TARTU UNIVERSITY
FACULTY OF BIOLOGY AND GEOGRAPHY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY
DEPARTMENT OF BIOTECHNOLOGY AND BIOMEDICINE

MARINA GRIGOROVA

Worldwide variation of follicle-stimulating hormone beta-subunit gene and its potential association with reproductive success

Master thesis

Supervisor: professor *extraordinarius* Maris Laan, Ph.D.

TARTU
2006

Contents

Abbreviations

cAMP	cyclic adenosine monophosphate
CG	chorionic gonadotropin
CGA	gonadotropine hormone α-subunit gene
DNA	deoxyribonucleic acid
EST	expressed sequence tag sequence
FSH	follicle-stimulating hormone
FSHB	follicle-stimulating hormone beta-subunit protein
FSHB	follicle-stimulating hormone beta-subunit gene
FSHR	follicle-stimulating hormone receptor
GnRH	gonadotropin releasing hormone
GPCR	G protein-coupled receptors
H1, HAP1	haplotype 1
H13, HAP13	haplotype13
HKA	Hudson-Kreitman-Aguade test
HWE	Hardy-Weinberg equilibrium
hCG	human chorionic gonadotropin
kb	thousand basepairs
L	liter
LD	Linkage Disequilibrium
LH	luteinizing hormone
LHR	luteinizing hormone and chorionic gonadotropine receptor
µL	microliter
mRNA	messenger RNA
MAF	minor allele frequency
ng	nanogram
pmol	picomol
PCR	Polymerase Chain Reaction
PCOS	Polycystic Ovarian syndrome
RFLP	Restriction Fragment Length Polymorphism analysis
RNA	Ribonucleic acid
SNP	Single Nucleotide Polymorphism
STP	“short-time-to-pregnancy” women
TSH	thyroid stimulating hormone
U	unit
UTR	Untranslated Region

Introduction

Over the past two decades, a number of genes that regulate the development and function of hypothalamic-pituitary action have been identified. Mutations in these genes are relatively rare, but it is important to study the genetic basis of disorders that are associated with gonadotropin hormones. An appropriate treatment can be provided to an affected individual if the pathological basis of the underlying disorder is known.

Follicle-stimulating hormone belongs to an evolutionarily conserved glycoprotein gonadotropine hormone family and is secreted from the pituitary to regulate reproduction in mammals. FSH is used clinically to treat women with anovulatory infertility, and high-dosage administration of FSH is used to improve disturbed sperm structures as well. Follitropin is a dimeric hormone that consists of α- and β-subunits. The α-subunit is a common for all glycoprotein gonadotropines, while the β-subunit determines the specificity of a hormone. *FSHB* gene coding for follicle-stimulating hormone β-subunit is highly conserved. To date, only eight subjects with inactivating mutations of human *FSHB* gene have been described. All mutations lead to the loss of functional product and infertility.

This resequencing study is the first detailed description of the worldwide variation of human follicle-stimulating hormone beta-subunit gene, revealing two major *FSHB* haplotypes apparently enriched due to balancing selection. Although the functional consequence of these gene variants is still to be determined, I also present preliminary data showing that these *FSHB* haplotypes might be associated with human reproductive success.

1. **Review of literature**

1.1 Gonadotropine hormones

1.1.1 Overview of Gonadotropine Hormones

The gonadotropine hormones belong to the Glycoprotein hormone family which includes the pituitary gland hormones follicle-stimulating hormone or follitropin (FSH), luteinizing hormone or lutropin (LH), and thyroid stimulating hormone (TSH). Primates also produce a placental hormone termed chorionic gonadotropin (CG, in human hCG).

The gonadotropine hormones control the production of male and female gametes and stimulation of the sex steroid hormones by signaling through G-protein coupled transmembrane receptors on the gonads (Bousfield *et al.*, 1994).

1.1.2 Structure of gonadotropine hormones FSH, LH and hCG

The gonadotropines are relatively large proteins (molecular mass, 30-40kDa) consisting of specific β-subunits that are non-covalently bound to a common α-subunit. The three-dimensional structure of gonadotropine hormones is similar and can be presented by the structure of deglycosylated follicle-stimulating hormone (Fig. 1).

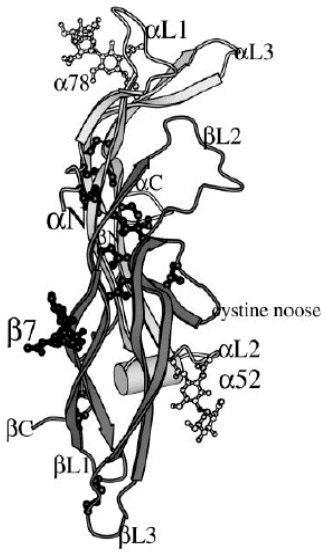


Figure 1. Schematic overview of three-dimensional structure of human follitropin with termini and main loops identified (Fox *et al*., 2001).

In mammals the α-subunit is encoded by a single gene that is expressed in the anterior pituitary gland in every species, in the placenta of primates and equines and in other tissues (Table 1) (Moyle and Campbell, 1996). The α-subunit is one of the first indicators of pituitary cell differentiation during fetal development (Rosenfeld *et al*., 2000). It consists of 116 amino acid residues and is encoded by a single gene, comprising four exons, that is located on chromosome 6q12.21. The LH and FSH β-subunits are each encoded by a separate gene, whereas six genes encode the CG β-subunits of humans. The β-subunit genes are located on different chromosomes: the *LH/hCGB* gene cluster on chromosome 19q13.32, *FSHB* on chromosome 11p13. The *LH/hCGB* gene cluster consists of one *LHB* gene and six *hCGB* genes (Table 1).

The structures of α- and β-subunits are similar. Both subunits contain a cysteine knot, a structure formed by four polypeptide chains and three disulfide bonds. The α-subunit protein contains 10 cysteines, which are involved in intrasubunit disulfide linkages and two N-linked glycosylation sites. The β-subunit differs from the α-subunit by an additional carboxyterminal segment that contains 10^{th}, 11^{th}, and 12^{th} cysteine residues. This loops over the α-subunit to stabilize αβ-heterodimer (Lapthorn *et al.*, 1994).

Crystal structures of human gonadotropine hormones reveal elongated molecules with similar fold for α- and β-chains (Lapthorn *et al*., 1994; Fox *et al*., 2001). The interactions between the glycoprotein hormones and their corresponding receptors are highly selective, with very few cases of cross activity (Themmen and Huhtaniemi, 2000).

Glycosylation of the gonadotropin hormones has been shown to be important in circulatory persistence and clearance, and in bioactivity (Ulloa-Aguirre *et al*., 1999). β-subunit glycosylation has been shown to affect disulfide bond formation and rate of secretion (Feng *et al*., 1995).

Table 1. Overview of structure and function of gonadotropine hormones.

hormone	major action	structure	gene	expression	chromosome location
follicle-stimulating hormone (follitropin, FSH)	♀: stimulation of follicular growth	alpha-subunit	*CGA*	different cell types	6q12.21
	♂: stimulation of Sertoli cell proliferation	beta-subunit	*FSHB*	pituitary	11p13
luteinizing hormone (lutropin, LH)	♀: promotion of the ovulation and luteinization of mature Graafian follicles	alpha-subunit	*CGA*	different cell types	6q12.21
	♂: stimulation of steroidogenesis in thecal and interstitial cells	beta-subunit	*LHB*	pituitary	19q13.32
chorionic gonadotropin (hCG)	♀: prevention of the disintegration of the *corpus luteum* of the ovary and maintain progesterone production	alpha-subunit	*CGA*	different cell types	6q12.21
	♂: -	beta-subunit	*CGB (six* homologous genes)	placenta	19q13.32

1.1.3 Gonadotropins act through gene-specific receptors

The gonadotropine hormones act through specific G-protein-coupled receptors (GPCRs) on target cell surfaces (Dias *et al*., 2002). FSH binds to FSHR, LH and CG both bind to the same receptor, LHR. These receptors belong to a subfamily of GPCRs, which have a rhodopsin-like domain of seven transmembrane helices (7TM) and a large extracellular domain, and bind protein ligands including gonadotropin hormones (Fan and Hendrickson, 2005). The extracellular domains of gonadotropin hormone receptors are responsible for the specificity of ligand binding, whereas the transmembrane domains are responsible for the receptor activation and signal transduction (Dias *et al*., 2002). Important discriminating determinants of receptor binding have been localized to β-chain C-terminal segments that correspond to a unique seat-belt feature of the hormone structures (Lapthorn *et al*., 1994; Fox *et al*., 2001; Keutmann *et al*., 1989). Hormone binding to the ectodomain of gonadotropine hormone receptors prompts changes in the 7TM domain that propagate across the plasma membrane to obtain guanine nucleotide exchange in a heterodimeric G_s protein. Activation of adenylyl cyclase for the production of cAMP follows, thereby initiating a signaling cascade that leads to steroid synthesis (Dias *et al*., 2002).

FSH transmits its signal via the 75 kDa FSH receptor (675 amino acids). The gene encoding the FSH receptor consists of 10 exons. The first 9 exons encode the extracellular domain and the last exon encodes the membrane-spanning region. The crystal structure of FSH bound to hormone-binding domain of the FSH receptor has been solved to reveal that FSH and its receptor interact in a manner that resembles a handclasp (Fig. 2). Ten parallel β-strands of the receptor and additional loops just C-terminal to the β-strands surround and contact the FSH. When FSH interacts with the receptor, the hormone undergoes a series of conformational adjustments and adopts a rigid structure that appears to be required for signaling. As a result of hormonal binding to the ectodomain of the receptor, structural changes occur in the seven membrane-spanning domains that promote guanine nucleotide exchange in associated G_S proteins. There is also evidence that ligand binding causes dimerization of plasma membrane FSH receptors through contacts that are

limited to the cytoplasmic domains and the receptor dimerization contributes to signaling (Dias, 2005; Fan and Hendrickson, 2005).

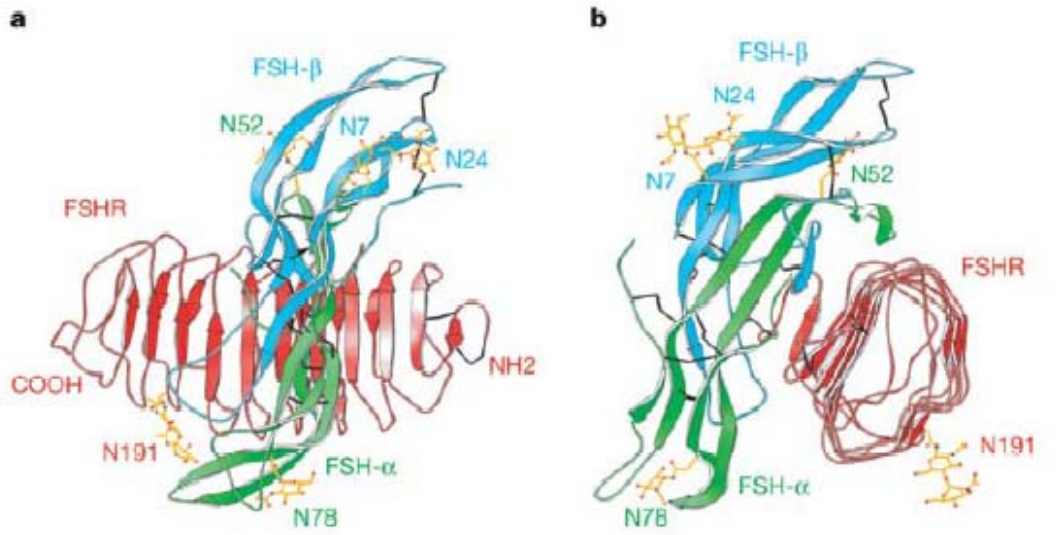


Figure 2. Crystal structure of human FSH bound to $FSHR_{HB}$ (extracellular hormone-binding domain of follicle-stimulating hormone receptor). **a,b** Ribbon diagram of the complex structure shown in two views related by a 90° rotation about the vertical axis. FSH α-chains and β-chains are in green and cyan, respectively. $FSHR_{HB}$ is in red. The N-linked carbohydrates at N52 of FSH-α, N7 and N24 of FSH-β, and N191 of $FSHR_{HB}$ are in yellow. Disulphide bonds are in black. (Fan and Hendrickson, 2005).

1.1.4 Function of gonadotropine hormones

The major role of follitropin in women is stimulation of the development of follicles that have resumed meiosis (Table 1). Binding of FSH to granulosa cell FSH receptors causes the proliferation of granulosa cells, convertion of androgens to estrogens, and production of inhibin (Moyle and Campbell, 1996; Howles, 2000). In the absence of sufficient FSH, follicles fail to develop and ovulation does not occur. During the later stages of follicular maturation, FSH stimulation promotes the formation of the antrum and induction of granulosa cell LH receptors (Hsueh *et al.*, 1984).

Follicle-stimulating hormone is also needed for spermatogenesis (Table 1). The specific role of FSH in testicular function is still somewhat unclear, but functions such as stimulation of Sertoli cell proliferation in the immature testis and maintenance of

qualitatively and quantitatively normal spermatogenesis, through indirect effects mediated by Sertoli cells, have been proposed (Heckert and Griswold, 1991; Plant and Marshall, 2001). The roles of FSH in males in regulating testicular function continue to be debated (Moudgal and Sairam, 1998; Kumar *et al*. 1997).

Luteinizing hormone acts on the ovary to promote the ovulation and luteinization of mature Graafian follicles and to stimulate steroidogenesis in thecal and interstitial cells (Table 1). The actions of LH on thecal cells produce the androgen substrates for follicular estradiol synthesis (Hsueh *et al.*, 1984). LH has an important role in follicular estrogen formation and acts synergistically with FSH. In males LH stimulates Leydig cell androgen production and thereby maintain the endocrine (extratesticular) and paracrine (spermatogenic) effects of androgens.

Chorionic gonadotropin is a gonadotropine closely related to LH in structure and activity that is produced by the placenta of only a few mammals, notably primates and a few equines. The start of human chorionic gonadotropin (hCG) synthesis appears near the time of implantation (Table 1). hCG is produced throughout the pregnancy. Although hCG may have multiple functions during pregnancy, the most important function is to prolong the life of the corpus luteum, in that way permitting the production of progesterone essential to prevent uterine rejection of the developing embryo (Moyle and Campbell, 1996).

1.1.5 Evolution of genes coding for glycoprotein gonadotropine hormones

Glycoprotein gonadotropine hormones have been found in all five vertebrate classes. The amino acid sequences of all α-subunits are highly conserved. Analysis of the sequence conservation of nine mammalian α-subunit genes revealed strict conservation of 30 residues in mature proteins (Fox *et al*., 2001). 10 of these residues are cysteines and all are disulfide-bonded. Of these, three disulfides form the cysteine knot (Fig. 1).

Comparative analysis of glycoprotein hormone subunits showed that many conserved amino acid residues exist among different species. These conserved amino acid residues may be involved in functions that are common to all these subunits, such as peptide folding, formation of the heterodimer, receptor binding or biological activities. Mutations in these conserved residues lead to a loss of a particular function. Differences in the primary sequences of FSH, LH and CG β-subunits cause the ability of the β-subunit to give the biologic activity unique to each hormone (Li and Ford, 1998). The amino acids that have been identified as critical for receptor binding are striking in similarity (Dias *et al*., 1998).

Although the amino acid sequences of subunits are different, the overall structures of the α- and β-subunits are similar (Moyle and Campbell, 1996). On the basis of knowledge of the molecular structure of mammalian pituitary hormones, it was predicted that α- and β-subunits of the glycoprotein hormones evolved from a common ancestral gene through gene duplications. The first duplication produced the α-subunit and the ancestor of all β-subunits. Duplication of the ancestor gene of the β-subunit subsequently generated the LH beta-subunit gene and the ancestor for the *TSHB* (thyroid-stimulating hormone, pituitary glycoprotein hormone) and *FSHB* groups, which eventually gave rise to the genes for these latter β-subunits (Fig. 3) (Li and Ford, 1998).

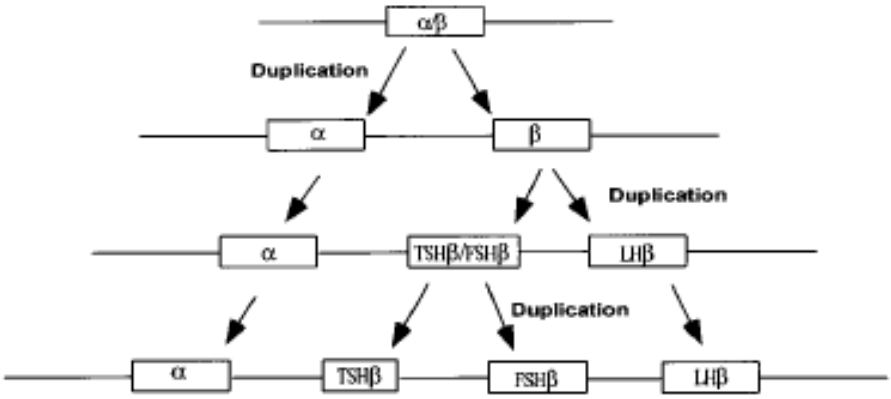


Figure 3. Proposed model for the evolution of the gonadotropin glycoprotein hormone family (Li and Ford, 1998).

1.2 Follicle-stimulating hormone

1.2.1 Structure of the FSHB gene

The structure of the human *FSHB* gene was first described in detail by Jameson *et al*. (1988). Thus far, *FSHB* gene has been isolated and sequenced for human, bovine, mouse, rat, frog, zebrafish, porcine, gray short-tailed opossum, sheep, and chicken organisms. The human FSH β-subunit gene, which codes for a 129-aminoacid preprotein and is located on chromosome 11p13, spans over 4.2 kb and is composed of three exons, the first of which encodes a 5' untranslated sequence, exon 2 the 18-aminoacid signal sequence and residues 1-35, and exon 3 encodes residues 36-111 of mature peptide. The 18-aminoacid signal sequence is hydrophobic and contains four cysteine residues. Although the signal sequences vary substantially, the amino acid sequences in the mature proteins have 87-94% identity among the various species.

Phylogenetic comparisons of mammalian *FSHB* genes revealed five regions of highly conserved sequence homology: the proximal 5' promoter region, exon 2, the 5' translated region of exon 3, and two regions at the 3' untranslated end of exon 3 that include putative polyadenylation and transcriptional termination signals (Kumar *et al*., 2006).

There are four forms of *FSHB* mRNAs that arise by a combination of alternate splicing and polyadenylation (Fig. 4). *FSHB* transcripts with 5'-untranslated tracts that differ in length by 30 nucleotides are produced by using an alternate splice donor site within the first exon. There are multiple consensus polyadenylation signals (AAUAAA), one proximal coinciding with the stop codon in exon 3 and a cluster of three additional sites about 1-1.2 kb downstream of the stop codon resulting in an unusually long 3'-untranslated region (UTR) (Fig. 4). Analyses of mRNA size by Northern blots, alignments of genomic and expressed sequence tag (EST) sequences, and phylogenetic comparisons of multiple mammalian *FSHB* gene 3' UTRs suggest that the most distal AAUAAA site at nucleotides +1218 to +1223 is the predominantly utilized signal (Jameson *et al*., 1988; Brockman *et al*., 2005). Alternate splicing and polyadenylation appear to be regulated independently. A functional role for such processes could potentially involve RNA stability, intracellular transport of RNA, or translational efficiency (Jameson *et al*., 1988).

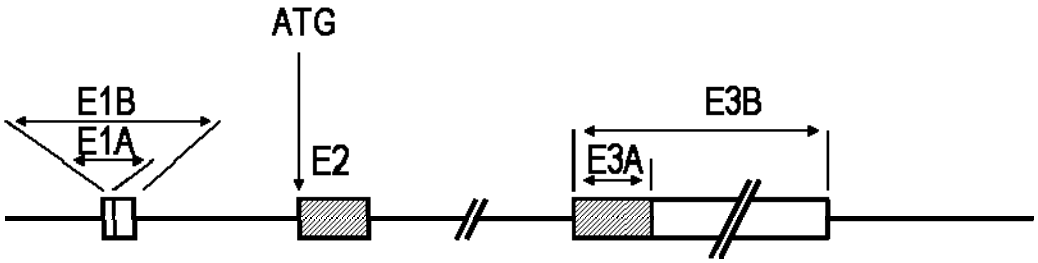


Figure 4. Schematic representation of human *FSHB* gene. Structure is drawn to an approximate scale. Exons (E1-E3) are indicated by boxes. Coding and non-coding areas are denoted by hatched and open boxes, respectively.

1.2.2 Regulation of FSH

The control of FSH synthesis is complex and involves interplay between the gonads, pituitary and hypothalamus. FSH acts on ovaries and testes to regulate folliculogenesis, ovulation, spermatogenesis and steroidogenesis. Gonadal steroids and peptides, in turn, act at the hypothalamus and/or pituitary to regulate FSH synthesis and secretion either positively or negatively.

The synthesis and secretion of FSH are regulated by the hypothalamic decapeptide gonadotropin-releasing hormone (GnRH), which is secreted in a pulsatile manner (Miller *et al*., 2002). Some investigators have immunobioneutralized GnRH in animals to show that ~50-67% of serum FSH depends on GnRH (Culler and Negro-Vilar, 1986). Studies with *hpg* mice that lack GnRH show that serum follicle-stimulating hormone level is 60% lower in the female and even lower (87%) in *hpg* males (Mason *et al*., 1986).

Gonadal peptides activin, inhibin, and follistatin also alter FSH activity on the ovary. Inhibin and activin are structurally related members of the TGF-β supefamily of ligands that initiate their actions by binding to a complex of transmembrane serine and threonine kinase receptors. At the level of pituitary, activin stimulates and inhibin inhibits the release of FSH (Suszko *et al*., 2003; Rivier *et al*., 1985). *FSHB* mRNA and circulating FSH are reduced in mice deficient in the activin receptor (Weiss *et al*., 1993; Matzuk *et al*., 1995; Kumar *et al*., 2003). Follistatin is a known activin binding protein and inhibitor of activin action (Krummen *et al*., 1993). It is co-expressed with activin in essentially all tissues (Kawakami *et al*., 2001). Activin and follistatin are produced in gonadotrophs and pituitary folliculostellate cells, and perhaps by somatotrophs, lactotrophs and thyrotrophs. Activin and follistatin can be induced by GnRH, depending on the pulse rate and amplitude (Dalkin *et al*., 1999).

In addition to gonadal peptides, gonadal steroids are known to regulate the *FSHB* gene and thereby FSH synthesis (Burger *et al*., 2004). Estrogen is thought to mediate its repressive effects indirectly via modulation of GnRH secretion. Androgen administration results in species-specific effects on direct pituitary regulation of *FSHB* transcription. Conserved androgen response elements were identified in the ovine *FSHB* promoter (Spady *et al*., 2004).

1.2.3 FSH, FSHB mutations and fertility

Although an increasing number of factors are associated with inherited disorders of gonadotropin release function, relatively few patients have been identified with mutations in these genes. Mutations of genes concerned with hypothalamic-pituitary-gonadal function, due to their critical role in the development and regulation of

reproductive functions, are very rare and therefore not of major concern within the clinical practice of infertility treatment.

Below I discuss evidences for the essential role of FSH in fertility.

1.2.3.1 Evidence from inactivating FSHB mutations

A total of eight subjects with different inactivating mutations of the *FSHB* gene have so far been described in the literature (Table 2).

Homozygous or compound heterozygous mutations in the *FSHB* gene have been reported in five women who presented delayed puberty, absent breast development, and primary amenorrhea (Val61X; Val61X/Cys51Gly, Tyr79X) (Themmen and Huhtaniemi, 2000; Layman *et al*., 1997, Layman *et al*., 2002). In these cases, serum FSH was undetectable, and serum LH was elevated. The mutations reported in these patients affect different regions of the molecule (Lapthorn *et al*., 1994). Val61X causes a deletion of the C-terminus, including the so-called "seat-belt" region (residues 90-110) that is essential for heterodimer formation (Fig. 5). C51 is involved in the "cysteine-knot" motif that arranges the core of the protein subunit and establishes the remaining folds (Fig. 5).

To date, three male patients with FSH β-subunit mutations have been described. Two of these patients, who harbored distinct mutations, presented with small testes, azoospermia, and normal testosterone level. The third male patient presented with absence of pubertal development, low testosterone levels, small and soft testes, and azoospermia (Phillip *et al*., 1998).

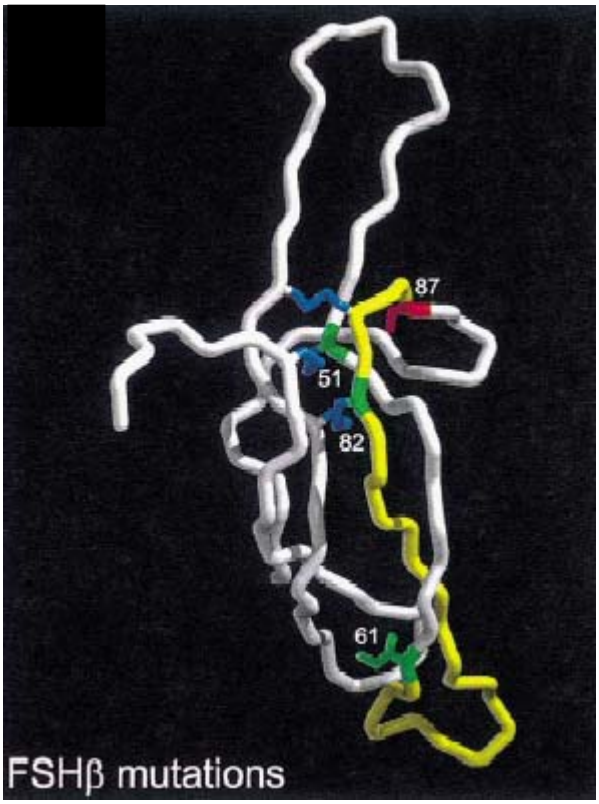


Figure 5. Locations of reported mutations in the FSH β-subunit. Identified mutations in hFSH β-subunit were mapped onto the crystal structure of hCG β-subunit (Lapthorn *et al*., 1994) relative to the positions of the cysteine residues, which are fully conserved among the two β-subunits. Mutated residues are shown in green. Residue numbers (in white) refer to the position in the FSH β-subunit rather than the number of the residue in the hCG structure. The three disulfide bridges involved in the 'cysteine knot' motif are shown in blue. A 2 bp deletion at residue 61 changes the ensuing sequence (colored yellow) up to residue 86 and results in a stop codon at residue 87 (colored red). (Achermann *et al*., 2001).

Table 2. Clinical data of five women and three men with selective FSH deficiency due to FSH β-subunit gene mutations (Berger *et al.* 2005)

Case	CA (y)	Sex/origin	Clinical presentation	Molecular defects
1	27	Female/Italian	Primary amenorrhea, no breast development, eunuchoidal proportions, infertility	Val61X homozygosis
2	15	Female/American	Primary amenorrhea, no breast development	Val61X/Cys51Gly
3	22	Female/Israeli	Primary amenorrhea, partial breast development and eunuchoidal proportions	Val61X homozygosis
4	32	Female/Brazilian	Primary amenorrhea, partial breast development, infertility	Tyr76X homozygosis
5	16	Female/Brazilian	Primary amenorrhea, partial breast development, normal proportions	Tyr76X homozygosis
6	18	Male/Israeli	Delayed puberty, small soft testes, azoospermia	Val61X homozygosis
7	28	Male/Serbian	Infertility, small soft testes, azoospermia	Cys82Arg homozygosis
8	30	Male/Brazilian	Infertility, small testes, azoospermia	Tyr76X homozygosis

CA = chronological age

1.2.3.2 Evidence from transgenic deficient or overexpressing animals

Using gonadotrophin deficient (hpg) mice Singh *et al*. (1995) observed that in the absence of blood FSH, the number of produced germ cells per Sertoli cell was reduced. To study the isolated deficiency of FSH, Kumar *et* al. (1997) deleted most of the coding region of follicle-stimulating hormone beta-subunit gene in embryonic stem cells and subsequently generated *FSHB* knockout mice from these cells. As a result, FSH was absent in these mice. *FSHB* heterozygotes were normal. *FSHB* knockout female mice were infertile and demonstrated decreased ovary size. However, *FSHB* knockout male mice were fertile despite reduced testes size and volume of the seminiferous tubules. Although all stages of spermatogenesis appeared qualitatively normal, sperm number and sperm motility were reduced.

To assess the role of FSH in gonadal growth, differentiation and tumorigenesis, Kumar *et al*. (1999) generated transgenic mice overexpressing human FSH using a mouse metallothionein-1 promoter, achieving hormone levels that exceed levels in postmenopausal women. Male transgenic mice were infertile despite normal testicular development and demonstrated enlarged seminal vesicle due to the high testosterone levels. It suggests that over-production of human FSH does not affect testicular growth and differentiation nor spermatogenesis. FSH-overexpressing female mice demonstrated hemorrhagic, enlarged, cystic ovaries, and no signs of folluculogenesis. These symptoms are typical to human Polycystic Ovarian syndrome (PCOS). All of the transgenic female mice were infertile. In addition, the female animals had urinary tract abnormalities such as enlarged kidneys and urinary bladders, and died by 13 weeks of age. To determine the role of FSH in gonadal tumor development, Kumar *et al*. (1999) generated double-mutant mice that lacked both inhibin and FSH. Most of the double-mutant mice developed slow-growing gonadal tumors.

1.2.3.3 Evidence from experimental stimulation or inhibition of FSH

The study of Matsumoto *et al*. (1986) has shown that blocking endogenous FSH secretion in normal men leads to significant inhibition in sperm production which can be

reversed by exogenous FSH but not testosterone supplementation. More remarkably the spermatozoa of FSH immunized monkeys and men as well as FSH receptor immunized monkeys exhibited a marked reduction in acrosomal glycoprotein content and defective chromatin packaging (Xing *et al*., 2003). Both the above parameters are known to be associated with human male infertility (Cross *et al*., 1986; Evenson *et al*., 1994).

A study by Shetty *et al*. (1996) reported that inmmunoneutralization of follicle-stimulating hormone in both immature and adult rats causes increased apoptosis, with spermatogonia and pachytene spermatocytes being particularly susceptible. These studies all support a role for FSH in cell survival.

2. Aims

The aims of present study were:

1. to survey *FSHB* gene variation in worldwide populations (Estonians, Czech Utah/CEPH, Mandenka, Hans, Koreans) ;
2. to investigate whether the identified *FSHB* gene variants have functional consequence on reproductive success

3. Materials and methods

3.1 Samples

The study has been approved by the Ethics Committee of Human Research of the University Clinic of Tartu, Estonia (permission no. 117/9, 16.06.03).

Estonian samples (n=47) originated from the DNA bank of the Department of Biotechnology. Mandenka (n=24) and Han Chinese (n=25) samples were obtained from HGDP-CEPH Human Genome Diversity Cell Line Panel (http://www.cephb.fr/HGDP-CEPH-Panel/). Czech (n=50), and Korean (n=45) population samples were shared by Dr. Viktor Kozich (Charles University First Faculty of Medicine, Institute of Metabolic Disease) and Dr. Woo Chul Moon (GoodGene Inc. Seoul, Korea), respectively. Unrelated individuals from CEPH/Utah families (n=30) were used as a reference. Common chimpanzee (*Pan troglodytes*) DNA was extracted from sperm material obtained from Tallinn Zoo, Estonia. The sources of orangutan (*Pongo pygmaeus*) and gorilla (*Gorilla gorilla*) DNAs were primary cell lines AG12256 and AG05251B, purchased from ECACC.

Blood samples for DNA analysis of pregnant women (n=48, mean age 26.6±6.9), who had conceived during sexual debut or within three months after stopping contraception, were collected at the Tartu University Clinics Women's Clinic by dr. Kristiina Rull after informed consent was obtained from every participant. Male subjects (all residents of Estonia, of European origin) were recruited at the Andrology Unit, Tartu University Clinics by dr. Margus Punab, during a period of 2 years and an informed consent was obtained from every participant. The study group (mean age 30.5 years) consisted of 94 male subjects of couples with the diagnosis of infertility for at least 12 months and without any known causes of male subinfertility and sperm pathology.

3.2 PCR, resequencing and RFLP analysis

For full resequencing of the *FSHB* gene (2909 bp: 1898 bp of coding region, 456 bp upstream and 555 bp downstream; genomic DNA (100 ng) was amplified in four overlapping fragments using primers (suppl. info Table S1) designed with the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and Smart-Taq Hot DNA polymerase (Naxo, Estonia).

Amplifications were performed in a GeneAmp PCR System 2700 thermal cycler (Applied Biosystems). Reaction was as follows:

initial denaturation	95 ºC 15 min	
denaturation	95 ºC 20 sec	10 cycles
annealing	68 ºC 30 sec	
	-1 ºC per cycle	
extension	72 ºC 1 min	
denaturation	95 ºC 20 sec	10 cycles
annealing	56 ºC 30 sec	
extension	68 ºC 1 min	
denaturation	95 ºC 20 sec	10 cycles
annealing	54 ºC 30 sec	
extension	68 ºC 1 min	
final extension	68 ºC 5 min	

To remove unincorporated PCR primers and mononucleotides, PCR products were treated with exonuclease I (1 U; MBI Fermentas) and shrimp alkaline phosphatase (1.5 U)

and incubated in a GeneAmp PCR System 2700 thermal cycler (Applied Biosystems) at 37°C for 20 min followed by enzyme inactivation at 80°C for 15 min.

The purified PCR product (1.5–3 μL) served as a template in sequencing reactions (10 μL) with sequencing primer (2 pmol) and DYEnamic ET Terminator Cycle Sequencing Kit reagent premix (Amersham Biosciences Inc.) as recommended by the supplier. *FSHB* gene was sequenced from both strands using eight PCR primers and six different sequencing primers. Sequencing reactions (1.5 μL) were run on an ABI 377 Prism automated DNA sequencer (Applied Biosystems) using ReproGel 377 gels (Amersham Biosciences Inc.). The reaction was as follows:

initial denaturation	95˚C	35 sec	} 35 cycles
annealing	50˚C	15 sec	
extension	60˚C	1 min	

For each gene and each population, the sequence data were assembled into a contig using phred and phrap software (Ewing *et al.*, 1998), and the contig was edited in a consed package (Gordon *et al.*, 1998) to ensure that the assembly was accurate (http://www.phrap.org/phredphrapconsed.html).

Polymorphisms were identified using the polyphred program (Version 4.2) (Nickerson *et al.*, 1997) and confirmed by manual checking. A genetic variant was verified only if it was observed in both the forward and the reverse orientations.

Alternatively, as *FSHB* markers were in strong linkage disequilibrium (LD), a tag-SNP approach (5 SNPs) combining genotyping by re-sequencing (from - 456 to 288 relative to ATG) and RFLP analysis were used. SNPs rs594982 and rs6169 were typed by RFLP analysis as they result in the formation of recognition sites for restriction enzymes *XapI* (MBI Fermentas) and *Bst1107*I (MBI Fermentas), respectively. Allelic status of SNPs rs550312, rs611246 and rs609896 was determined by resequencing.

3.3 Data analysis

3.3.1 Basic population genetics parameters

Allele frequencies were estimated and conformance with Hardy-Weinberg equilibrium (HWE) was computed by an exact test (α=0.05) using Genepop 3.1d program (Raymond and Rousset, 1995). LD was evaluated by a descriptive statistic r^2 estimated for pairs common SNPs (minor allele frequency, MAF>10%) using Arlequin 2.000 (Schneider *et al.*, 2000) and the significance of LD between markers was computed with the Genepop 3.1d (Raymond and Rousset, 1995).

Haplotypes were inferred from unphased genotype data using the Bayesian statistical method in the program PHASE 2.1 (Stephens *et al.*, 2001; http://www.stat.washington.edu/stephens/). For haplotype reconstruction, the model allowing recombination was used. Running parameters were: number of iterations = 1000, thinning interval = 1, burn-in = 100, for increasing the number of iterations of the final run of the algorithm the –X10 parameter, making the final run 10 times longer than other runs, was used. The algorithm was run 10 times resulting in identical outputs of the parallel analysis, thus the median of the values obtained from one of the runs was used. Relationships between inferred haplotypes were investigated using the Median-Joining (MJ) network algorithm (Bandelt *et al.*, 1999) within NETWORK 4.0 software.

To estimate the significance level for statistics I used the probability value. The probability value (*p*-value) of a statistical hypothesis test is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0, is true. The null hypothesis, H_0 represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. Small p-values suggest that the null hypothesis is unlikely to be true. The smaller it is, the more convincing is the rejection of the null hypothesis. A p-value <0.001 was considered as highly significant, a p-value between 0.001 and 0.01 as significant, and 0.01<p<0.05 as suggestive.

Fisher's exact probability test was used as appropriate to evaluate proportions for statistical significance. Fisher's exact test is the test for fourfold (2 by 2) tables and it examines the relationship between the two dimensions of the table (classification into

rows vs. classification into columns). The null hypothesis is that these two classifications are not different.

Alignment of human and great apes' *FSHB* genomic sequences was performed with a web-based implementation of CLUSTALW at the EBI (http://www.ebi.ac.uk/clustalw/).

3.3.2 Statistical tests

Sequence diversity parameters were calculated with DnaSP 4.10.3 (Rozas and Rozas, 1999). The direct estimate of per-site heterozygosity (π) was derived from the average pairwise sequence difference. Tajima's D (D^T), Fu and Li's D^{FL}, and Fu and Li's F^{FL} statistics were performed to determine if the observed patterns of intraspecies diversity are consistent with the standard neutral model. Significant positive D^T, D^{FL} and F^{FL} values may indicate an excess of high-frequency SNPs referring to either balancing selection or population bottlenecks. Conversely, significant negative D^T, D^{FL} and F^{FL} values may reflect an excess of rare polymorphisms in a population indicating either positive selection or an increase in population size.

The relative amount of within-species polymorphisms should reflect the amount of between-species fixation under neutrality (Kimura, 1983). The interspecies data was used for the Hudson, Kreitman, and Aguade (HKA) test (Hudson *et al*,. 1987), to determine whether the ratio of polymorphism to divergence across *FSHB* coding regions was consistent with that of noncoding regions (Verrelli and Tishkoff, 2004). Neutrality of *FSHB* was tested by comparing genetic diversity of the human gene with fixed differences between human and primate sequences.

For comparison of testicular (combined testis volume, sperm concentration, total sperm count, semen volume) and hormonal (FSH, LH, testosterone, inhibin B) parameters among the groups of male patients (defined based on their *FSHB* genotype), statistical analysis was carried out using non-parametric Mann-Whitney U-test. It tests the null hypothesis that two samples come from the same population (i.e. have the same median), or alternatively, whether observations in one sample tend to be larger than observations in the other. The advantage of Mann-Whitney U-test is that it allows differences in sample sizes. The analysis was performed with a web-based implementation of the Mann-

Whitney U-test (http://eatworms.swmed.edu/~leon/stats/utest.html). A p-value <0.05 was considered as significant and a p-value <0.1 as suggestive.

4. Results

4.1 *FSHB* gene variation

4.1.1 Detailed FSHB SNP patterns investigated by resequencing

In order to determine fine-scale variation of the entire *FSHB* gene region I chose the resequencing approach. Analyzed population samples originated from Europe (Estonians), Africa (Mandenka) and Asia (Hans), and therefore it was possible to determine worldwide variation of the *FSHB* genomic region.

Resequencing of 2.909 kb of *FSHB* genomic sequence for 196 human chromosomes supported the conservative nature of the gene as no non-synonymous mutations were identified. The resulting *FSHB* genomic sequence collection contained seven common single nucleotide polymorphisms (minor allele frequency > 10%) seen in all populations and only five singleton variants (Fig. 5; Table 3), presented on a single chromosome in the population. One synonymous change (Tyr → Tyr, rs6169) was identified in the coding region of exon 3. Eleven polymorphisms are found to localize in the untranslated regions (5' UTR, intron, 3' UTR). Most identified polymorphisms were localized in the intron. Four (two common and two singleton) 5' UTR region DNA sequence variants were identified. One new synonymous change in the third exon was found in the Estonian population. This change (Lys → Lys) was presented as a singleton. The single nucleotide polymorphism rs506197 was found in the Estonian, Mandenka and Han populations, but was not included in subsequent analyses. *FSHB* SNPs (ss# 49785048-49785060) and population allele frequencies as well as sequences for primate *FSHB* genes (DQ304480-DQ304481, DQ302103) were submitted to NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and NCBI Genbank. All the data will be available after Grigorova *et al*. (Ann Hum Gen, in press) is published online.

Resequencing elucidated that *FSHB* gene SNP patterns appeared to be similar among populations. It might be indicative for the ancient evolution and critical functions of *FSHB* gene. This evidence is also supported by the finding of enrichment of

intermediate-frequency polymorphisms and the lack of coding DNA sequence variations that change the primary structure of FSHB protein.

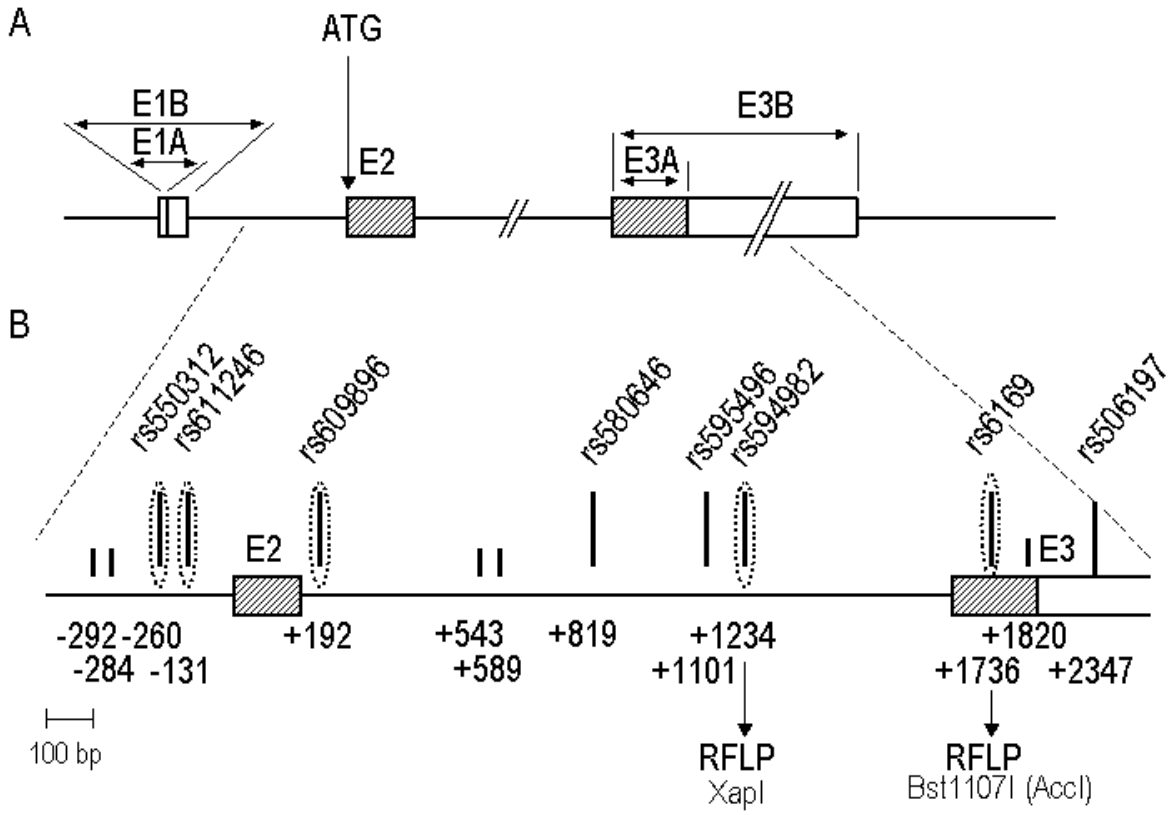


Figure 5a,b Schematic representation of the re-sequenced human *FSHB* region. **a** Gene structure drawn to an approximate scale. Exons (E1-E3) are indicated by boxes with the coding and non-coding sequences denoted by grey and open areas, respectively. E1 contains an alternate splice donor site resulting in two forms, A and B; E3 contains alternate polyadenylation sites, A and B (Jameson *et al.* 1988). The intron processing and polyadenylation are regulated independently. **b** Human SNPs (*vertical black bars*) identified in Estonians, Mandenkas and Hans are marked as *long bars* for common SNPs (MAF >10%) and *short bars* for rare SNPs (<10%). The SNP positions are relative to ATG. Five tag-SNPs (*dotted ellipses*) were chosen for genotyping Korean, Czech, CEPH and STP-sample. Two of the tag-SNPs (rs594982 and rs6169) also represent RFLPs (Grigorova *et al.*, in press).

Table 3. *FSHB* gene SNP minor allele frequencies

			Full Resequencing			RFLP/ Partial Resequencing[b]		
Position[a]	SNP	dbSNP rs no	Estonians (n=48)	Mandenka (n=24)	Han (n=25)	Korea (n=50)	Czech (n=50)	CEPH/Utah (n=30)
-292	A/T	–	–	–	S	–	–	–
-284	A/T	–	–	S	–	–	–	–
-260	G/T	550312	0.447	0.375	0.240	0.367	0.510	0.700
-131	T/A	611246	0.426	0.333	0.200	0.356	0.520	0.700
192	C/T	609896	0.447	0.375	0.180	0.322	0.520	0.700
543	T/G	–	–	–	S			
589	A/G	–	–	S	–			
819	G/T	580646	0.436	0.354	0.180			
1101	C/T	595496	0.447	0.375	0.200			
1234	C/A	594982	0.415	0.271	0.180	0.322	0.520	0.700
1806	T/C	6169	0.457	0.167	0.180	0.256	0.500	0.700
1820	G/A	–	S	–	–			
2347	G/A	506197	0.447					

S = Singleton polymorphism; n = No of individuals; Empty cells = SNP not analyzed; '–' = SNP not identified; [a]Relative to ATG; [b]See text for explanation.

4.1.2 FSHB gene might be under positive selection

To investigate whether *FSHB* gene has evolved under neutral evolution circumstances, different sequence diversity parameters were calculated. The direct estimate of per-site heterozygosity (π) (Tajima, 1983) was derived from the average pairwise sequence difference, and Watterson's θ represents as the estimate of the expected per-site heterozygosity based on the number of segregating sites (S) (Watterson, 1975).

Under the null hypothesis, the observed and expected sequence diversity parameters should be roughly equal in value. D^T, D^{FL} and F^{FL} statistics are the most widely used ways of comparing the allele frequency spectrum against the expectations of the null model. Tajima's D (D^T) (Tajima, 1989), Fu and Li's D^{FL}, and Fu and Li's F^{FL} (Fu and Li, 1993) statistics were performed to determine if the observed patterns of intraspecies diversity are consistent with the standard neutral model. Significant positive D^T, D^{FL} and F^{FL} values may indicate an excess of high-frequency SNPs referring to either balancing selection or population bottlenecks. Conversely, significant negative D^T, D^{FL} and F^{FL} values may reflect an excess of rare polymorphisms in a population indicating either positive selection or an increase in population size. The significance of values was assessed using p-value.

The estimated positive D^T, D^{FL} and F^{FL} values for *FSHB* gene (Table 4) fell into the upper range of the distribution determined in a recent study for 132 different human genes in European- and African-Americans (Akey *et al*. 2004), where only a few analyzed genes (e.g. *ABO, ACE2, IL10RB, IL11A*) resulted in estimates as high as determined for *FSHB*. This indicates an enrichment of intermediate-frequency alleles for *FSHB* polymorphisms consistent with either balancing selection or population demography characterized by subdivision or reduction in size (reviewed by Bamshad and Wooding, 2003). For Estonians and Mandenkas significant positive values in two tests (D^T; F^{FL}) rejected the hypothesis of neutrality. Failure to reject the hypothesis of neutrality by Fu and Li's D^{FL} test may be caused by weaker power (Simonsen *et al*. 1995).

Table 4. *FSHB* nucleotide diversity parameters and neutrality tests

	Estonians	Mandenka	Han
Sample size	47	24	25
Diversity estimates and neutrality tests			
π^1	0.00123	0.00109	0.00079
θ^1	0.00048	0.00056	0.00055
Tajima's D^T	3.224[2]	2.523[3]	1.130
Fu and Li's D^{FL}	1.198	1.244	1.242
Fu and Li's F^{FL}	2.397[4]	1.938[4]	1.416

[1] An estimate of nucleotide diversity per site from average pairwise difference among individuals (π) and number of segregating sites (θ)
[2]$p<0.01$, [3]$p<0.05$ for Tajima's D^T statistics
[4]$p<0.02$ for Fu and Li's D^{FL} and F^{FL} statistics

4.1.3 FSHB gene exhibits strong intergenic LD

To examine patterns of linkage disequilibrium, I calculated squared correlation coefficient r^2 for all pairwise comparisons of segregating sites for which the minor allele was found in two or more individuals. r^2 is the squared correlation in allelic state between the two loci as they occur in haplotypes.

FSHB exhibited strong intergenic linkage disequilibrium for both applied statistics: p-values from Fisher's exact test and correlation coefficient r^2, consistent with expectations over short distances (Table 5a,b,c).

Allelic associations were significant and strong throughout the gene for Estonians ($p<0.0001$, $0.795<r^2<1$) (Table 5a), Mandenka ($p<0.001$, $0.62<r^2<1$) (Table5b) and Han ($p<0.001$, $0.53<r^2<1$) (Table 5c).

The strength of LD is weaker in Mandenka and Han populations (Table 5b,c). It could be explained by smaller sample size, and therefore lower statistical power.

The evidence of strong allele association supports hypothesis that *FSHB* gene or given genomic region is under the influence of evolutionary forces like selection.

Table 5. LD values (r^2) in studied population samples for all pairwise comparisons of segregating sites.

(a) Estonian population

rs#							
550312	xxxxx						
611246	0,958***	xxxxx					
609896	1***	0,958***	xxxxx				
580646	0,876***	0,833***	0,876***	xxxxx			
595496	0,835***	0,8***	0,835***	0,795***	xxxxx		
594982	0,835***	0,8***	0,835***	0,795***	0,97***	xxxxx	
6169	0,958***	0,911***	0,958***	0,837***	0,797***	0,799***	xxxxx
	550312	**611246**	**609896**	**580646**	**595496**	**594982**	**6169**

* 0,01<P<0,05 ; ** 0,001<P<0,01 ; *** P<0,001

(b) Mandenka population

rs#							
550312	xxxxx						
611246	0,914***	xxxxx					
609896	1***	0,914***	xxxxx				
580646	0,83***	0,914***	0,83***	xxxxx			
595496	0,83***	0,914***	0,83***	0,83***	xxxxx		
594982	0,62***	0,677***	0,619***	0,62***	0,62***	xxxxx	
6169	0,333***	0,365***	0,333***	0,333***	0,333***	0,538***	xxxxx
	550312	**611246**	**609896**	**580646**	**595496**	**594982**	**6169**

* 0,01<P<0,05 ; ** 0,001<P<0,01 ; *** P<0,001

(c) Han population

rs#							
550312	xxxxx						
611246	0,43***	xxxxx					
609896	0,695***	0,458***	xxxxx				
580646	0,695***	0,173*	0,531***	xxxxx			
595496	0,792***	0,4***	0,651***	0,651***	xxxxx		
594982	0,695***	0,458***	1***	0,531***	0,651***	xxxxx	
6169	0,695***	0,458***	1***	0,531***	0,651	1***	xxxxx
	550312	**611246**	**609896**	**580646**	**595496*** **	**594982**	**6169**

* 0,01<P<0,05 ; ** 0,001<P<0,01 ; *** P<0,001

4.1.4 FSHB has worldwide two major 'yin-yang' haplotypes

The presence of intergenic linkage disequilibrium makes it possible to infer *FSHB* gene haplotype pairs for each individual from genotype data.

Haplotype frequency estimated by the PHASE algorithm revealed one prevalent *FSHB* gene variant (no. 1a, HAP1; Table 6a) spread around the world with frequencies ranging from 51.1 % in Estonians to 62.5 % in Mandenka. Interestingly, the next most frequent gene variant (no. 13a, HAP13; Table 6a) is composed of completely mismatching SNP alleles compared to this dominant haplotype, i.e. nucleotides differ at every SNP in this haplotype pair, a feature termed yin-yang haplotypes (Zhang *et al*., 2003). This second gene variant is represented in Estonians with a frequency (38.3 %) more than two times greater than that observed in the other resequenced population samples (Mandenka 12.5 %, Han 14 %).

To analyze additional populations for *FSHB* diversity, 5 tag-SNPs sufficient to represent all major gene variants were chosen (Fig. 5; Table 6a).

The five *FSHB* SNPs selected for this analysis were typed in additional populations from Europe (Czech, unrelated CEPH/Utah individuals) and Asia (Korea) (Table 6b). The allelic status for the three proximal SNPs (rs550312/from ATG -260, rs611246/ -131 and rs609896/ +192) was defined by resequencing and for the two distal SNPs (rs594982/ +1234/XapI and rs6169/ +1763/ Bst1107I) by RFLP analysis.

The data confirmed the presence of the two principal gene variants composed of completely mismatching SNP alleles in all populations, together covering 96.6% of CEPH, 96% of Czech, 92.6% of Estonian, 86% of Han, 79.2% of Mandenka, and 76% of Korean 5-SNP haplotypes (Table 6b). The second most frequent haplotype (HAP13) appeared to be enriched in populations of European-origin, ranging from 39.4% (Estonians) to 68.3% (CEPH), compared to non-Europeans for which frequencies ranged from 14% (Han) to 21% (Korea). In Median-Joining network of all *FSHB* haplotypes the two groups of haplotypes are clearly separated, with HAP1 approximately equally distributed among Europeans and non-Europeans and HAP13 enriched in populations of European origin (Fig. 6).

Table 6. Haplotype structure of *FSHB* gene.

(a) *FSHB* haplotypes defined on the basis of 7 common SNPs (MAF>10%); (b) *FSHB* haplotypes defined on the basis of 5 'tag' SNPs (Fig, 5b).

(a)

HAP	rs numbers of SNPs with MAF > 10 %							**Population frequency**		
no	550312	611246	609896	580646	595496	594982	6169	Est	Man	Han
1a	G	T	C	G	C	C	T	0.511	0.625	0.680
1b	G	T	C	G	T	C	T	S	–	S
1c	G	T	C	T	C	C	T	S	–	S
2	G	T	C	G	C	C		S	–	–
3	G	T	C	G	T	A	T	S	–	–
5	G	A	C	G	C	C	T	–	–	0.040
6	T	T	C	T	T	C	T	–	S	0.040
7	T	A	C	G	C	C	T	–	–	S
8	T	T	T	T	T	A	C	S	–	0.040
10	T	A	T	T	T	C	T	–	0.080	–
11a	T	A	T	G	C	C	C	S	–	–
11b	T	A	T	T	C	C	C	S	–	–
11c	T	A	T	T	T	C	C	0.020	–	–
12	T	A	T	T	T	A	T	–	0.100	–
13b	T	A	T	G	T	A	C	S	0.040	–
13a	T	A	T	T	T	A	C	0.383	0.125	0.140
chi	T	A	T	T	T	C	T			
gor	T	A	T	T	T	C	T			
oran	T	del	T	T	T	C	T			

Est = Estonians, Man = Mandenka; chi = chimpanzee; gor = gorilla; oran = orangutan; S = Singleton haplotype; '–' = haplotype not identified. Two core haplotypes are boxed. SNP alleles identical to HAP1 are lettered black on white background and alleles identical to HAP13 are lettered white on black background. 'a, b, c' denote haplotypes, which pool together when five tag-SNPs (Fig. 5) are used.

(b)

HAP	*'tag' SNP rs number*					*Population frequency*					
no	550312	611246	609896	594982	6169	Est	Man	Han	Kor	Cze	CEPH
1	G	T	C	C	T	0.532	0.625	0.720	0.550	0.480	0.283
2	G	T	C	C	C	S	–	–	0.020	–	S
3	G	T	C	A	T	S	–	–	–	–	–
4	G	T	T	C	T	–	–	–	–	S	–
5	G	A	C	C	T	–	–	0.040	–	–	–
6	T	T	C	C	T	–	S	0.040	S	–	–
7	T	A	C	C	T	–	–	S	0.030	–	–
8	T	T	T	A	C	S	–	0.040	–	–	–
9	G	A	T	A	C	–	–	–	–	S	–
10	T	A	T	C	T	–	0.083	–	–	–	–
11	T	A	T	C	C	0.043	–	–	–	–	–
12	T	A	T	A	T	–	0.104	–	0.080	0.020	S
13	T	A	T	A	C	0.394	0.167	0.140	0.210	0.480	0.683
chi	T	A	T	C	T						
gor	T	A	T	C	T						
oran	T	del	T	C	T						

Est = Estonians; Man = Mandenka; Kor = Koreans; Cze = Czech; CEPH = unrelated CEPH/Utah individuals; chi = chimpanzee; gor = gorilla; oran = orangutan; S = Singleton haplotype; '–' = haplotype not identified; Two core haplotypes are boxed. SNP alleles identical to the worldwide *FSHB* variant are lettered black on white background and alleles identical to the 'European' variant are lettered white on black background.

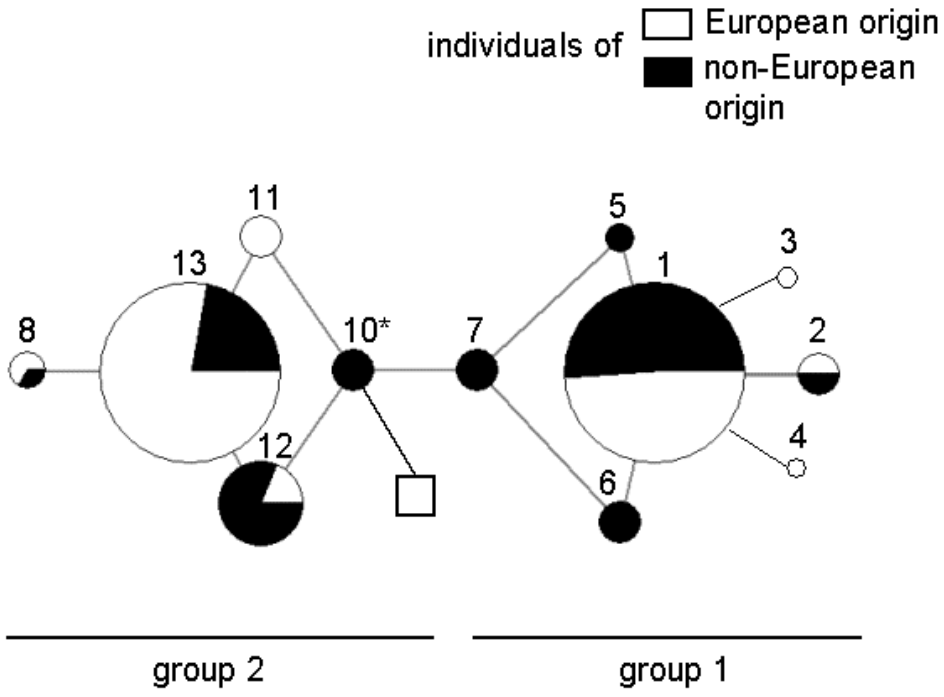


Figure 6. Median-Joining (MJ) network for predicted *FSHB* haplotypes. Haplotypes have been constructed based on five *FSHB* tag-SNPs (Fig. 5b; Table 6b) typed for three European (Estonian, Czech, unrelated CEPH/Utah), two Asian (Han, Korean) and one African population (Mandenka). The size of each node is proportional to the haplotype frequency in the total dataset. Branch lengths represent one nucleotide substitution. The relative distribution of each haplotype among individuals of non-European and European origin is indicated with black and white, respectively. Haplotype nomenclature is identical to Table 6. Haplotypes form two groups, clustering around the major variants HAP1 and HAP13 differing at every SNP.

*Human HAP10 is shared with chimpanzee and gorilla *FSHB* sequence. Orangutan haplotype is denoted by a square.

4.1.5 'Haplotype 13' is related to ancestral variant in primates

In order to uncover the ancestral *FSHB* variant among primates, the chimpanzee (C), gorilla (G) and orangutan (O) gene was sequenced (suppl. info fig. S2). Divergence of primate *FSHB* from the human (H) sequence falls into the range of previous estimates: for H/C 1.28%, for H/G 1.92% and for H/O 3.38% compared to a report for 53 intergenic regions 1.24±0.07%, 1.62±0.08% and 3.08±0.11%, respectively (Chen and Li, 2001). Three amino acid differences in great apes relative to human *FHSB* were identified: a change in signal peptide (amino acid no. 4) from Leu to Val between human and orangutan (H/O), and changes in the mature protein from Tyr to His (exon 2, amino acid no. 49) between H/C and Lys to Asn (exon 3 amino acid no. 64) between human and all of the other studied species. None of the differences are located within the region of the hormone adjacent to the receptor in the recently described FSH-FSHR co-crystal structure (Fan and Hendrickson, 2005).

Primate haplotypes formed from the positions of human common SNPs were identical among the chimpanzee, gorilla and orangutan, except for a single basepair deletion in the orangutan gene for SNP rs611246 (Table 6a). When compared to human *FSHB* the conserved great apes' haplotype is seen to be most similar to human HAP13, as opposed to HAP1 (Fig. 6). Only 2 changes are required from the conserved great ape haplotype to human HAP13, whereas HAP1 is different in five positions from other primates (Table 6a).

The three primate (chimpanzee, gorilla and orangutan) sequences were used as references for the Hudson-Kreitman-Aguade test (Hudson *et al.*, 1987). The HKA test is based on the most basic prediction of the neutral model of molecular evolution, that DNA sequence polymorphisms within a species, and DNA sequence divergence between species, will be proportional to the neutral mutation rate (Kimura, 1983). The HKA test resulted in rejecting the scenario of positive selection on the *FSHB* gene in all three resequenced human populations (H/C χ^2=0.825-0.869, 0.35<p<0.37; H/G χ^2=0.063-0.085, 0.77<p<0.81; H/O χ^2=0.415-0.470, 0.49<p<0.52).

4.2 *FSHB* gene variants and reproductive success

It was determined that *FSHB* gene has two worldwide spread variants. Standard neutrality tests have shown the evidence that the gene might be under influence of balancing selection. To test whether one of the *FSHB* gene variants is associated with human fertility and reproductive success, two studies were carried out.

4.2.1 FSHB gene variants are associated with effectiveness of sperm production

Five human *FSHB* tagSNPs (Fig. 5) were genotyped for a cohort (n=94) of men visiting the andrology laboratory at the Andrology Unit of the Tartu University Clinics in the frame of 2 years. In studied cohort of individuals, *FSHB* gene was represented by two major haplotypes formed on the basis of the five genotyped SNPs (Fig. 5b). A recombinant haplotype was defined only for one case. The subject, who carried recombinant haplotype, was excluded from the following analysis.

Participating subjects were grouped based on the *FSHB* genotypic status as H1/H1 homozygotes (n=18), H13/H13 homozygotes (n=22) or H1/H13 heterozygotes (n=53).

The groups did not differ (Mann-Whitney U-test, $p>0.1$) in the parameters of possible confounding effects on sperm quality, the age (H1/H1: median 31 years, range 25-42; H13/H13: median 30, range 21-40, H1/H13: median 31, range 24-46) and the ejaculatory abstinence period (H1/H1: median 4 days, range 3-10; H13/H13: median 4, range 1.5-7, H1/H13: median 4, range 1-10). As a major outcome of the study, *FSHB* H1 and H13 homozygotes exhibited significantly higher sperm concentration (Mann-Whitney U-test, $p=0.006$ and $p=0.074$, respectively) and total sperm count ($p=0.002$ and $p=0.066$, respectively) compared to H1/H13 heterozygotes (Fig. 7).

The total testis volume of H13 homozygote individuals exceeded the measures of other study groups. The difference between H13 and H1 homozygotes exhibited a borderline statistical significance by Mann-Whitney test ($p=0.070$; Fig. 7).

Thus, for H13 homozygotes the higher sperm count may result from a larger testicular size, while H1 homozygosity might be associated with more effective sperm production.

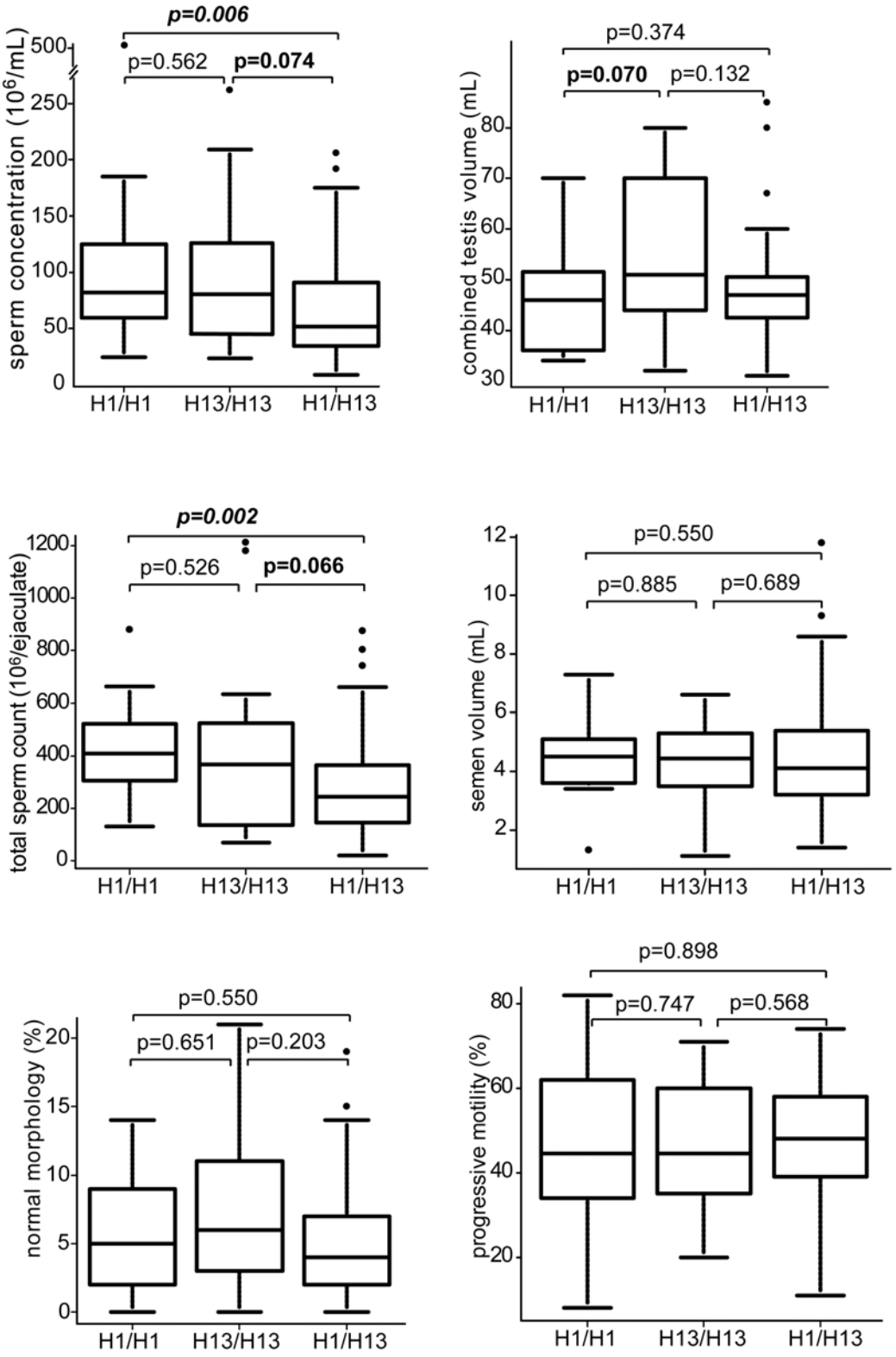


Figure 7 continued and legend on p.40

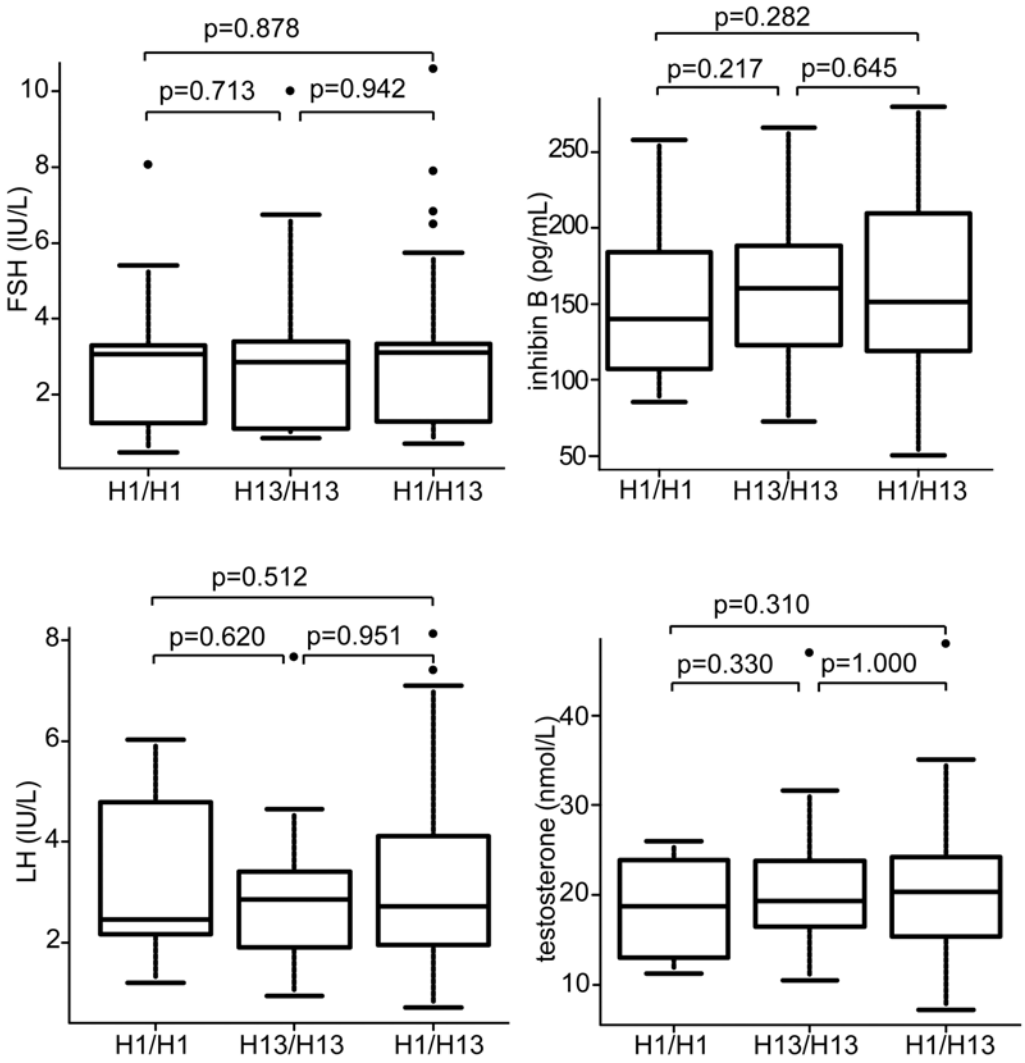


Figure 7. Boxblot for the distribution of sperm concentration (analyzed for subjects n=93), combined testicular volume (n=93), total sperm count (n=93), semen volume (n=93), sperm morphology (n=91) and progressive motility (n=93) and serum FSH (n=93), LH (n=92), Inhibin B (n=53) and testosterone values (n=87) of men subgrouped according to their *FSHB* genotype. H1/H1 (n=18), H13/H13 (n=22) and H1/H13 (n=53) designate homozygotes and heterozygotes for *FSHB* haplotype 1 and 13. The boxes represent the 25th and 75th percentiles; whiskers are lines extending from each end of the box covering the extent of the data on 1.5 X interquartile range. The median value is denoted as the line that bisects the boxes. Circles represent the outlier values. *p*-values were computed using Mann-Whitney U-test, which is a test that compares medians of two unpaired groups. $p<0.05$ were interpreted as significant, $p<0.1$ were interpreted as borderline difference, $p>0.1$ – non-significant.

There was no statistical difference (Mann-Whitney U-test, p>0.1) between the study groups in testosterone and LH levels, semen volume, sperm morphology and motility (Fig. 7) and Inhibin B levels. Notably, FSH concentrations did not differ significantly between the carriers of the alternative combinations of *FSHB* variants either (p>0.1; Fig. 7). However, this could also reflect the lack of power in the applied statistical test due to inadequate group sizes, or the nature of the used standard laboratory assays based on quantitative antibody-based measurements and not addressing hormone quality and activity (Rose *et al*., 2000).

4.2.2 HAP13 may favour rapid conceiving and short time to pregnancy

In order to test the hypothesis that the identified *FSHB* core haplotypes have functional consequence to female reproductive success, Estonian women who had conceived within three months after stopping contraception (STP=short time to pregnancy), were analyzed. When STP-women were compared to a random Estonian sample, the distribution of the two core *FSHB* variants was significantly different (χ^2=3.982; p<0.05). The frequency of the worldwide variant (HAP1) was 38.5% in STP-women and 53.2% in the random population sample, whereas the prevalence of the HAP13 was 52.1% in the STP group and 39.4% in the random sample. Notably, an even more significant difference between the two samples was detected for the distribution of the homozygotes for these haplotypes (χ^2=10.471; p<0.002). HAP1 homozygote frequency was 14.6% for STP-women compared to 27.7% in random individuals, whereas HAP13 homozygotes formed 33.3% of STP-women and only 14.9% of population sample (Fig. 8).

The enrichment of HAP13 among STP-women remains significant when HAP13 homozygotes are compared to the rest of the sample (χ^2=8.882, p<0.003; Fig.8). Consequently, we hypothesize that this *FSHB* variant might possibly be associated with rapid success in achieving conception after cessation of contraception.

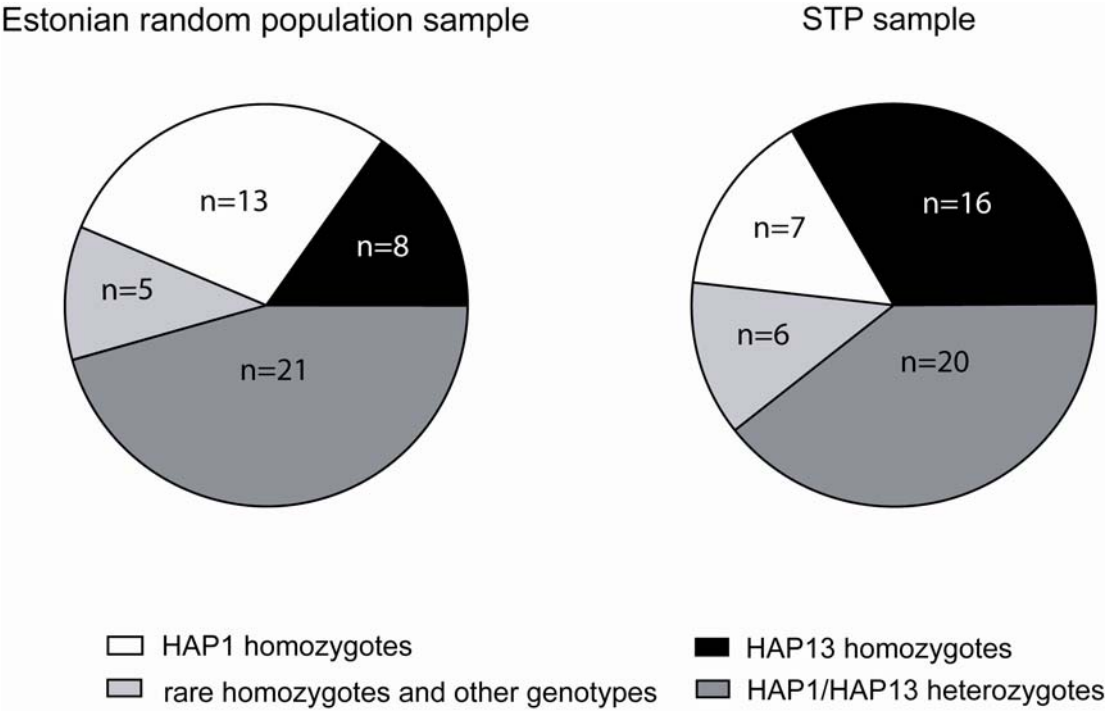


Figure 8. The distribution of homo- and heterozygotes in a random Estonian sample and Estonian women who conceived within three months after stopping contraception (STP=short time to pregnancy).

5. Discussion

5.1 Evidence for the balancing selection on *FSHB* gene

Possible selection on the maintenance of *FSHB* gene function is revealed by its evolutionarily conserved sequence in mammals and teleost fish (Li and Ford, 1998) as well as by the very small number of non-synonymous mutations (n=4, cases=8, all leading to infertility), identified in humans (Berger *et al*., 2005). Consistent with this evidence for conservation, the re-sequencing study of 96 individuals, representing populations from Europe, Africa and Asia, did not uncover any non-synonymous changes. The majority of the *FSHB* SNPs were represented as common polymorphisms with worldwide occurrence. Two of the variants overlapped with previous studies: (i) a common synonymous change Tyr-Tyr (rs6169) in exon 3 present in Asia (Han Chinese, Malays Indians, Koreans), Europe (Finns, Danes, Estonians, Czech, CEPH/Utah) and Africa (Mandenka) (Lamminen *et al*., 2005; Liao *et al*. 1999); and (ii) rs609896 also found in Asia (Han Chinese, Japanese, Koreans), Europe (Estonians, Czechs, CEPH/Utah, Finns, Danes) and Africa (Mandenka, Yoruba) (The International HapMap Consortium, 2005; Lamminen *et al*., 2005). Two of the population-specific synonymous changes described for Finns and Danes were not detected (Lamminen *et al*., 2005).

Two worldwide *FSHB* core haplotypes (HAP1, HAP13 carried by 76 to 96.6% of each population's individuals), the sequences of which are clearly separated from each other, were identified (Fig. 6). Statistical tests suggested significant deviations for *FSHB* from neutrality. Three neutrality tests (D^T, D^{FL}, F^{FL}) indicated an enrichment of SNPs with intermediate minor allele frequencies. The results are consistent with the scenario of balancing selection acting on *FSHB* gene. Although similar results can be caused by population subdivision, in which haplotypes are restricted to specific subpopulations or reduction in population size, this is not the case for *FSHB* as all studied populations share the two major haplotypes as well as show similar outcomes of the neutrality tests.

5.2 *FSHB* gene variants and reproductive success

The requirement for follicle-stimulating hormone to regulate reproduction and fertility has been demonstrated through the use of gene knockouts (Kumar *et al*., 1997), specific FSH (McLachlan *et* al., 2002; Moudgal and Sairam, 1998; Plant and Marshall, 2001) and FSH receptor (Moudgal *et al*., 1997) antibodies as well as expression studies that have been carried out on mouse using LbetaT2 gonadotroph cell line that expresses and secretes FSH (Graham *et al*., 1999; Pernasetti *et al*., 2001).

As fertility affects an organism's fitness most directly, the carriers of most common haplotypes have apparently had more success in human history to contribute to the next generation. *FSHB* HAP1 was dominant in almost all studied populations, consistent with the notion that it might favor an individual's reproductive success.

In men, present data suggest potential differential effects of the core human *FSHB* gene variants on testicular parameters. A major determinant of the total sperm output is the number of Sertoli cells in the testis, the reduction in germ cell count in the *FSHB* knockout was shown to be the outcome of not only impaired Sertoli cell proliferation, but also of the reduced capacity of Sertoli cells to nurture germ cells (Andersen *et al*., 2000). Thus, it is possible that human *FSHB* HAP13 might have a positive effect on Sertoli cell proliferation, while HAP1 may be related to more effective Sertoli cell function responsible for spermatogenesis. Hence, the increase in total testis volume for HAP13 homozygotes might reflect a compensatory effect to guarantee sufficiently high sperm counts for successful reproduction. The heterozygote variant is determined as a disadvantaged variant having lower total sperm count in comparison to both HAP1 and HAP13 haplotype homozygotes.

In women, there is a preliminary observation for a significant statistical difference in the quantity of HAP13 homozygote carriers between a normal Estonian population sample and Estonian women who had conceived within three months after stopping contraception. It was found that HAP13 homozygotes were enriched in STP-women sample. The data of this study allows postulating of a hypothesis that one of the worldwide spread *FSHB* gene variants, the so-called 'European' variant may be

associated with rapid conceiving. Thus, HAP13 might have been temporarily advantageous during human history.

HAP13 found to be enriched in Europe and that is potentially associated with rapid conceiving, is also the human haplotype most closely related to the ancestral *FSHB* inferred from the sequences present in all the extant great apes. The emergence of HAP1 might be related to differences between humans and great apes in fertility and reproductive behavior. It is possible that the current dominant *FSHB* haplotype (HAP1) has been favored by selection because it might be associated with enhanced reproductive success characterized by optimal birth intervals necessary for higher rates of mother and child survival (Blurton Jones *et al*., 1989).

In summary, there is growing evidence that an individual's *FSHB* haplotypic/genotypic content may have functional consequence and two core haplotypes of *FSHB* gene could provide a selective advantage in reproduction.

5.3 Possible functional consequence of *FSHB* gene haplotypes

Present study elucidated that two worldwidely spread *FSHB* gene variants have been favored by selection. These variants were found in all studied populations. The data of two studies on men and women allow to hypothesize about indirect evidence on functional consequences of the two core haplotypes of *FSHB* gene.

None of the polymorphisms defining the common *FSHB* gene variants HAP1 and HAP13 are amino acid changing mutations that would directly affect the structure of the *FSHB* protein. Hence, two core haplotypes might be associated with regulatory polymorphisms that are localized outside of the gene coding region. Due to strong linkage disequilibrium in the *FSHB* gene region, the HAP1 and HAP13 could have been driven to high frequency by hitchhiking with alternative alleles of a critical polymorphism within a putative upstream or downstream gene regulatory region contributing to the differential expression of *FSHB*.

A variety of mechanisms modulate the ability to translate mRNA. The length of 3' untranslated region (UTR) and poly(A) tail play important role in determining both the translational efficiency and the stability of mRNA.

It is known that there are at least four *FSHB* gene mRNA transcripts having different length of 3'UTR. The identified core haplotypes of *FSHB* gene might be associated with regulatory polymorphisms in 3'UTR that determine the length of mRNA transcripts. A recent in vivo study reported the presence of critical regulatory elements in human *FSHB* gene mapping to the 1 kb genomic segment downstream of position +2138, relative to the translational stop codon in exon 3 (Kumar *et al*., 2006). Analysis of the transgenic mouse lines revealed that none of the *FSHB* transgenes that contained truncations of the 3' genomic flanking sequences to either positions +2138, +1262, or +1227 demonstrated expression despite the inclusion of necessary 5' promoter elements.

6. Perspectives

This study suggests that *FSHB* gene variants may be associated with human reproductive success. Future projects include:

1. Replication studies in larger sample collections for the association of *FSHB* gene variants with reproductive parameters.
2. Cell culture experiments targeting functional consequences of identified *FSHB* gene haplotypes.

Summary

Follicle-stimulating hormone (FSH) is essential for human reproduction. The unique functions of this hormone are provided by the FSH receptor-binding beta-subunit encoded by the *FSHB* gene. This study presents the sequence variation in *FSHB* across human populations and its possible functional consequences on human fertility. A comprehensive review of research papers on gonadotropine hormones, their function in reproduction and described mutations is given.

Resequencing of one African, 2 Asian and 3 European populations revealed lack of rare single nucleotide polymophisms and two dominant *FSHB* haplotypes: one frequent worldwide, and the other enriched in Europe as a probable consequence of a past founder effect and/or temporary selective advantage.

In order to test whether these variants are associated with human reproductive parameters, I analyzed *FSHB* variants in a cohort of males along with testicular and hormonal parameters; and groups of females selected based on conceiving success. In men, the data suggests potential differential effects of the core human *FSHB* gene variants on testicular parameters. The total testis volume of homozygote carriers of one of the *FSHB* gene haplotype exceeded the measures for other groups. Significantly higher sperm concentration was estimated for homozygotes of both *FSHB* gene haplotypes if compared to heterozygotes. In women, homozygosity of one of the core variants might be associated with rapid conceiving. This was the haplotype that is also related to an ancestral *FSHB* variant shared with the ancestor of the great apes (chimpanzee, gorilla and orangutan).

The determination of the functional consequence of the *FSHB* variants may have implications for understanding and regulating human fertility as well as in assisting infertility treatments.

Kokkuvõte

Folliikuleid stimuleeriv hormoon ehk follitropiin kuulub glükoproteiinsete gonadotroopsete hormoonide perekonda. FSH stimuleerib naistel munasarja folliikuli arenemist. FSH seondumine granuloosa rakkude FSH retseptoritega põhjustab folliikulite vohamist, folliikulite poolt androgeenide östrogeenideks muutmist ja inhibiini produtseerimist. Meestel follitropiini seondumine FSH retseptoriga reguleerib spermatogeneesi. Nagu kõik gonadotropiinid, koosneb follitropiin α- ja β-subühikust. α-subühik on kõikidel hormoonidel ühine, β-subühik on spetsiifiline iga hormooni jaoks. Follitropiini β-subühuku geeni produkt mängib ülitähtsat rolli inimese seksuaalses arengus ja reproduktsioonis. *FSHB* geen on väga konserveerunud struktuuriga ja funktsiooniga. Kõik senini leitud muutused, mis kajastuvad aminohappelisel järjestusel, põhjustavad viljatust. Mitte-kodeerivad muutused on kõrge sagedusega ning see viitab nende polümorfismide evolutsioonilise vanusele ja/või tugevale selektsioonile antud geenis.

Käesoleva magistritöö kirjanduse osas antakse ülevaade gonadotroopsete hormoonide perekonnast, nende funktsioonidest ja struktuurist. Pikemalt on kirjutatud follitropiinist ja selle β-subühiku kodeerivast geenist ning mutatsioonidest, mis põhjustavad viljatust. Samuti on esitatud põgus ülevaade FSH ja *FSHB* molekulaargeneetiliste ja funktsionaalsete uuringute kohta.

Töö tulemusena tuvastati *FSHB* geeni detailne polümorfismide (SNP-de) muster Eesti, Tsehhi, Utah/CEPH, Mandenka, Han ja Korea populatsioonides resekveneerimise ja restriktsiooni fragmentide pikkuse analüüsi teel ning defineeriti *FSHB* geeni põhilised populatsioonides esinevad tuumhaplotüübid.

Saadud andmete põhjal püstitati hüpotees *FSHB* geenivariantide seotusest reproduktiivse edukusega. Oli analüüsitud androloogia kabinetti tulnud meestest koosnev valim. Tulemused viitavad sellele, et *FSHB* geeni variandid on seotud mehe reproduktiivse tervise kliiniliste parameetritega. Samuti analüüsiti valim, kuhu kuulusid naised, mis rasestusid kolme kuu jooksul pärast rasestumisvastaste ravimite võtmist. Üldpopulatsioonil põhineva valimiga võrdluse tulemused näitasid, et kiiresti rasestunud naiste valimis on rikastunud kindla tuumhaplotüübi homosügoodid.

Edasiste plaanide hulka kuuluvad: (i) uurida *FSHB* geenivariantide jaotuvuse seotus reproduktiivse edukusega suurema indiviidide arvuga valimites; (ii) geeniekspressiooni uuringud, mis võimaldaks uurida detailsemalt geeniregulatsiooni ja ekspressiooni mehhanisme.

Acknowledgements

I would like to thank my supervisor, Maris Laan, for her encouragement and support, especially for sharing her enthusiasm for reproductive genetics research with me. I also thank dr. Kristiina Rull and dr. Margus Punab for DNA samples as well as timely discussions and advises on human female and male fertility issues. Drs. Howard Cann, Viktor Kozich, Woo Chul Moon are thanked for providing Mandenka and Chinese Han, Czech and Korean DNA samples, respectively. I owe a thank to Tõnu Margus for assistance in sequence assembly and Viljo Soo for technical help. I would also like to thank all the co-workers for supportive and friendly atmosphere.

References

Achermann JC, Weiss J, Lee E-J, Jameson JL Inherited disorders of the gonadotropin hormones. Mol Cell Endocrinol 2001 179:98-96

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. 2004 2, e286

Ammerman AJ, Cavalli-Sforza L The Neolithic Transition and the Genetics in Europe. 1984 (Princeton University Press, Princetown, NJ)

Andersen AG, Jensen TK, Carlsen E, Jørgensen N, Ansersson AM, Krarup T, Keiding N, Skakkebaek NE High frequency of sub-optimal semen quality in an unselected population of young men. Hum Reprod 2000 15(2):366-372

Bamshad M, Wooding SP Signatures of natural selection in the human genome. Nat Rev Genet 2003 4:99-111

Bandelt HJ, Forster P, Rohl A Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 1999 16:37-48

Berger K, Souza H, Brito VN, d'Alva CB, Mendonca BB, Latronico AC Clinical and hormonal features of selective follicle-stimulating hormone (FSH) deficiency due to FSH beta-subunit gene mutations in both sexes. Fertil Steril 2005 83:466-70

Blurton Jones N, Hawkes K, O'Connell J in Comparative socioecology of humans and other mammals., ed. Standen V, Foley RA (Basil Blackwell, London), 1989, pp. 367-390

Bousfield GR, Perry WM, Ward DN Gonadotropins. Chemistry and biosynthesis. In: Knobil E, Neill JD (eds) The physiology of Reproduction, ed 2. Raven Press, New York, pp 1749-1792

Brockman JM, Singh P, Liu D, Quinlan S, Salisbury J, Graber JH PACdb: PolyA cleavage site and 3'-UTR database. Bioinformatics 2005 21:3691-3693

Burger LL, Haisenleder DJ, Dalkin AC, Marshall JC Regulation of gonadotropin subunit gene transcription. J Mol Ecdocrinol 2004 33:559-584

Chen FC, Li WH Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 2001 68:444-56

Cross NL, Morales P, Overstreet JH, Hanson FW The simple methods for detecting acrosome-reacted human sperm. Gamete Res 1986 15:213-226

Culler MD, Negro-Villar A Evidence that pulsatile follicle-stimulating hormone secretion is independent of endogenous luteinizing hormone-releasing hormone. Endocrinology 1986 118:609-612

Dalkin AC, Haisenleder DJ, Gilrain JT, Aylor K, Yasin M, Marshall JC Gonadotropin-releasing hormone regulation of gonadotropin subunit gene expression in female rats: actions on follicle-stimulating hormone beta messenger ribonucleic acid (mRNA) involve differential expression of pituitary activin (beta-B) and follistatin mRNAs. Endocrinology 1999 140(2):903-8

Dias JA, Cohen BD, Lindau-Shepard B, Nechamen CA, Peterson AJ, Schmidt A Molecular, structural, and cellular biology of follitropin and follitropin receptor. Vitam Horm 2002 64,249-322

Dias JA, Lindau-Sheparn B, Hauer C, Auger I Human follicle-stimulating hormone structure-activity relationships. Biol Reprod 1998 58:1331-1336

Dias JA Endocrinology: fertility hormone in repose. Nature 2005 433:203-204

Evenson DP, Jost LK Sperm chromatin structure assay: DNA denaturability. In Darzynkiewicz Z, Robinson JP, Crissman HA (eds), Flow Cytometry, edn 2, A volume of Methods in Cell Biology. 1994 Academic Press, Orlando, pp. 159-175

Ewing B, Green P Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998 8:186-194

Fan QR, Hendrickson WA Structure of human follicle-stimulating hormone in complex with its receptor. Nature 2005 433, 296-277

Feng W, Matzuk MM, Mounjoy K, Bedows E, Ruddon RW, Boime I The asparagine-linked ologosaccharides of the human chorionic gonadotropin β subunit facilitate correct disulfide bond pairing. J Biol Chem 1995 270:11851-11859

Fox KM, Dias JA, Van Roey P Three-dimensional structure of human follicle-stimulating hormone. Mol Endocrinol 2001 15, 378-389

Fu YX, Li WH Statistical tests of neutrality of mutations. Genetics 1983 133:
693–709

Gordon D, Abajian C, Green P Consed: a graphical tool for sequence finishing. Genome Res 1998 8:195-202

Graham KE, Nusser KD, Low MJ LbetaT2 gonadotroph cells secrete follicle stimulating hormone (FSH) in response to activin. A J Endocrinol 1999 162;R1-R5

Grigorova M, Rull K, Laan M Haplotype structure of *FSHB*, the beta-subunit gene for fertility-associated follicle-stimulating hormone: possible influence of balancing selection. Ann Hum Genet, in press

Jameson JL, Becker CB, Lindell CM, Habener JF Human follicle-stimulating hormone beta-subunit gene encodes multiple messenger ribonucleic acids. Mol Endocrinol 1988 2:806-815

Heckert LL, Griswold MD Expression of follicle-stimulating hormone receptor mRNA in rat testes and Sertoli cells. Mol Endocrinol 1991 5:670-677

Howles CM Role of LH and FSH in ovarian function. Mol Cell Endocrinol 2000 161:25-30

Hsueh AJW, Adasfi EY, Jones PBC, Welsh TH Jr Hormonal regulation of differentiation of cultured ovarian granulosa cells. Endocr Rev 1984 5:76-127

Hudson RR, Kreitman M, Aguade M A test of neutral molecular evolution based on nucleotide data. Genetics 1987 116:153-159

Kawakami S, Fujii J, Winters SJ Follistatin production by skin fibroblasts and its regulation by dexamethasone. Mol Cell Endocrinol 2001 172:157-167

Keutmann HT, Mason KA, Kitzmann K, Ryan RJ Role of β 93-100 determinant loop sequence in receptor binding and biological activity of human luteinizing hormone and chorionic gonadotropin. Mol Endocrinol 1989 3:526-531

Kimura M The Neutral Theory of Molecular Evolution. 1983 Cambridge University Press, Cambridge, MA

Krummen LA, Woodruff TK, DeGuzman G, Cox ET, Baly DL, Mann E, Garg S, Wong WL, Cossum P, Mather JP Identification and characterization of binding proteins for inhinin and activin in human serum and follicular fluids. Endocrinology 1993 132,431-443

Kumar TR, Wang Y, Lu N, Matzuk MM Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. Nat Genet 1997 15:201-204

Kumar TR, Schuff KG, Nusser KD, Low MJ Gonadotroph-specific expression of the human follicle stimulating hormone β gene in transgenic mice. Mol Cell Endocrinol 2006 247:103-115

Kumar TR, Agno J, Janovick JA, Conn PM, Matzuk MM Regulation of *FSHB*eta and GnRH receptor gene expression in activin receptor II knockout male mice. Mol Cell Endocrinol 2003 212:19-27

Kumar TR, Palapattu G, Wang P, Woodruff TK, Boime I, Byrne MC, Matzuk MM Transgenic Models to Study Gonadotropin Function: The Role of Follicle-Stimulating Hormone in Gonadal Growth and Tumorigenesis. Mol Endocrinol 1999 13(6):851-865

Lamminen T, Jokinen P, Jiang M, Pakarinen P, Simonsen H, Huhtaniemi I Human FSH{beta} subunit gene is highly conserved. Mol Hum Reprod 2005 11:601-605

Lapthorn AJ, Harris DC, Littlejohn A, Lustebader JW, Canfield RE, Machin KJ, Morgan FJ, Isaacs NW Crystal structure of human chorionic gonadotropin. Nature 1994 369, 455-461

Layman LC, Lee EJ, Peak DB, Namnoum AB, Vu KV, van Lingen BL, *et al*. Delayed puberty and hypogonadism caused by mutations in the follicle-stimulating hormone beta-subunit. N Engl J Med 1997 337:607-611

Layman LC, Porto AL, Xie J, da Motta LD, Weiser W, *et al*. FSH beta gene mutations in a female with partial breast development and a male sibling with normal puberty and azoospermia. J Clin Endocrinol Metab 2002 87:3702-3707

Layman LC Mutations in the follicle-stimulating hormone-beta (FSH beta) and FSH receptor genes in mice and humans. Semin Reprod Med. 2000;18(1):5-10

Lerchl A, Sotiriadou S, Behre HM, Pierce J, Weinbauer GF, Kliesch S, Nieschlag E Restoration of spermatogenesis by follicle-stimulating hormone despite low intratesticular testosterone in photoinhibited hypogonadotropic djungarian hamster (*P. surgorus*). Biol Reprod 1993 49:1108-1117

Li MD, Ford J A comprehensive evolutionary analysis based on nucleotide and amino acid sequences of the α- and β-subunits of glycoprotein hormone gene family. Journ of Endocr 1998 156, 529-542

Mason AJ, Pitts SL, Nikolics K, Szonyi E, Wilcox JN, Seeburg PH, Stewart TA The hypogonadal mouse: reproductive functions restored by gene therapy. Science 1986 234(4782):1372-8

Matsumoto AM, Karpas AE, Bremner WJ Chronic human chorionic gonadotropin administration in normal men: evidence that follicle stimulating hormone is necessary for the maintenance of quantitative normal spermatogenesis in men. J Clin Endocrinol Metab 1986 62:1184-1190

Matzuk MM, Kumar TR, Bradley A Different phenotypes for mice deficient in either activins or activin receptor type II. Nature 1995 374:356-3560

McLachlan RI, O'Donnel L, Meachem SJ, Stanton PG, de Kretser DM, Pratis K, Robertson DM Identification of specific sites of hormonal regulation in spermatogenesis in rats, monkeys, and man. Recent Progress in Hormone Reseacrh 2002 57;149-179

Miller W, Shafiee-Kermani F, Strahl B, Huang H-J The nature of FSH induction by GnRH. Trends Endocrinol Metab 2002 13(6):257-62

Moudgal NR, Sairam MR In there is true requirement for follicle stimulating hormone in promoting spermatogenesis and fertility in primates? Hum Reprod 1998 13:916-919

Moudgal NR, Murthy GS, Prasanna Kumar KM, Martin F, Suresh R, Medhamurthy R, Patil S, Sehgal S, Saxena BN Responsiveness of human male volunteers to immunization with ovine follicle stimulating hormone vaccine: results of a pilot study. Hum Reprod 1997 12(3):457-63

Moyle WR, Campbell RK Gonadotropins. In: Adashi, E.Y., Rock, J.A, and Rosenwaks, Z (eds) Reproductive Endocrinology, Surgery and Technology. Lippincott-Raven Publishers, Philadelphia 1996, pp 684-724

Nickerson DA, Tobe VO, Taylor SL PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 1997 25:2745-2751

Pernasetti F, Vasilyev VV, Rosenberg SB, Bailey JS, Huang HJ, Miller WL, Mellon PL Cell-specific transcriptional regulation of follicle-stimulating hormone-beta by activin and gonadotroph-releasing hormone in the LbetaT2 pituitary gonadotrope cell model. Endocrinology 2001 142;2284-2295

Phillip M, Arbelle JE, Segev Y, Parvari R Male hypogonadism due to a mutation in the gene for the beta-subunit of follicle-stimulating hormone. N Engl J Med 1998 338:1729-32

Plant TM, Marshall GR The functional significance of FSH in spermatogenesis and the control of its secretion in male primates. Endocr Rev 2001 22:764-786

Raymond M, Rousset F GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J Heredity 1995 86:248-249

Rivier J, Spiess J, McClintock R, Vaughan J, Vale W Purification and partial characterization of inhibin from porcine follicular fluid. Biochem Biophys Res Commun 1985 133,120-127

Rose MP, Gaines Das RE, Balen AH Definition and measurement of follicle stimulating hormone. Endocr Rev 2000 21:5-22

Rosenfeld MG, Briata P, Dasen J, Gleiberman AS, Kioussi C, Lin C, O'Connell SM, Ryan A, Szeto DP, Treier M Multistep signaling and transcriptional requirements for pituitary organogenesis *in vivo*. Recent Progress in Hormone Research 2000 55:1-13

Rozas J, Rozas R DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 1999 15:174-175

Schneider S, Roessli D, Excoffier L Arlequin: A software for population genetics data analysis. Ver 2.000 edn, Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, 2000

Shetty J, Marathe GK, Dinghe RR Specific immunoneutralization of FSH leads to apoptotic cell death of pachytene spermatocytes and spermatogonial cells in the rat. Endocrinology 1996 137:2179-2182

Simonsen KL, Churchill GA, Aquadro CF Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 1995 141:413-429

Singh J, O'Neill C, Handelsman DJ Induction of spermatogenesis by androgens in gonadotropin-deficient (hpg) mice. Endocrinology 1995 136(12):5311-21

Spady TJ, Shayya R, Thackray VG, Ehrensberger L, Bailey JS, Mellon PL Androgen regulates follicle-stimulating hormone beta subunit gene expression in an activin-dependent manner in immortalised gonadotropes. Mol Endocrinol 2004 18:925-940

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han J-H, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale

CM, Nandabalan K, Judson RS, Ruano G, Vovis GF Haplotype variation and linkage disequilibrium in 313 human genes. Science 2001 293:489-493

Suszko MI, Denise JLO, Suh H, Camper SA, Woodruff TK Regulation of the rat follicle-stimulating hormone β-subunit promoter by activin Mol Endocrinol 2003 17(3):318-332

Tajima F Evolutionary relationship of DNA sequences in finite populations. Genetics 1983 105:437-460

Tajima F Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989 123: 585–595

The International HapMap Project. Nature 2003 426:789-96

Themmen APN, Huhtaniemi IT Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. Annu Rev Biochem 2000 21, 551-583

Ulloa-Aguirre A, Timossi C, Damian-Matsumura P, Dias JA Role of glycosylation in function of follicle-stimulating hormone. Endocrine 1999 11:205-215
Winters SJ, Moore JP Intra-pituitary regulation of gonadotrophs in male rodents and primates. Reproduction 2004 128:13-23

Verrelli BC, Tishkoff SA Signatures of selection and gene conversion associated with human color vision variation. Am J Hum Genet 2004 75:363-375

Watterson GA On the number of segregating sites in genetical models without recombination. Theor Popul Biol 1975 7:256-276

Weiss J, Guendner MJ, Halvorson LM, Jameson JL Transcriptional activation of the follicle-stimulating hormone beta-subunit gene by activin. Endocrinology 1995 136:1885-1891

Xing W, Krishnamurthy H, Sairam MR Role of follitropin receptor signaling in nuclear protein transitions and chromatin condensation during spermatogenesis. Biochem Biophys Res Commun 2003 312(3):697-701

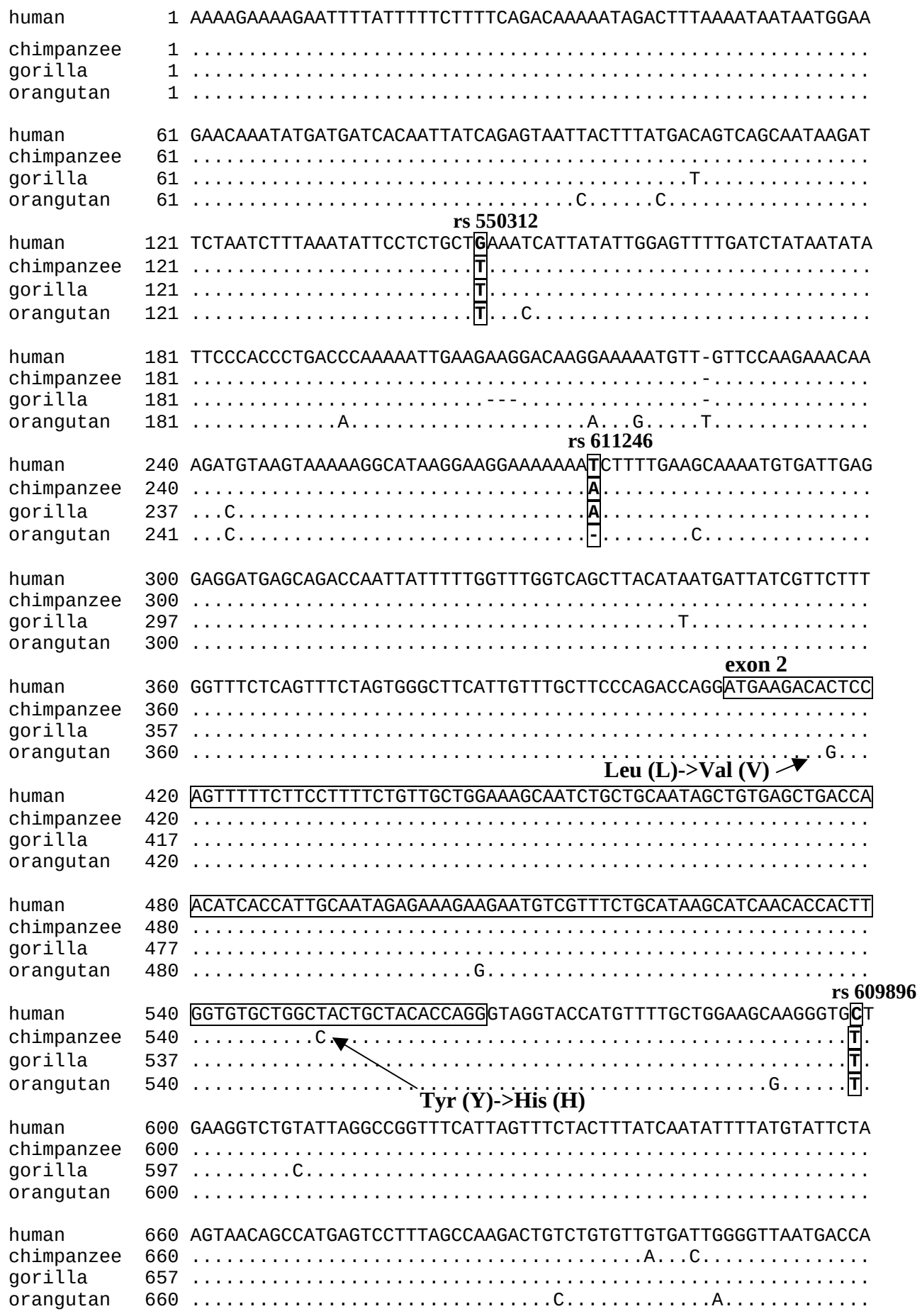
Zhang J, Rowe WL, Clark AG, Buetow KH Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. Am J Hum Genet 2003 73:1073-81

Supplementary information

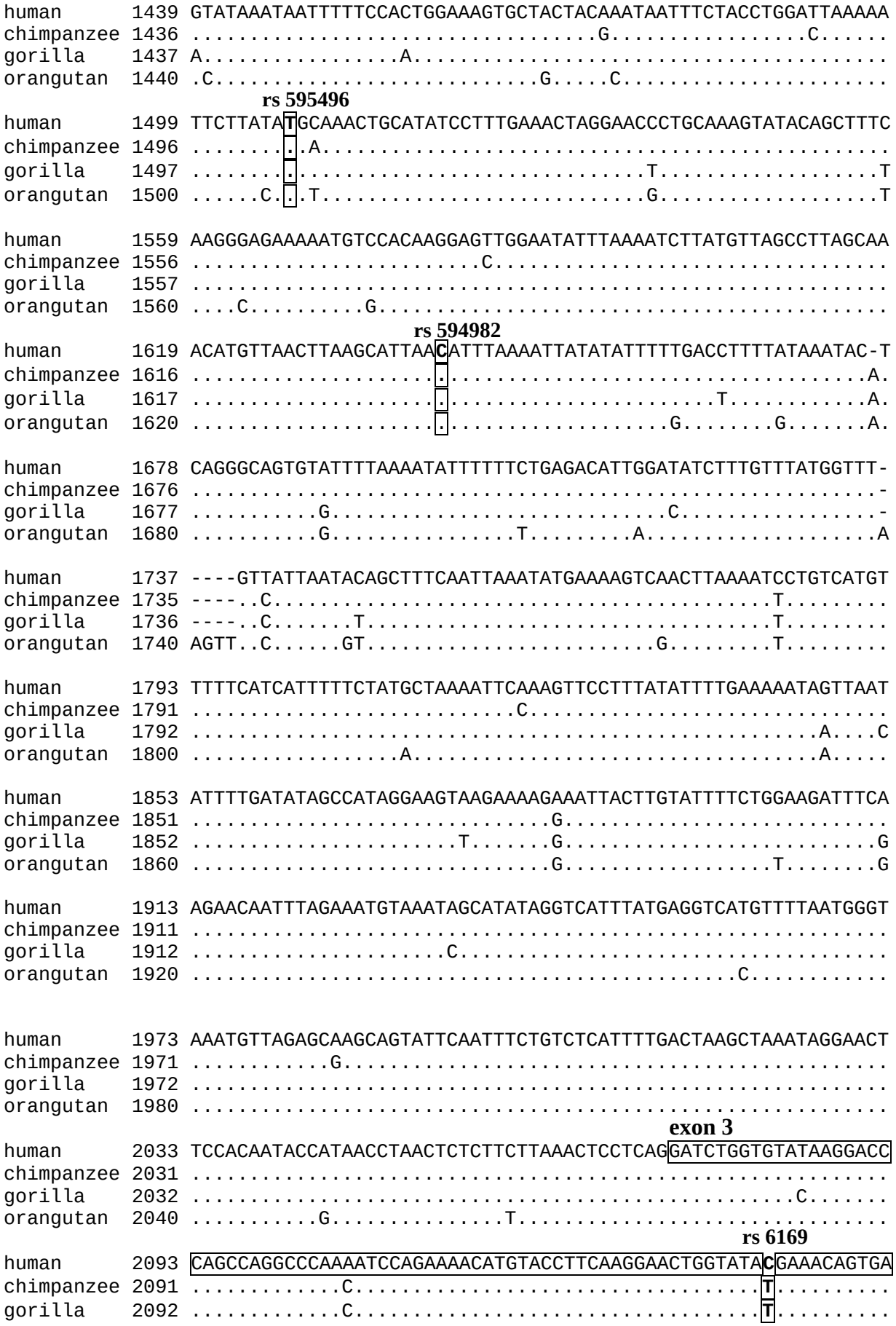
Table S1 Primers for *FSHB* PCR amplification and sequencing reactions

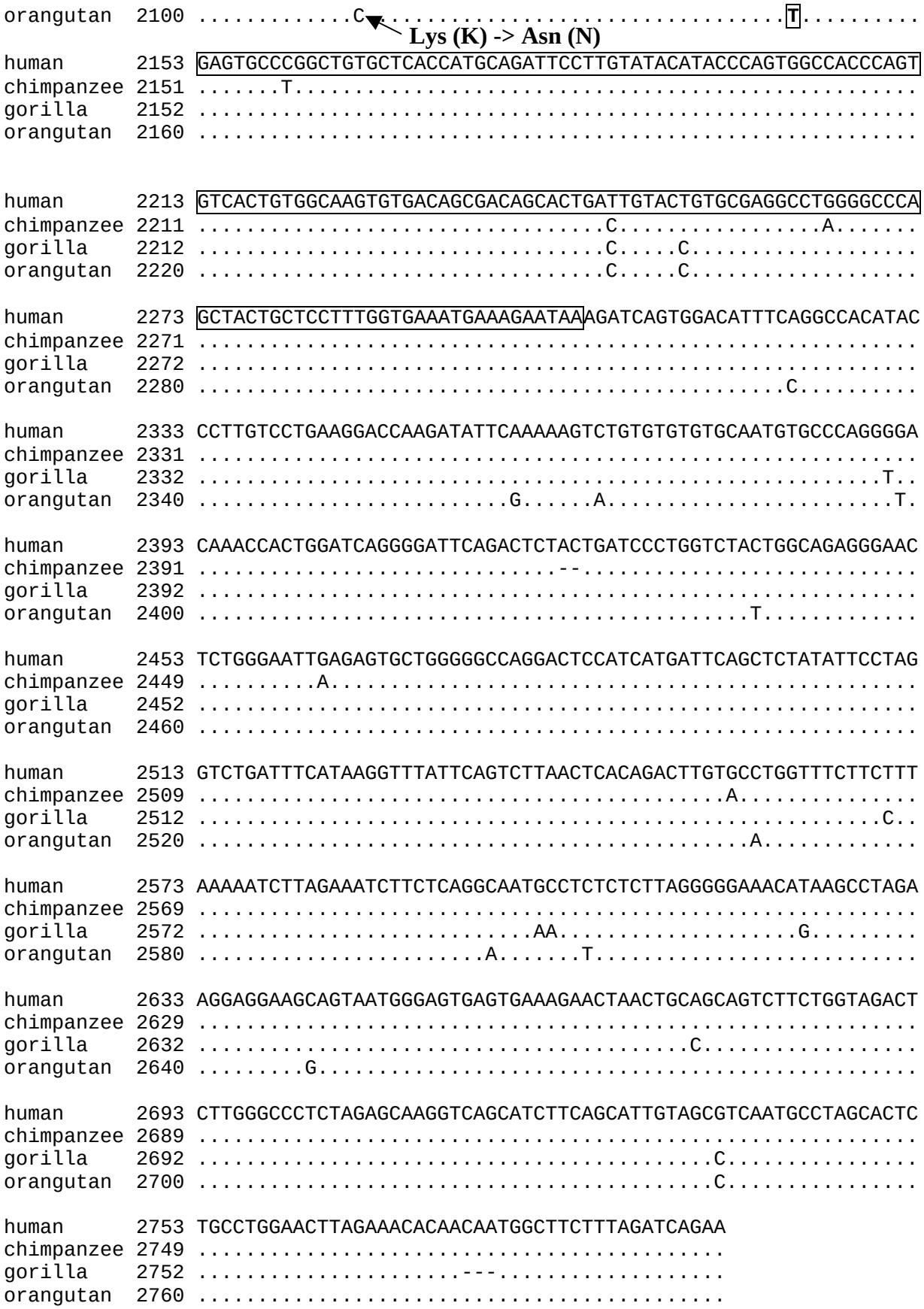
Primer	Sequence
*FSHB*_1AF	5’ GGTTGAAGATGCAGTGATTC 3’
*FSHB*_1AR	5’ ACACAGACAGTCTTGGCTAAAGG 3’
*FSHB*_2AF	5’ ATTGTTTGCTTCCCAGACCAG 3’
*FSHB*_2AR	5’ CCCAAGAAGGTGCAATAAATTGTG 3’
*FSHB*_3AF	5’ CAGTTGCTAGTCTGTGTTTGCAG 3’
*FSHB*_3AR	5’ CCTTGAAGGTACATGTTTTCTGG 3’
*FSHB*_4AF	5’ GCCATAGGAAGTAAGAAAAGA 3’
*FSHB*_4AR	5’ TACCTCAAACATCGTCTTCCAGT 3’
*FSHB*_2.1Rseq	5’ GGGCCAGGAAGACAATTTTT 3’
*FSHB*_2.2Fseq	5’ TTGGGCTTGGATTTGATTTG 3’
*FSHB*_3.1Rseq	5’ TTCCAACTCCTTGTGGACATT 3’
*FSHB*_3.2Fseq	5’ CCCCAATTTCTCTACGCAG 3’
*FSHB*_4.1Rseq	5’ GAGTCTGAATCCCCTGATCC 3’
*FSHB*_4.2Fseq	5’ CCAGCTACTGCTCCTTTGGT 3’
*FSHB*_2.1Rseq	5’ GGGCCAGGAAGACAATTTTT 3’
*FSHB*_2.2Fseq	5’ TTGGGCTTGGATTTGATTTG 3’
*FSHB*_3.1Rseq	5’ TTCCAACTCCTTGTGGACATT 3’

Figure S2. DNA sequence alignments of human, chimpanzee, gorilla and orangutan *FSHB* gene. Location of exons, identified SNPs and fixed differences among species causing non-synonmous changes are indicated.


rs 550312
rs 611246
exon 2
Leu (L)->Val (V)
rs 609896
Tyr (Y)->His (H)

```
human       720 CGATATCACTTAGATGTTTGGGCTTGGATTTGATTTGGGTAAATTTAGGAAAGCCTCAGA
chimpanzee  720 T............................................................
gorilla     717 T............................................................
orangutan   720 T...........................................................G

human       780 TTTAATCTGATCAATTTGGTACTAGTCCAACTTTGCATCTACAGGGAAAAAGTATTTCTA
chimpanzee  777 ...................................T............................
gorilla     777 ...................................T......A.....................
orangutan   780 ...C.................A.............T......A.....................

human       840 TGTTACGTTTTTACACATAGAGAGATAAACATGGAAACATACATATATTTAATCATAAAG
chimpanzee  837 ..............T.....................................................
gorilla     837 ..............T.....................................................
orangutan   840 .....T.........T.....................C.............................

human       900 GACCTATAATATTCTCATAAAGGCAATTTCTTTAACTGACACTACATCTTTGACACAAAA
chimpanzee  897 ..............C.....................................................
gorilla     897 ............................................................
orangutan   900 ......C.G.............................TG.............................

human       960 ATCACACCAAAATATGTCTCCAAGTCACATAAAAACATAGACAGCCACTTAAAAAAATTG
chimpanzee  957 ...................................C.............T.........
gorilla     957 ...................................C.....................
orangutan   960 ....T...............................C........C.............

human      1019 TCTTCCTGGCCCTACTAAATACAAATGCCAAAAAACAGCCTGAGAACACAATCAATTCTT
chimpanzee 1017 ............................................................
gorilla    1017 ............................................................
orangutan  1020 ............................C...T.....................

human      1079 GCAGACTGTTAGAACAAAAATGAATCAGCAAACCCACTCCCTTCGTTATAGCATTGAGAA
chimpanzee 1077 ..........................................A.................
gorilla    1077 ............................................................
orangutan  1080 ..........................................A.................

human      1139 AACCAAGACATAGAGGCATCAGTTGCTAGTCTGTGTTTGCAGTTTCCTTGCATTAATACA
chimpanzee 1137 ............................................................
gorilla    1137 ............................................................
orangutan  1140 ............................................................
                                          rs 580646
human      1199 AGTAGAGAAATAGTTTCCATGGTGCTGTCTTTTTTCTCTGCAGCACCCCTAATTATCTAT
chimpanzee 1197 ........................T...........................................
gorilla    1197 ........................T.............A.............................
orangutan  1200 ....A...................T...........................................

human      1259 GCAGAATTTCATTCTATAAACTAAAATTGAAAATGGCAACTTTTTAAATGAACGATACTT
chimpanzee 1257 ..................T....-.....................................
gorilla    1257 ..............C.....................................................
orangutan  1260 ..............G.........C.....................T.......


human      1319 TATTTGACGGTAAATGAGTTTGATCAAACTCCATTTATTACACAATTTATTGCACCTTCT
chimpanzee 1316 ............................................................
gorilla    1317 ............................................................
orangutan  1320 C......T.....................................................

human      1379 TGGGATATACATTTGGTAGGATGATATTAAAATAAACAGAAGCCCCAATTTCTCTACGCA
chimpanzee 1376 ..........................................G.............
gorilla    1377 ..................................C.............T......
orangutan  1380 ........................C.........C.............T..T...
```


rs 595496
rs 594982
exon 3
rs 6169


Lys (K) -> Asn (N)