Genet*. 1997;6:397–402.
9. Franks S. Polycystic ovary syndrome. *N Engl J Med*. 1995;333:853–861.
10. Ciaraldi TP, el-Roeiy A, Madar Z, Reichert D, Olefsky JM, Yen SS. Cellular mechanisms of insulin resistance in polycystic ovarian syndrome. *J Clin Endocrinol Metab*. 1992;75:577–583.

11. Dunaif A, Segal KR, Shelley DR, Green G, Dobrjansky A, Licholai T. Evidence for distinctive and intrinsic defects in insulin action in polycystic ovary syndrome. *Diabetes*. 1992;41:1257–1266.

12. Dunaif A, Xia J, Book C-B, Schenker E, Tang Z. Excessive insulin receptor serine phosphorylation in cultured fibroblasts and in skeletal muscle: a potential mechanism for insulin resistance in the polycystic ovary syndrome. *J Clin Invest*. 1995;96:801–810.

13. Kiddy DS, Hamilton-Fairley D, Bush A, et al. Improvement in endocrine and ovarian function during dietary treatment of obese women with polycystic ovary syndrome. *Clin Endocrinol*. 1992;36:105–111.

14. O'Rahilly S, Choi WH, Patel P, Turner RC, Flier JS, Moller DE. Detection of mutations in the insulin receptor gene in non-insulin dependent diabetic patients by analysis of single-stranded conformation polymorphisms. *Diabetes*. 1991;40:777–782.

15. Taylor SI, Cama A, Accili D, et al. Mutations in the insulin receptor gene. *Endocr Rev*. 1992;13:566–595.

16. Conway GS, Avey C, Rumsby G. Genetics: the tyrosine kinase domain of the insulin receptor gene is normal in women with hyperinsulinaemia and polycystic ovary syndrome. *Hum Reprod*. 1994;9 Suppl 9: 1681–1683.

17. Krook A, Kumar S, Laing I, Boulton AJ, Wass JA, O'Rahilly S. Molecular scanning of the insulin receptor gene in syndromes of insulin resistance. *Diabetes*. 1994;43:357–368.

18. Sorbara LR, Tang Z, Cama A, et al. Absence of insulin receptor gene mutations in three women with the polycystic ovary syndrome. *Metabolism*. 1994;43:1568–1574.

19. Krook A, O'Rahilly S. Mutant insulin receptors in syndromes of insulin resistance. *Baillieres Clin Endocrinol Metab*. 1996;10:97–122.

20. Carey AH, Chan KL, Short F, White D, Williamson R, Franks S. Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin Endocrinol*. 1993;38:653–658.

21. Legro RS, Driscoll D, Strauss JF, Fox J, Dunaif A. Evidence for a genetic basis for hyperandrogenemia in the polycystic ovary syndrome. *Proc Natl Acad Sci U S A*. 1998;95:14956–14960.

22. Govind A, Obhrai M, Clayton R. Polycystic ovaries are inherited as an autosomal dominant trait: analysis of 29 polycystic ovary syndrome and 10 control families. *J Clin Endocrinol Metab*. 1999;84 Suppl 1:38–43.

23. Jahanfar S, Eden JA, Warren P, Seppala M, Nguyen TV. A twin study of polycystic ovary syndrome. *Fertil Steril*. 1995;63:478–486.

24. Urbanek M, Legro RS, Driscoll DA, et al. Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. *Proc Natl Acad Sci U S A*. 1999;96:8573–8578.

25. Thomas PD, Campbell MC, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13:2129–2141.

26. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A*. 2004;101(43):15398–15403.

27. Urbanek M, Wu X, Vickery KR, et al. Allelic variants of the follistatin gene in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2000; 85 Suppl 12:4455–4461.

28. Jones MR, Wilson SG, Mullin BH, Mead R, Watts GF, Stuckey BGA. Polymorphism of follistatin gene in polycystic ovary syndrome. *Mol Hum Reprod*. 2007;13(4):237–241.

29. Asteria C. Identification of follistatin as a possible trait-causing gene in polycystic ovary syndrome. *Eur J Endocrinol*. 2000;143 Suppl 4: 467–469.

30. Calvo R, Villuendas G, Sancho J, San Millan JL, Escobar-Morreale HF. Role of the follistatin gene in women with polycystic ovary syndrome. *Fertil Steril*. 2001;75:1020–1023.

31. Lin S-Y, Morrison JR, Phillips DJ, de Krester DM. Regulation of ovarian function by the tgf-ß superfamily and follistatin. *Reproduction*. 2003;126:133–148.

32. Phillips D, de Krester DM. Follistatin: a multifunctional regulatory protein. *Front Neuroendocrinol*. 1998;19:287–322.

33. Liao WX, Roy AC, Ng SC. Preliminary investigation of follistatin gene mutations in women with polycystic ovary syndrome. *Mol Hum Reprod*. 2000;6 Suppl 7:587–590.

34. Odunsi K, Kidd KK. A paradigm for finding genes for a complex human trait: polycystic ovary syndrome and follistatin. *Proc Natl Acad Sci U S A*. 1999;96 Suppl 15:8315–8317.