THE MOLECULAR GENETICS OF POLYCYSTIC OVARY SYNDROME

by

Neda Gharani

A thesis submitted for the degree of

Doctor of Philosophy

in the

University of London

January 2000

Department of Reproductive Science and Medicine,

Imperial College School of Medicine,

St Mary's Hospital,

London W2 1PG.

LONDON UNIV.

To mum, dad and my sisters

Nona and Nargess

ABSTRACT

Polycystic ovary syndrome (PCOS) is a common endocrinopathy affecting up to 10% of women of reproductive age. PCOS is characterised by hyperandrogenaemia and is responsible for over 70% of cases of anovulatory infertility and 85% of cases of idiopathic hirsutism. PCOS is a familial disorder with a complex genetic basis.

In this thesis, a candidate gene approach was undertaken to study the molecular genetics of PCOS. The presence of hyperandrogenaemia as the most consistent finding in women with PCOS suggests an underlying disorder of androgen biosynthesis or metabolism. Therefore, investigations were focused on the genes of key components in the androgen biosynthesis/metabolism pathway in the ovary.

Linkage analyses were performed on three genes: *CYP11a* (encoding the rate-limiting enzyme P450 side chain cleavage), *CYP19* (the gene for P450 aromatase) and the luteinizing hormone (LH) receptor gene (*LHR*), using 23 multiply affected PCOS/male pattern baldness (MPB) pedigrees. Five microsatellite markers from the *CYP19* locus on chromosome 15q21.1 and six microsatellite markers spanning the *LHR* locus (chromosome 2p21) were used to obtain clear evidence for exclusion of a major disease gene from these two chromosomal regions.

In contrast, significant evidence for both linkage and association was found at the *CYP11a* locus. Linkage analysis provided evidence for linkage at the *CYP11a* locus, present on chromosome 15q24. Mutation screening of the promoter region of *CYP11a*, revealed the polymorphic nature of a pentanucleotide repeat microsatellite marker, *CYP11a* $(tttta)_n$, present in this region, but failed to identify any other functional mutations. The *CYP11a* $(tttta)_n$ was subsequently used to carry out an association study in a case control data set of Europid women with PCOS and matched controls. A significant association between allelic variants of this gene and total serum testosterone level was observed.

In a previous study, our group had demonstrated association, but not linkage, of a 5' polymorphism of the gene *CYP17* (encoding 17 α-hydroxylase, 17-20 Lyase) with PCOS. This polymorphism was further evaluated using the case control data set. No association between this polymorphism and serum testosterone level was found, furthermore, the original association with PCOS was no longer observed.

Finally, the gene *SRD5A1*, encoding the enzyme 5 α-reductase type 1, was investigated for a role in the aetiology of hirsutism in PCOS. Two single base change polymorphisms, present in the coding sequence of this gene, were assessed in hirsute and non-hirsute subjects from the case control data set. Preliminary results of this study have shown association between allelic variants of this gene and hirsutism.

In summary, significant evidence for both linkage and association at the *CYP11a* gene has been demonstrated, suggesting that variation in *CYP11a* plays an important role in the aetiology of the hyperandrogenaemia associated with PCOS. *CYP11a* is one of the first susceptibility genes identified for PCOS. Furthermore, this study has demonstrated that *SRD5A1* may have a role in the cutaneous manifestation of hyperandrogenaemia and that the genes for the *LHR*, *CYP19* and *CYP17* may be excluded as major determinants of risk for PCOS.

ACKNOWLEDGEMENTS

I would like to start by thanking my supervisors Steve Franks and Mark McCarthy for all their guidance and enthusiasm throughout this project and for their continual encouragement and support. I would also like to express my gratitude to Bob Williamson who, together with Steve, gave me the opportunity to start this PhD and who found me the perfect substitute supervisor (Mark) before leaving for Australia.

My time at St Mary's and especially in the PCO group has been very special and enjoyable, largely due to the great people I have worked with. Firstly, I want to thank Dawn Waterworth for all her help in the lab when I first started, our inspiring discussions and for her friendship. I also want to say a big thanks to Brien Riley, Cemal Cemal and others from the Biochemistry Department; everyone from Obs. and Gynae.: Debbie Willis, Hazel Watson, Sari Batty, Helen Mason, Ester Cela, Lisa Webber, Phil Bates, Anne and Zoshia and all the people I have worked with over the last year in the CRB at the Hammersmith Hospital - Stewart Huxtable, Lema Haddad, Tasha Weerasekera, Cath Williamson, Eleni Kousta, Phil Saker, Santiago Lewis and Sasha Saveliev, for their help, advice, generosity, friendship and lively gossips. Special thanks go to Stew, for making me laugh, broadening my musical horizons and for cooking me such delicious meals.

Thanks also to Marjan and Mahmoud for lending me their computer to write up this thesis and to Nargess, Shahab and Shefali for putting up with it taking up most of the kitchen table for so long.

Most of all I want to thank my parents Massoud and Lili and my sisters Nona and Nargess for their constant love, encouragement and support.

CONTENTS

Title		1
Dedication		2
Abstract		3
Acknowledgements		5
Table of Contents		6
List of Figures		11
List of Tables		13
Abbreviations		14
1.0	**INTRODUCTION**	17
1.1	**BACKGROUND**	17
	1.1.1 History	17
	1.1.2 Symptomatology and prevalence	19
	1.1.3 Diagnosis and differential diagnosis	20
	1.1.4 Endocrine features	21
	1.1.5 Metabolic abnormalities	23
	Insulin resistance and PCOS	24
	β-cell dysfunction in PCOS	25
	Insulin and androgens	26
1.2	**POLYCYSTIC OVARIAN MORPHOLOGY**	26
	1.2.1 Folliculogenesis in the normal ovary	27
	Initiation and growth	27
	Selection and maturation	27
	Ovulation	30
	1.2.2 Altered folliculogenesis in the polycystic ovary	31
1.3	**CURRENT TREATMENTS**	32
	1.3.1 Anovulation	32
	1.3.2 Cutaneous manifestations of hyperandrogenism	33
	1.3.3 Improvement of insulin sensitivity	34
1.4	**ISSUES IN PATHOGENESIS**	36
	1.4.1 Is the abnormality central or peripheral ?	36
	Primary hypothalamic defect	36
	Primary role for androgens	37
	1.4.2 Ovarian or adrenal source of hyperandrogenism ?	38

1.5 GENETICS OF PCOS 40
1.5.1 Family studies of PCOS 41
1.5.2 Current hypothesis 43

1.6 INSULIN GENE *VNTR* AND PCOS 44

1.7 CANDIDATE GENES FOR PCO / PCOS 45
1.7.1 LH receptor 46
1.7.2 *CYP11a* 46
1.7.3 *CYP17* 47
1.7.4 *CYP19* 48
1.7.5 *SRD5A1* 49

1.8 EXPERIMENTAL STRATEGY 50
1.8.1 Linkage analysis in pedigree data set 51
Definition of linkage 51
Parametric linkage analysis 52
Markers for linkage analysis 53
Non-parametric linkage analysis 54
GENEHUNTER 56
Options for parametric analysis 58
1.8.2 Case control association studies 59

2.0 SUBJECTS, MATERIALS AND METHODS 61

2.1 SUBJECTS 61
2.1.1 Pedigree data set 62
Diagnostic criteria 62
Pedigree data set size and structure 63
2.1.2 Case control data set 71

2.2 MATERIALS 72
2.2.1 Markers for linkage studies 72
Chromosome 15 72
Chromosome 2 74
2.2.2 Chemicals and reagents 75
2.2.3 Solutions 76

2.3 METHODS 79
2.3.1 Extraction of high molecular weight DNA from human nucleated cells 79

2.3.2 DNA quantification 79
2.3.3 Polymerase chain reaction (PCR) 80
Microsatellite markers 80
RFLP markers and LHR 1a/1b PCR 81
PCR Conditions for S1,S2 and S3 (CYP11a promoter primers) 82
2.3.4 Polyacrylamide gel electrophoresis 82
2.3.5 Agarose gel electrophoresis 83
2.3.6 Mutation screening of *CYP11a* promoter region 84
2.3.7 Purification of PCR product 86
2.3.8 Preparation of single-stranded DNA 86
2.3.9 Sequencing analysis 87
2.3.10 Stanford G3 radiation hybrid panel screening 88
2.3.11 *CYP11a* $(tttta)_n$ genotyping 90
2.3.12 *CYP17* assay 90
2.3.13 *SRD5A1* RFLP assays 91
SRD5A1 NspI assay 91
SRD5A1 Hinf I assays 92

2.4 COMPUTER PROGRAMS AND STATISTICAL ANALYSES 93
2.4.1 MLINK 93
2.4.2 Model parameters 93
Model parameters used for the parametric analyses 94
2.4.3 GENEHUNTER 95
Input parameters used for all GENEHUNTER analyses 95
Key to GENEHUNTER output results 96
Interpretation of GENEHUNTER results 96
2.4.4 Power limitations of parametric and non-parametric linkage analysis 97
2.4.5 Case control association studies 98
Statistical analyses 98
Limitations of case control studies 99
2.4.6 z-Test 100
2.4.7 M-test 101
2.4.8 Bonferroni correction 102
2.4.9 Measure of linkage disequilibrium 102

3.0 RESULTS 104

3.1 CHROMOSOME 15q 104
3.1.1 *CYP19* locus linkage results 104
3.1.2 *CYP11a* locus linkage results 115
3.1.3 *CYP11a* promoter screening 125
3.1.4 *CYP11a* (tttta)n association results 127

3.2 CHROMOSOME 2p 132
3.2.1 Radiation Hybrid panel screening with *LHR* 132
3.2.2 Chromosome 2 linkage analysis 136

3.3 *CYP17* ASSOCIATION STUDY 146

3.4 *SRD5A1* ASSOCIATION RESULTS 149

4.0 DISCUSSION 155

4.1 EVALUATION OF LINKAGE RESULTS 155
4.1.1 Assessment of the PCOS families 155
4.1.2 Assessment of the linkage approach taken 157

4.2 *CYP19* 159
4.2.1 Evaluation of linkage results 159
4.2.2 Significance of *CYP19* results 159

4.3 *CYP11a* 160
4.3.1 Evaluation of linkage results 160
4.3.2 Mutation screening of *CYP11a* promoter 162
4.3.3 *CYP11a* $(tttta)_n$ evaluation in the case control data set 162
4.3.4 What is the role of *CYP11a* $(tttta)_n$ 163
Is the aetiological mutation elsewhere in the CYP11a gene ? 164
Is there a function for CYP11a (tttta)n ? 165
4.3.5 Expression studies on *CYP11a* promoter 167
Interaction between CYP11a promoter elements 170
4.3.6. DNA bending and gene transcription 170
4.3.7 Is there a role for *CYP11a* (tttta)n in DNA bending ? 172
Does CYP11a (tttta)n fit the A-tract model ? 174
Proposed role of CYP11a (tttta)n in transcriptional regulation 176
4.3.8 Evidence for general upregulation of steroidogenic enzymes 178

4.4 ***LHR*** 180
4.4.1 Evaluation of linkage results 180
4.4.2 Significance of *LHR* gene results 180

4.5 ***CYP17*** 181
4.5.1 Objective of the follow-up study 181
4.5.2 Outcome of this extended study 181
4.5.3 Significance of *CYP17* results 182

4.6 ***SRD5A1*** 183
4.6.1 Summary of association results 183
4.6.2 Interpretation of the data 184
4.6.3 Assessment of the role of *SRD5A1* in hirsutism 185

4.7 **FUTURE WORK** 186
4.7.1 Sample collection 187
Sibling-pairs 187
Parent-offspring trios 188
Detailed phenotypic and biochemical characterisation 188
4.7.2 Experimental strategies 189
4.7.3 *CYP11a* 191
Polyacrylamide gel mobility analysis 192
Expression studies 194
4.7.4 *SRD5A1* 196

4.8 **CONCLUSIONS** 197

5.0 **REFERENCES** 200

6.0 **PUBLICATIONS** 230

LIST OF FIGURES

1.1	Section through a polycystic ovary	18
1.2	Folliculogenesis	28
1.3	Two gonadotrophin-two cell type model of oestrogen synthesis	30
1.4	Androgen biosynthesis/metabolism in the ovary	39
2.1	Pedigree trees	65
2.2.1	Physical and genetic maps of chromosome 15	72
2.2.2	Physical and genetic maps of chromosome 2	74
2.3.6	Sequence of 5' region of human *CYP11a*	85
3.1.1	Autoradiograph of *CYP19*(ttta)n genotypes	104
3.1.2	Information plot for the *CYP19* locus linkage results	107
3.1.3	NPL plot for the *CYP19* locus linkage results	108
3.1.4	LOD plots for the *CYP19* locus linkage results	109
3.1.5	*CYP19* locus haplotypes in the PCOS/MPB pedigrees	110
3.1.6	Autoradiograph of *CYP11a*(ac)n genotypes	115
3.1.7	Information plot for the *CYP11a* locus linkage results	117
3.1.8	NPL plot for the *CYP11a* locus linkage results	118
3.1.9	LOD plots for the *CYP11a* locus linkage results	119
3.1.10	*CYP11a* locus haplotypes in the PCOS/MPB pedigrees	120
3.1.11	Sequence of S3 fragment showing presence of the *CYP11a* (tttta)n repeat	126
3.1.12	Autoradiograph of *CYP11a* (tttta)n genotypes	127
3.1.12	Association of *CYP11a* with geometric mean of total serum Testosterone levels in the case control data set	129
3.2.1	Screening of G3RH panel with the *LHR* 1a/1b primer set	13 3
3.2.3	Information plot for the chromosome 2 linkage results	138
3.2.4	NPL plot for the chromosome 2 linkage results	139
3.2.5	LOD plots for the chromosome 2 linkage results	140
3.2.2	Radiation hybrid mapping of the *LHR*	135

3.2.6 *LHR* locus haplotypes in the PCOS/MPB pedigrees. 141

3.3.1 Photograph of gel showing *CYP17 MspA1* RFLP genotypes 146

3.3.2 Comparison of the geometric mean of total serum testosterone levels in the three CYP17 genotype groups for the "Total group" and for the "PCO group" alone 147

3.4.1 Photograph of gel showing *SRD5A1 NspI* RFLP genotypes 149

3.4.2 Photograph of gel showing *SRD5A1 HinfI* RFLP genotypes 151

4.1 Schematic representation of the *CYP11a* promoter region 168

4.2 Bend site at the *Leishmania tarentolae* kinetoplast DNA 173

LIST OF TABLES

2.1.1 Summary of affection status in members of the pedigrees 64

2.1.2 Summary of symtomatology in the case control data set 71

2.2.1 Chromosome 15 microsatellite marker information 73

2.2.2 Cnromosome 2 microsatellite marker information 74

2.3.6 Summary of information of the PCR fragments S1, S2 and S3 84

3.1.1 Recombination data for the *CYP19* locus linkage results 105

3.1.2 GENEHUNTER results for *CYP19* locus 106

3.1.3 Recombination data for the *CYP11a* locus linkage results 115

3.1.4 GENEHUNTER results for *CYP11a* lucus 116

3.1.5 Comparison of *CYP11a* genotype distribution between different sub-groups in the case control data set 130

3.2.1 Recombination data for the *LHR* locus linkage results 136

3.2.2 GENEHUNTER results for *LHR* locus 137

3.3.1 Results of screening for the presence of the A2 alleles 148

3.4.1 *SRD5A1 NspI* RFLP genotype distribution in the hirsute and non-hirsute groups 150

3.4.2 *SRD5A1 HinfI* RFLP genotype distribution in the hirsute and non-hirsute groups 151

3.4.3 *HinfI/NspI* haplotype distribution in the total data set 152

3.4.4 *HinfI/NspI* haplotype distribution in the hirsute and non-hirsute groups 153

4.3 *CYP11a* (tttta)n genotype and allele distributions in the top and bottom quartiles of Ln total serum testosterone 176

ABBREVIATIONS

ACTH	Adrenocorticotrophic hormone
AFBAC	Affected family-based controls
anov	Anovulatory
aPCO	asymptomatic woman with polycystic ovaries
APM	Affected pedigree member
Apo(a)	Apolipoprotein (a)
AR	Androgen receptor
ASP	Affected sibpair
BMI	Body mass index (kg/m^2)
bp	Base pair
BSA	Bovine serum albumin
χ^2	Chi square
CAH	Congenital adrenal hyperplasia
cAMP	Cyclic adenosine monophosphate
CAT	Chloramphenicol acetyltransferase
cM	Centimorgans
CRE	cAMP-response element
CVD	Cardiovascular disease
Δ	Linkage disequilibrium correlation coefficient
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddATP	Dideoxyadenosine triphosphate
ddCTP	Dideoxycytidine triphosphate
ddGTP	Dideoxyguansosine triphosphate
ddTTP	Dideoxythymidine triphosphate
DHEAS	Dehydroepiandrosterone-sulphate
dGTP	Deoxyguansosine triphosphate
DHT	Dihydrotestosterone

DNA	Deoxyribonucleic acid
d.s.	Double-stranded
dTTP	Deoxythymidine triphosphate
EDTA	Ethylene diamine tetra-acetic acid
FSH	Follicle stimulating hormone
Gn RH	Gonadotrophin-releasing hormone
GRR	Genotype relative risk
HLOD	LOD score under a test for heterogeneity
HOTT	Human ovarian thecal-like tumour cells
IBD	Identical by descent
IBS	Identical by state
IGF	Insulin-like growth factor
INS	Insulin gene
K4	Kringle 4
Kb	Kilobase pairs
LD	Linkage disequilibrium
LH	Luteinizing hormone
LHR	Luteinizing hormone receptor gene
LOD	Logarithm of the odds
Lp(a)	Lipoprotein (a)
λs	Sibling relative risk
MPB	Male-pattern baldness
mRNA	Messenger RNA
NIDDM	Non-insulin dependent diabetes mellitus
NPL	Non-parametric linkage
OD	Optical density
OR	Odds ratio
P450c17α	Cytochrome P450 17α-hydroxylase and 17,20 lyase enzyme
P450scc	Cytochrome P450 side chain cleavage enzyme
P_c	Bonferroni corrected p-value

PCO	Polycystic ovaries
PCOS	Polycystic ovary syndrome
PCR	Polymerase chain reaction
PIC	Polymorphism information content
pm	Postmenopausal
RFLP	Restriction fragment length polymorphism
RH	Radiation hybrid
SDS	Sodium dodecyl sulphate
SF-1	Steroidogenic factor 1
SHBG	Sex hormone-binding globulin
SL	Single locus
SNP	Single-nucleotide polymorphism
s.s.	Single-stranded
StAR	Steroidogenic acute regulatory protein
θ	Recombination fraction, "Theta"
Taq	Taq polymerase from *Thermus aquaticus*
TDT	Transmission disequilibrium test
TAE	Tris-Acetate EDTA
TBE	Tris-Borate EDTA
VNTR	Variable number tandem repeats

1.0 INTRODUCTION

1.1 BACKGROUND

1.1.1 History

Polycystic ovary syndrome (PCOS) is a highly prevalent endocrine disorder affecting up to 10% of women of reproductive age (Dunaif, 1992b; Franks, 1995). It is characterised by hyperandrogenaemia (raised serum androgen levels), and represents the most common cause of anovulatory infertility and hirsutism (excessive growth of terminal hair in a male-like pattern) (Adams *et al.*, 1986; Hull, 1987).

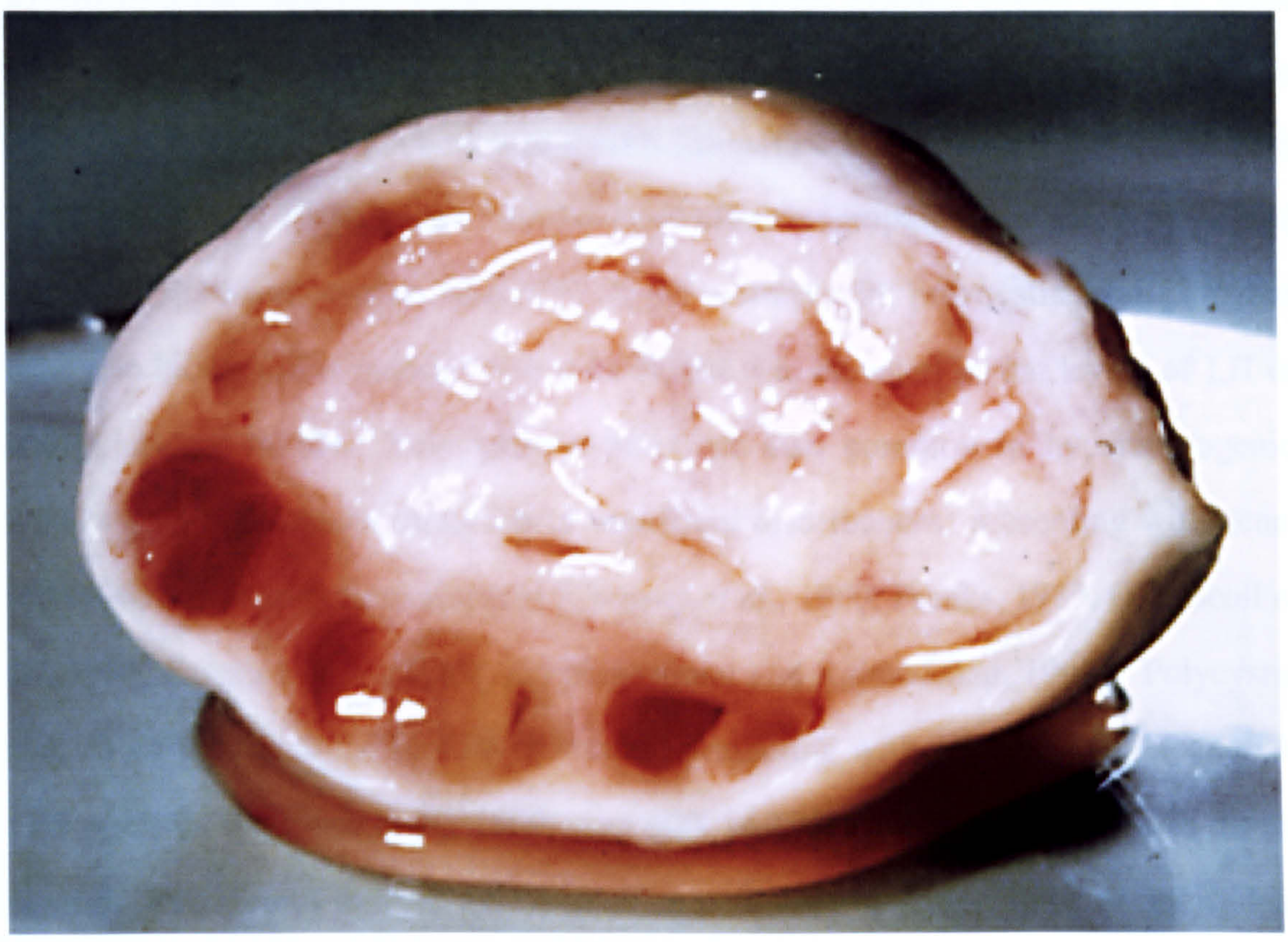
PCOS was first described by Stein and Leventhal in 1935 who noted the association of menstrual disturbances, hirsutism, obesity and infertility with bilaterally enlarged polycystic ovaries (Stein and Leventhal, 1935). Since then it has been recognised that there is considerable clinical and biochemical heterogeneity associated with the disorder, ranging from an incidental finding of polycystic ovaries on pelvic ultrasonography in an asymptomatic subject, to the classic manifestation of hyperandrogenic anovulation (Franks, 1989). Between these two extremes are women with anovulatory infertility without hirsutism and hirsute women with regular, ovulatory cycles. The biochemical markers associated with PCOS include raised serum concentrations of testosterone (and/or androstenedione), hypersecretion of luteinizing hormone (LH), and hyperinsulinaemia in anovulatory PCOS women (Franks, 1995).

This broad spectrum of presentation has led to controversy regarding the clinical diagnosis of the syndrome, and fuelled debate as to the relative importance of various diagnostic criteria such as the presence or absence of anovulation, raised serum LH or the polycystic ovarian morphology.

It was not until the introduction of high-resolution pelvic ultrasound scanning (which provided a non-invasive means of defining the typical ovarian morphology) that this issue could be properly addressed, by allowing more extensive investigations into the

prevalence of polycystic ovaries and associated symptoms. A polycystic ovary has been defined as one which contains 10 or more cysts, 2-8mm in diameter arranged peripherally in the ovary with an increase in the ovarian stroma (Adams *et al*., 1985; Adams *et al*., 1986) (**figure 1.1**).

FIGURE 1.1 Section through a polycystic ovary.



The polycystic ovary (PCO) has been defined as one which contains, in one plane, at least 10 follicles (each between 2-8mm in diameter) and an increased amount of stroma (Franks, 1989). The follicles are generally peripherally arranged but may also be scattered throughout the stroma (Swanson et al, 1981; Parisi et al., 1984; Adams et al., 1985; Adams et al., 1986). The ovary is usually enlarged with a volume of > 9ml, which is two standard deviations from the mean volume of normal ovaries (Adams et al., 1985).

1.1.2 Symptomatology and prevalence

Several studies based on ultrasound imaging of the ovaries have been carried out in an attempt to evaluate the presentation and prevalence of polycystic ovaries in women with symptoms associated with PCOS (Adams *et al.*, 1986; Franks, 1989; Fox *et al.*, 1991; Conway *et al.*, 1989; O'Driscloll *et al.*, 1994; Bunker *et al.*, 1991). In a study of over 200 women, who presented with either anovulation and/or hirsutism to a gynaecological endocrine clinic (Adams *et al.*, 1986; Franks, 1989), 30% of those with amenorrhea (absent menses), 87% of those with oligomenorrhea (irregular menstrual cycles) and 87% of cases with idiopathic hirsutism (i.e hirsutism with regular menstrual cycles) were found to have ultrasonographic evidence of polycystic ovaries. Moreover, greater than 85% of the women with polycystic ovaries had at least one biochemical marker typical of PCOS (i.e. elevated serum levels of LH or androgens (or both)). Polycystic ovaries are also common in hyperandrogenic women (with or without hirsutism or menstrual disturbances) presenting with acne, seborrhoea or male-pattern alopecia (Conway *et al.*, 1989; Franks, 1989; O'Driscoll *et al.*, 1994; Peserico *et al.*, 1989; Timpatanapong and Rojanasakul, 1997). Polycystic ovaries have been found in 45% of women presenting with acne alone compared to 17% controls (Peserico *et al.*, 1989), and 60% of 109 women presenting with difuse alopecia were found to have evidence of an endocrine disorder, most of whom had PCOS (Futterweit *et al.*, 1988).

Assessment of the population prevalence of polycystic ovaries has revealed that they occur in a significant proportion (22%) of the normal female population (Clayton *et al.*, 1992; Polson *et al.*, 1988). One study of 257 volunteers (none of whom had sought medical attention for gynaecological or dermatological symptoms) found a striking correlation between clinical (and biochemical) features and the presence of polycystic ovaries (Polson *et al.*, 1988). 75% of the women with polycystic ovaries had evidence of an irregular cycle and 45% demonstrated objective evidence of hirsutism (compared to less than 1% and 7% respectively of those with normal ovaries). Overall, greater than 90% of the "normal" women with polycystic ovaries

had at least one symptom considered to be a marker of PCOS, even though they had not sought medical attention.

These studies have highlighted the significance of the polycystic ovarian morphology in the diagnosis of the syndrome. The identification of polycystic ovaries in women with regular menstrual cycles demonstrates that the typical ovarian morphology is not simply a result of a non-specific response to chronic anovulation. Also, the fact that ovulatory women with polycystic ovaries share common biochemical features with those with anovulation and polycystic ovaries suggests that the former group represents a particular presentation of the same underlying disorder. This therefore calls into question the validity of the classic definition of the syndrome (the association of hyperandrogenism with chronic anovulation (Zawadzki and Dunaif, 1992)), which regards the presence of anovulation as an essential requirement and does not consider ovarian morphology for diagnosis.

1.1.3 Diagnosis and differential diagnosis

As discussed above (**section 1.1.2**), the classic definition of PCOS is likely to be too narrow and will inevitably lead to the exclusion of subjects who clearly are part of the same spectrum. Therefore, the approach taken by the group at St Mary's has been to assign women as having "PCO" if they have ultrasound evidence of polycystic ovaries (regardless of presentation and symptomatology). They are assigned as having polycystic ovary *syndrome* (PCOS), if they also (in addition to having polycystic ovaries on ultrasound) have one or more of the associated clinical and or biochemical features such as anovulation, hirsutism or elevated serum androgens and/or LH. Although these form the basic criteria for diagnosis, further sub-classifications may be made according to symptomatology (e.g. hirsute or non-hirsute and ovulatory or anovulatory) which will allow differential analysis during experimental investigations.

The diagnosis of PCOS (or PCO) also involves the exclusion of other "polycystic-ovary-like" syndromes (Franks, 1989; Yen, 1980). These include Cushing's syndrome (a disorder of either the adrenal or pituitary glands), congenital adrenal hyperplasia (CAH) (a disease of the adrenal gland), hyperprolactinaemia (a pituitary disorder) and androgen-secreting tumours (of the ovary or adrenal gland). These hyperandrogenic disorders may present with a PCO morphology and tend to be associated with symptoms of menstrual disturbances and hirsutism, but can be identified by the presence of other, distinctive, clinical and biochemical features.

1.1.4 Endocrine features

The endocrine features associated with PCOS are raised serum androgens (testosterone and/or androstenedione) and LH (Franks, 1995), with normal to low levels of follicle-stimulating hormone (FSH) (Holte *et al.*, 1994b; Yen *et al.*, 1970) and abnormalities of oestrogen secretion (Baird *et al.*, 1977; Polson *et al.*, 1987a; Yen, 1980).

In women with anovulatory PCOS, although serum oestradiol concentrations lie within the early to mid-follicular phase ranges of the normal menstrual cycle, the pattern of secretion differs from the normal cycle in that there is no pre-ovulatory rise in oestradiol levels (refer to **section 1.2.1**). There is also a lack of cyclical progesterone secretion, which leaves the oestrogens unopposed (Yen, 1980). This effect is amplified in obese subjects where there is exaggerated peripheral conversion of androgens to oestrone in adipose tissue. Unopposed oestrogens are thought to be important in the mechanism of chronic anovulation (Baird *et al.*, 1977; Yen, 1980), and both acyclical oestrogen production and progesterone deficiency may contribute to the mechanism of hypersecretion of LH (Kiddy *et al.*, 1989).

Elevated serum LH levels has been reported in both ovulatory and anovulatory women with PCO. However, this is not a universal observation and many women with all the other clinical and biochemical features of PCOS will have normal serum

LH levels. In fact, the observed prevalence of LH hypersecretion in women with PCOS will depend on both the diagnostic criteria as well as the measurement method used by the particular study. Because of this, different groups have reported greatly varying results (Conway *et al.*, 1989; Fauser *et al.*, 1992; Franks, 1989; Polson *et al.*, 1988; Waldstreicher *et al.*, 1988).

The most consistent biochemical finding in all series of women with polycystic ovaries, regardless of their clinical presentation, is an elevation of serum testosterone (and/or androstenedione). The degree of hyperandrogenaemia tends to increase with increasing symptomatology (Franks, 1991). However, as with LH, there is variation between individual women (Conway *et al.*, 1989; Franks, 1989). For example, some anovulatory non-hirsute women with PCO have similar levels of hyperandrogenaemia as hirsute ovulatory PCOS women, and there are hirsute subjects with polycystic ovaries who have normal serum concentrations of androgens. The former situation presumably reflects variable sensitivity of hair follicles to androgens (Barth, 1988; Lobo, 1991), and the latter an increased rate of androgen production and clearance (by peripheral tissue) (Bardin and Lipsett, 1967; Kirschner *et al.*, 1983; Kirschner and Bardin, 1972). Also, the fraction of bioavailable testosterone in, and its clearance from, the circulation is dependent on the serum concentrations of sex hormone-binding globulin (SHBG). SHBG is produced by the liver and its levels are regulated by insulin. There is an inverse relationship between circulating concentrations of SHBG and insulin (Dunaif *et al.*, 1987; Holte *et al.*, 1994b; Kiddy *et al.*, 1989). The higher plasma concentrations of insulin in obese women with PCO will result in lower plasma concentrations of SHBG, which will in turn increase the bioavailability of the free testosterone fraction, potentially leading to greater androgenic effects. Obese women with PCO are more likely to be anovulatory and hirsute than their lean counterparts (Kiddy *et al.*, 1990). In these obese subjects, calorie restriction has been shown to improve symptomatology, both in the degree of hirsutism and in some cases ovulation has even recommenced (Kiddy *et al.*, 1992).

1.1.5 Metabolic abnormalities

Polycystic ovary syndrome is also associated with a characteristic metabolic disturbance that may have important implications for long-term health. There is an increased frequency and degree of hyperinsulinaemia and peripheral insulin resistance (reduced insulin sensitivity) in both lean and obese women with the 'classic' form of PCOS (i.e. those with irregular or anovulatory cycles) compared with weight-matched controls (Chang *et al.*, 1983a; Conway *et al.*, 1990; Dunaif *et al.*, 1987; Dunaif, 1997; Holte, 1996). Up to 20% of young obese women with PCOS have impaired glucose tolerance or non-insulin dependent diabetes (NIDDM) (Conway *et al.*, 1990; Dunaif *et al.*, 1987), and anovulatory PCOS women have a seven-fold increased risk of developing NIDDM in later life (15% compared to a population risk of 2%) (Dahlgren *et al.*, 1992b). There is also evidence for dyslipidaemia in PCOS women, that is closely related to both hyperinsulinaemia and insulin insensitivity (Conway *et al.*, 1992; Graf *et al.*, 1990; Robinson *et al.*, 1996; Wild and Bartholomew, 1988). The typical lipid profile (high serum triglycerides and reduced serum HDL-cholesterol) is considered to be predictive of cardiovascular disease (CVD) risk in women, and some evidence of an increased risk of coronary artery disease in middle-aged PCOS women had been obtained by a study in Sweden (Dahlgren *et al.*, 1992a). However, in a recent long-term follow-up study of 786 women (followed for an average of 30 years, up to the age of 75), previously diagnosed with PCOS between 1930 and 1979 in the UK, no increased risk of mortality from circulatory disease could be found (Pierpoint *et al.*, 1998). It was therefore proposed that the characteristic endocrine profile of women with PCOS (specifically the unopposed oestrogens) may actually be protective against cardiovascular disease and this may balance out the deleterious effects of the insulin resistance and its associated dyslipidaemia.

Insulin resistance and PCOS

Insulin resistance in PCOS is characterised by a decreased sensitivity to insulin in peripheral tissues such as muscle and adipose tissue. However, in contrast to the insulin resistance of type 2 diabetes, there is normal hepatic insulin sensitivity (Peiris *et al.*, 1989). Furthermore, there is no evidence for ovarian insulin resistance (Willis *et al.*, 1996). Obesity (body mass index (BMI) of > 25kg/m^2) is a common feature of PCOS (Dunaif *et al.*, 1992a; Franks, 1989; Yen, 1980) with body fat generally more centrally (or on the upper body) rather than gluteally distributed compared to normal euandrogenic obese women. Although obesity, *per se*, and in particular, truncal-abdominal obesity, is associated with hyperinsulinaemia and insulin resistance (Evans *et al.*, 1984), the decrease in insulin sensitivity in PCOS is not entirely dependent on the effect of obesity (since it is also observed in a proportion of non-obese PCOS women (Chang *et al.*, 1983a; Dunaif *et al.*, 1989)). There is, however, a synergistic effect such that insulin resistance is most marked in the obese PCOS women (Peiris *et al.*, 1989; Robinson *et al.*, 1992).

The cellular mechanism of insulin insensitivity in PCOS remains unresolved. Reduced binding of insulin to its receptor has been suggested from studies of blood cells (Flier *et al.*, 1985; Jialil *et al.*, 1987). However, results from studies of peripheral adipocytes have shown normal binding but reduced insulin-mediated glucose transport, suggesting a post-receptor defect (Ciaraldi *et al.*, 1992; Dunaif *et al.*, 1992a). Further investigations by Dunaif and colleagues (Dunaif *et al.*, 1995) revealed a decrease in insulin receptor autophosphorylation in response to insulin in approximately 50% of anovulatory PCOS women. Since no mutations could be identified in the coding sequence of the insulin receptor gene in these women, it was proposed that the defect in insulin action may be due to a factor extrinsic to the insulin receptor that is involved in the regulation of receptor kinase activity, such as a serine/threonine kinase or a serine/threonine phosphatase inhibitor.

β-cell dysfunction in PCOS

There is also evidence for an insulin secretory defect in PCOS that is independent of the degree of insulin resistance, obesity and glucose intolerance. Holte and co-workers (Holte *et al.*, 1994a) have shown an increased early insulin response to glucose in both obese and non-obese PCOS women, which is not accounted for by the degree of insulin resistance. In a second study, they demonstrated that weight loss in severely insulin resistant obese PCOS women resulted in an improvement of insulin sensitivity, but the increase in early insulin secretion persistented (Holte *et al.*, 1995). These studies most likely indicate an increase in pancreatic β-cell insulin production as a fundamental defect contributing to the pathogenesis of PCOS. In contrast, evidence for a decreased insulin secretory response in PCOS women has also been provided. This has been demonstrated in women with PCOS and impaired glucose tolerance (Holte *et al.*, 1994a) (which may reflect a secondary β-cell dysfunction, resulting as a consequence of prolonged insulin resistance). However, inadequate insulin release for the degree of insulin resistance has also been demonstrated in both obese and non-obese PCOS women and this was found to be independent of glucose intolerance (Dunaif and Finegood, 1996b). Further evidence for β-cell dysfunction in PCOS has been provided by Ehrmann and colleagues, who demonstrated defects in β-cell entrainment to an oscillatory glucose infusion (Ehrmann *et al.*, 1995) and decreased meal-related insulin secretory responses (O'Meara *et al.*, 1993). Furthermore in the PCOS women who have a first-degree relatives with NIDDM, these defects were more pronounced (Ehrmann *et al.*, 1995), suggesting that they may be at a particularly high risk of developing glucose intolerance.

Thus, defects in insulin secretion (β-cell function) and insulin action (possible post-receptor defects), as well as abdominal obesity, are all likely to contribute to the metabolic aspects of PCOS. This pattern of metabolic disturbance, together with the greater predisposition of PCOS women to develop type 2 diabetes in later life, supports the idea that PCOS is a subphenotype of NIDDM.

Insulin and androgens

High serum insulin levels may affect androgen secretion and metabolism in a number of ways. As mentioned earlier in this section, insulin resistance in PCOS appears to be limited to peripheral tissues. However, the liver and ovaries remain insulin sensitive (Peiris *et al.*, 1989; Willis *et al.*, 1996). Hyperinsulinaemia affects the clearance and bioavailability of testosterone by suppressing hepatic SHBG production (Sharp *et al.*, 1991; Singh *et al.*, 1990) (reviewed in **section 1.1.4**). Insulin also has a gonadotrophic effect on ovarian steroidogenesis. The human ovary (theca, stroma and granulosa cells) has been shown to possess receptors for insulin and IGFs (Poretsky, 1994). *In vitro* studies have shown that physiological concentrations of insulin are capable of stimulating both basal and gonadotrophin-stimulated steroid production in theca cells (androgens) and granulosa cells (oestradiol and progesterone) from normal and polycystic ovaries (Nahum *et al.*, 1995; Willis *et al.*, 1996).

Hyperinsulinaemia and/or insulin resistance are predominantly features of anovulatory women with PCOS. Hyperandrogenic women with polycystic ovaries and regular menstrual cycles have normal insulin sensitivity and their fasting and glucose-stimulated serum insulin concentrations are indistinguishable from those in weight matched controls (Robinson *et al.*, 1993). This observation indicates that polycystic ovaries and hyperinsulinaemia are likely to be separate abnormalities which when combined are associated with anovulation.

1.2 POLYCYSTIC OVARIAN MORPHOLOGY

The typical polycystic ovarian morphology described earlier (**section 1.1.1, figure 1.1**) results from an alteration in the normal process of folliculogenesis. The exact mechanism of altered folliculogenesis is not completely understood, but before considering the aetiology of PCO it is necessary to briefly consider the process of folliculogenesis in normal women.

1.2.1 Folliculogenesis in the normal ovary

At birth, the human ovaries contain one to two million primordial follicles, each of which has the potential to mature, secrete oestrogens, and ovulate. The full process of follicular development (initiation, growth, selection and luteinization (**figure 1.2**)) is dependent on the appropriate stimulation of the ovaries by the gonadotrophic hormones follicle-stimulating hormone (FSH) and luteinizing hormone (LH). LH and FSH are secreted by the anterior pituitary gland and co-ordinate the temporal pattern of expression of the genes required to initiate and maintain this developmental program.

Initiation and growth

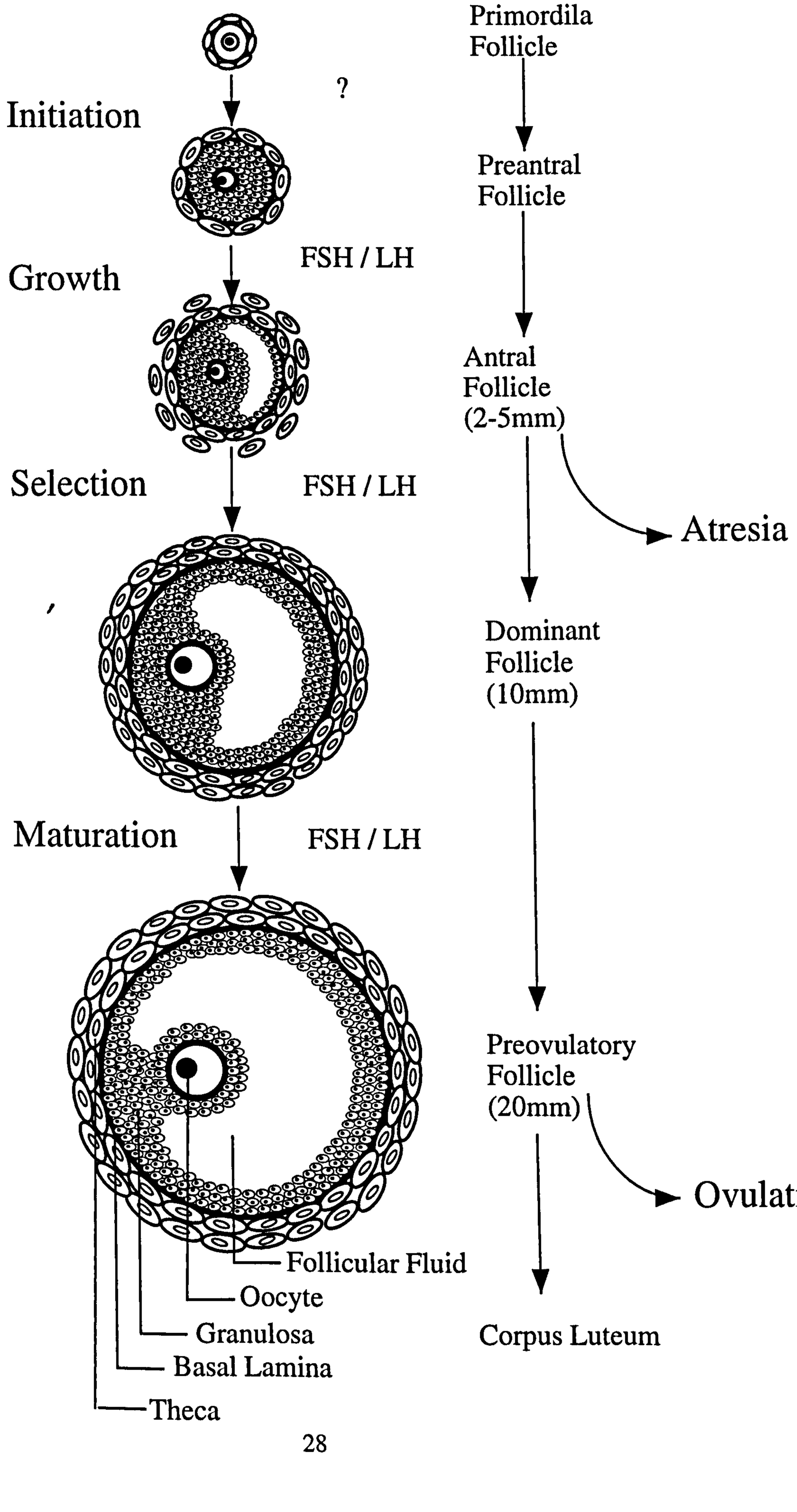
The mechanism responsible for the initiation or recruitment of primordial follicles to start growth and development is not known. However once a follicle is recruited it can either develop to ovulation or, as is the case for the vast majority, it will undergo atresia (a process of apoptosis or programmed cell death). Once recruited, follicles continue to grow and develop to the preantral stage (**figure 1.2**). It is at this point that they become gonadotrophin-responsive by acquiring LH and FSH receptors on the theca and granulosa cells respectively (Richards and Midgley, 1976; Shima *et al.*, 1987). The process from initial recruitment of a primordial follicle to development to the preantral stage is thought to take several months, and growth from this stage to the final stage of ovulation then takes about 85 days (Gougeon, 1986).

Selection and maturation

At the beginning of each new menstrual cycle, the ovaries contain about twenty intermediately mature (antral) follicles, which are available for recruitment to preovulatory development (**figure 1.2**). These follicles are present at varying maturational states and all continue to grow and develop to a lesser or greater extent in response to the increase in blood FSH levels. Usually only one (the most

Text cut off in original

Figure 1.2 Folliculogenesis



responsive) is 'selected' for ovulation (Gougeon, 1986; Hillier *et al.*, 1994; McNatty *et al.*, 1983).

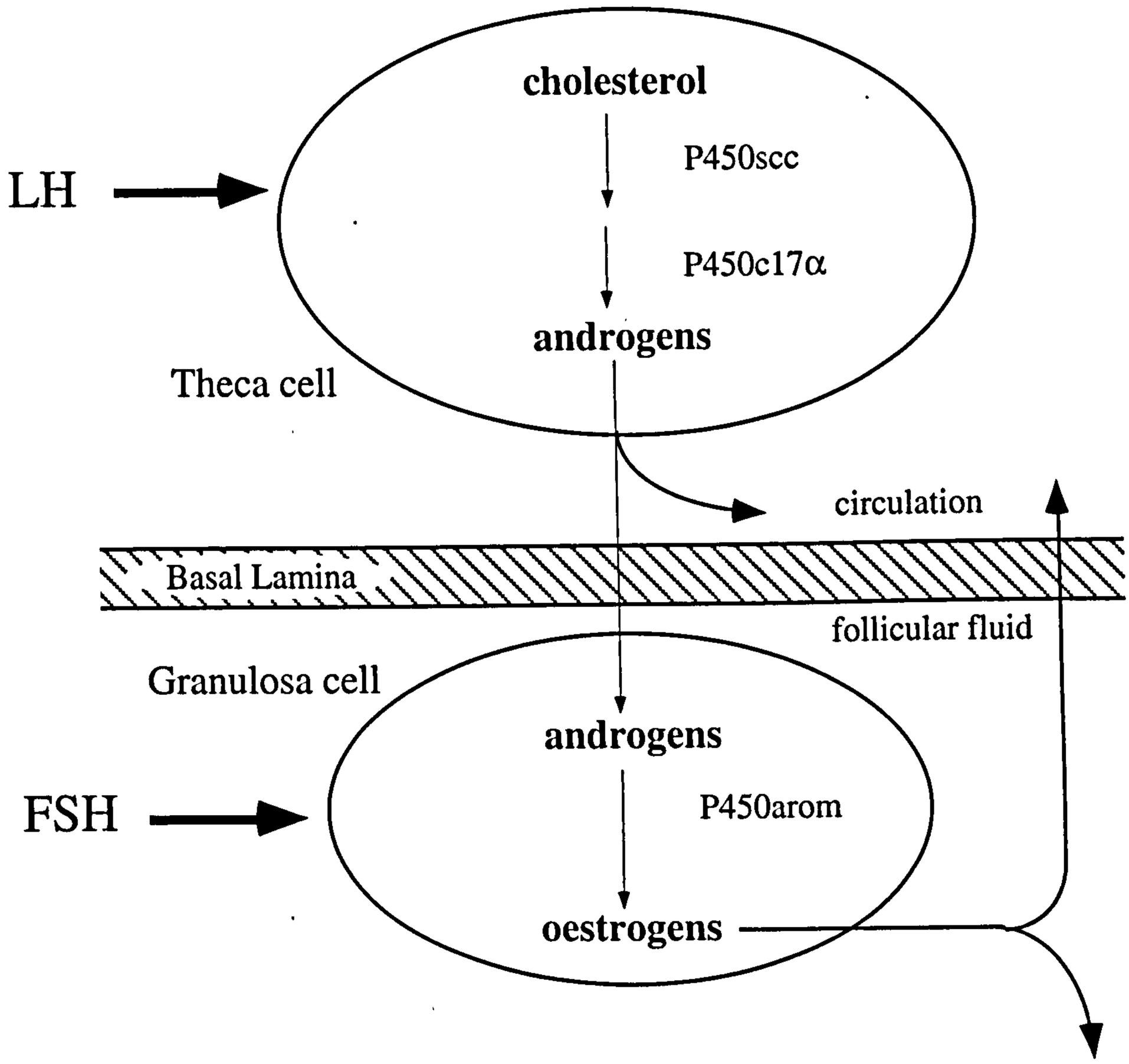
Oestradiol synthesis in maturing follicles is crucially dependent on the activity of the enzyme cytochrome P450 aromatase. Within the ovary, aromatase activity is exclusive to the granulosa cells (see later **figure 1.3**), where it is regulated by FSH (Whitelaw *et al.*, 1992). The relatively high responsiveness of the selected dominant follicle to FSH results in a rapid build up of high levels of aromatase activity in the granulosa cells. Levels of aromatase activity continue to increase, such that by late follicular phase, when ovarian oestrogen secretion peaks, aromatase activity in the preovulatory follicle is usually hundreds of times higher than in any other follicle. At this time the pre-ovulatory follicle is, briefly, the source of greater than 90% of the oestrogen being produced by the body (Hillier *et al.*, 1994).

Once the final stages of development of the dominant follicle have begun, oestrogen-mediated negative feedback on the anterior pituitory gland results in a decline in circulating blood FSH levels to the mid-follicular phase. The preovulatory follicle, which is both more sensitive and more responsive to FSH, is unaffected by this decline and continues to grow and secrete oestrogen. The other follicles in the cohort, being more dependent on FSH and therefore more vulnerable, cease to develop and undergo atresia.

The ability of the preovulatory follicle to synthesise and secrete oestrogen also depends on its ability to produce androgens (which are the precursors of oestrogens). Androgen synthesis in the ovary takes place in the theca cells where it is regulated by LH. Binding of LH to its receptor induces the expression of a number of genes involved in androgen production, including the genes encoding P450scc (side chain cleavage) and P450c17α (17 α-hydroxylase, 17-20 lyase) (see later **figures 1.3** and **1.4**). As circulating LH levels rise, increasing amounts of androgens are produced by the theca. This is then secreted into the venous effluent and also diffuse through the basal lamina into the granulosa cell layer, where they are available for aromatisation

to oestrogen, under the influence of FSH. This phenomenon is termed the 'two gonadotrophin-two cell type' mechanism of oestrogen biosynthesis (**figure 1.3**). Granulosa cells also express androgen receptors. Thus androgens also perform a direct regulatory function by enhancing follicular responsiveness to FSH (Hillier, 1991). Additionally, FSH induces granulosa cell LH receptors (Erickson *et al.*, 1979; Richards, 1994), further increasing the responsiveness of the pre-ovulatory follicle and reducing its dependence on FSH.

Figure 1.3 Two gonadotrophin-two cell type model of oestrogen synthesis.



Ovulation

About one week after the start of the menstrual cycle, the dominant follicle is the healthiest and largest follicle, having reached a diameter of about 10mm . During the second week, the development-related increase in responsiveness of this dominant follicle to FSH and LH allow it to continue to grow and secrete more and more

oestradiol. Once fully mature, having reached a diameter of >20mm, it ovulates in response to the mid-cycle LH surge .

Therefore in the ovary, androgens serve both as oestrogen precursors as well as performing a direct regulatory function by modulating granulosa cell responsiveness to FSH (Hillier *et al.*, 1994; Hillier, 1991). Hence, locally produced androgens play an important role in the normal functioning of the female gonad.

1.2.2 Altered folliculogenesis in the polycystic ovary

In the polycystic ovary, there are about twice the number of growing follicles relative to normal ovaries (refer to **figure 1.2**). There is, however, no evidence for a reduction in the number of primordial follicles (Hughesdon, 1982). This indicates that the initial steps of folliculogenesis, that is, recruitment and growth to antral stages, are normal in the polycystic ovary. The abnormality in folliculogenesis most likely occurs at a stage beyond these initial steps. In addition, in anovulatory women with PCO, there is also a lack of selection and/or progression of a dominant follicle. The overall effect is that there is an accumulation of a large number of small follicles in which the theca cells produce abnormally high amounts of androgen, resulting in a state of continued hyperandrogenism. The exact mechanism of altered folliculogenesis in the polycystic ovary is not known. There is, however, evidence to suggest that the typical ovarian morphology and the development of anovulation have separate causes (Robinson *et al.*, 1993). Alterations in the endocrine, and paracrine/autocrine control of folliculogenesis have been proposed for the former, and a role for insulin in the mechanism of anovulation has recently been postulated (Willis *et al.*, 1996).

Willis and co-workers have demonstrated that granulosa cells from normal and polycystic ovaries show increased subsequent basal and LH-induced oestradiol and progesterone production *in vitro,* if they are pre-incubated with insulin (Willis *et al.*,

1996). They have therefore proposed that in anovulatory women with PCOS, elevated insulin levels leads to an enhancement of LH action on granulosa cells of the maturing follicle. This results in an effective activity of LH that is equal to what is normally seen only at the onset of the LH surge. This effect not only enhances steroidogenesis, but also brings about an inappropriate advancement of granulosa cell differentiation and, hence, leads to premature arrest of follicular growth.

1.3 CURRENT TREATMENTS

1.3.1 Anovulation

In women with polycystic ovary syndrome who have anovulatory infertility, the treatment of first choice for ovulation induction is the antioestrogen clomiphene citrate. Of subjects treated in this way, 75-80% will ovulate and there is a high resultant pregnancy rate (Hull, 1992; Franks *et al.*, 1985). For patients resistant to clomiphene citrate, exogenous gonadotrophin preparations, principally urine-derived, have been used for more than 30 years with variable success. The use of conventional doses generally results in an overall low pregnancy rate of less than 30%, and is associated with significant problems such as a high incidence of multiple pregnancies (25-30%) and ovarian hyperstimulation syndrome (Lunenfeld and Insler, 1978; Wang and Gemzell, 1980). This has lead to the development of a chronic low-dose gonadotropin regimen that allows the induction of a single follicle ovulation and thus reduces the rate of multiple pregnancies (< 7%) and hyperstimulation syndrome while achieving conception rates of up to 45% (Brown, 1978; Polson *et al.*, 1987b; White *et al.*, 1996). Pulsatile administration of gonadotrophin-releasing hormone (GnRH) is another (less successful) method that may be used to induce unifollicular ovulation in women with PCOS (Jacobs, 1994).

The one factor known to have a negative influence on the outcome of treatment with low-dose gonadotrophin or GnRH (in terms of both the rate of ovulation and

miscarriage) is obesity (Jacobs, 1994; White *et al.*, 1996). The adverse effect of obesity however may be reversible with weight reduction, and this in some cases may itself be sufficient to result in spontaneous ovulation and pregnancy (Kiddy *et al.*, 1992). Therefore, calorie restriction is an important part of infertility management of obese women with PCOS.

Surgical treatment is another option that may be offered to women who do not respond to clomiphene citrate (reviewed by Donesky and Adashi, 1995). The traditional treatment of ovarian wedge resection (resulting in a temporary re-commencement of regular menstrual cycles by an unknown mechanism), has been replaced with the less invasive laparoscopic ovarian diathermy (wherein a laser "drilling" technique is used to burn small areas on the surface of the ovary). The efficacy of surgical treatment is comparable to treatment with exogenous gonadotrophins (Abdel-Gadir *et al.*, 1990), without the risks of multiple gestations or hyperstimulation although there is a risk of postoperative pelvic adhesion (Donesky and Adashi, 1995).

Management of dysfunctional uterine bleeding in women who do not desire pregnancy is usually achieved by either a low-dose oral contraceptive or by the administration of cyclical progestins (Franks *et al.*, 1985).

1.3.2 Cutaneous manifestations of hyperandrogenism

Localised, moderate hirsutism may be treated cosmetically by simple hair removal techniques. In more severe cases however, antiandrogen therapy may be offered. The most widely used antiandrogen in Europe is cyproterone acetate (Miller and Jacobs, 1986). This is also a progestin and is administered with ethinyl oestradiol to provide effective control of menses and contraception. Other antiandrogens used in the treatment of hirsutism include flutamide and spironolactone (Cusan *et al.*, 1994). Side effects of antiandrogen therapy include mood changes, loss of libido and lethargy and there is a rare risk of liver dysfunction. Recently, the 5 α- reductase

inhibitor finasteride, has been evaluated for its use in the treatment of hirsutism (and other androgen-dependent skin disorders). 5 α-reductase converts testosterone to the more potent androgen dihydrotestosterone (DHT), which then interacts with the androgen receptor. Of the two 5 α-reductase isozymes, the type 1 enzyme predominates in the pilosebacious unit of skin. Although finasteride (which has been used in the treatment of benign prostate hyperplasia) is preferentially a type 2 isozyme inhibitor, preliminary results from several studies have shown it to be effective in improving hirsutism scores (Castello *et al.*, 1996; Fasetti *et al.*, 1997; Gormley, 1995) (presumably the selectivity is partially overcome at therapeutic doses). Overall, finasteride appears to be effective and has minimal side effects. This should encourage the pharmaceutical industry to develop 5 α-reductase type 1 inhibitors, which are likely to be more effective than finasteride in the treatment of androgen-dependent skin disorders. Any medical therapy for hirsutism is likely to involve long-term treatment, with maximal effect usually only reached after a significant period of time (6 to 18 months). Also, the effects of therapy are not permanent and symptoms will usually return after treatment has ceased (e.g. if pregnancy is desired). It is important that while on any form of antiandrogen treatment, pregnancy should be avoided for at least 4 months after discontinuation of therapy, to avert possible interference of the drug with normal masculinization of the male fetus. Acne may be treated by the use of antibiotics or antiandrogens. Antiandrogens are also offered in the treatment of androgen-dependent alopecia, however satisfactory cosmetic results are often not achievable and complete reversal of symptoms are rare.

1.3.3 Improvement of insulin sensitivity

Recently, insulin sensitising agents such as metformin (Nestler and Jakubowicz, 1996) and troglitazone (a thiazolidinedione) (Dunaif *et al.*, 1996a) have been evaluated for their use in the treatment of both the reproductive and metabolic aspects of PCOS. Although some promising results have been demonstrated so far, long-term

results are not yet available. Metformin acts by enhancing the sensitivity of peripheral tissue to insulin and inhibiting hepatic glucose production. Metformin therapy has been shown to reduce plasma insulin levels and consequently lead to a substantial amelioration of the hyperandrogenism in anovulatory PCOS women as well as showing improvement in reproductive function (Diamanti-kandarakis *et al.*, 1998; Nestler and Jakubowicz, 1996; Velazquez *et al.*, 1994). There is, however, some controversy as to whether the beneficial effects observed are directly a result of metformin itself or may, in part, be related to the associated weight loss. The main criticism of these studies is the lack of appropriate control groups. Some studies have shown effectiveness (Diamanti-kandarakis *et al.*, 1998), whilst others have shown no effect (Ehrmann *et al.*, 1997a), in obese PCOS subjects whose weight was maintained throughout the study.

Thiazolidinediones are oral antihyperglycemic agents that improve insulin sensitivity (in muscle and adipose tissue by inhibition of hepatic gluconeogenesis) without significant weight changes. Improved insulin sensitivity, decreased insulin-mediated ovarian androgen excess and improved menstrual cyclicity have been noted in PCOS women treated with troglitazone (Dunaif *et al.*, 1996a; Ehrmann *et al.*, 1997b). However, some reports of hepatic dysfunction have been made with troglitazone and so this drug is no longer used in the UK (Day, 1999). In the case of both metformin and thiazolidinediones, although there is no documentation of teratogenicity, it is sensible to recommend that these drugs should be discontinued when pregnancy is diagnosed.

Therefore, although at present many of clinical and reproductive aspects of PCOS may be treated fairly effectively, there are still patients who do not respond to therapy. As a clearer understanding of the molecular basis of PCOS is gained, it is hoped that improved therapy may be designed that is more accurately tailored to the need of individual patients based on the particular aetiology of their symptoms.

1.4 ISSUES IN PATHOGENESIS

Despite the heterogeneity of clinical presentations, there are certain biochemical features that are common to all groups of women with a polycystic ovarian morphology, whether symptomatic, or not. As mentioned earlier (**section 1.1.4**) the most common endocrine markers include an elevation of serum androgens and to a lesser extent, hypersecretion of LH. This suggests that the distinctive pattern of increased follicle number and stromal hypertophy (Hughesdon, 1982) of the polycystic ovary may either be due to an intrinsic ovarian abnormality which results in excessive androgen production, or an extrinsic effect secondary to the abnormal hormonal/steroidal milieu to which it is exposed.

Primary disorders of the hypothalamus, adrenals and ovaries, have all been implicated in the pathogenesis of PCO/PCOS. There is also evidence of a major genetic component to the aetiology of this disorder (see later **section 1.5**).

1.4.1 Is the abnormality central or peripheral ?

Primary hypothalamic defect

Gonadotrophin-releasing hormone (GnRH), secreted by the hypothalamus, stimulates the pulsatile release of gonadotrophins (LH and FSH) from the anterior pituitary. A primary abnormality of hypothalamic regulation of LH in PCOS has been proposed based on the result of studies which have shown an increase in LH pulse amplitude (Franks, 1989; Rebar *et al.*, 1976) and frequency (Waldstreicher *et al.*, 1988). However, raised serum LH is unlikely to be the primary defect for several reasons. Firstly, elevated LH is only observed in some, but not all, women with polycystic ovaries (many have normal basal and pulsatile LH secretion) (Adams *et al.*, 1986; Conway *et al.*, 1989; Yen, 1980). Secondly, there is evidence that unopposed oestrogens (in the anovulatory PCOS women) as well as elevated serum androgens may have a role in sensitising the pituitary to gonadotrophin-releasing hormone (GnRH). This is supported by the following observations:

- Induction of ovulation in women with PCOS results in normalisation of LH secretion and the LH response to GnRH (Blankstein *et al.*, 1987).
- Other hyperandrogenaemic disorders may also be associated with hypersecretion of LH, e.g. congenital adrenal hyperplasia (CAH) and androgen-secreting tumours, (Barnes *et al.*, 1994).
- Furthermore, tumour removal has been shown to normalise both the androgen and LH levels as well as the sensitivity of LH to GnRH in a woman with an ovarian androgen-secreting tumour (Dunaif *et al.*, 1984).

Therefore it is unlikely that raised LH is the primary defect in women with PCO. It probably occurs secondary to abnormal steroid feedback on the hypothalamic-pituitary axis.

Primary role for androgens

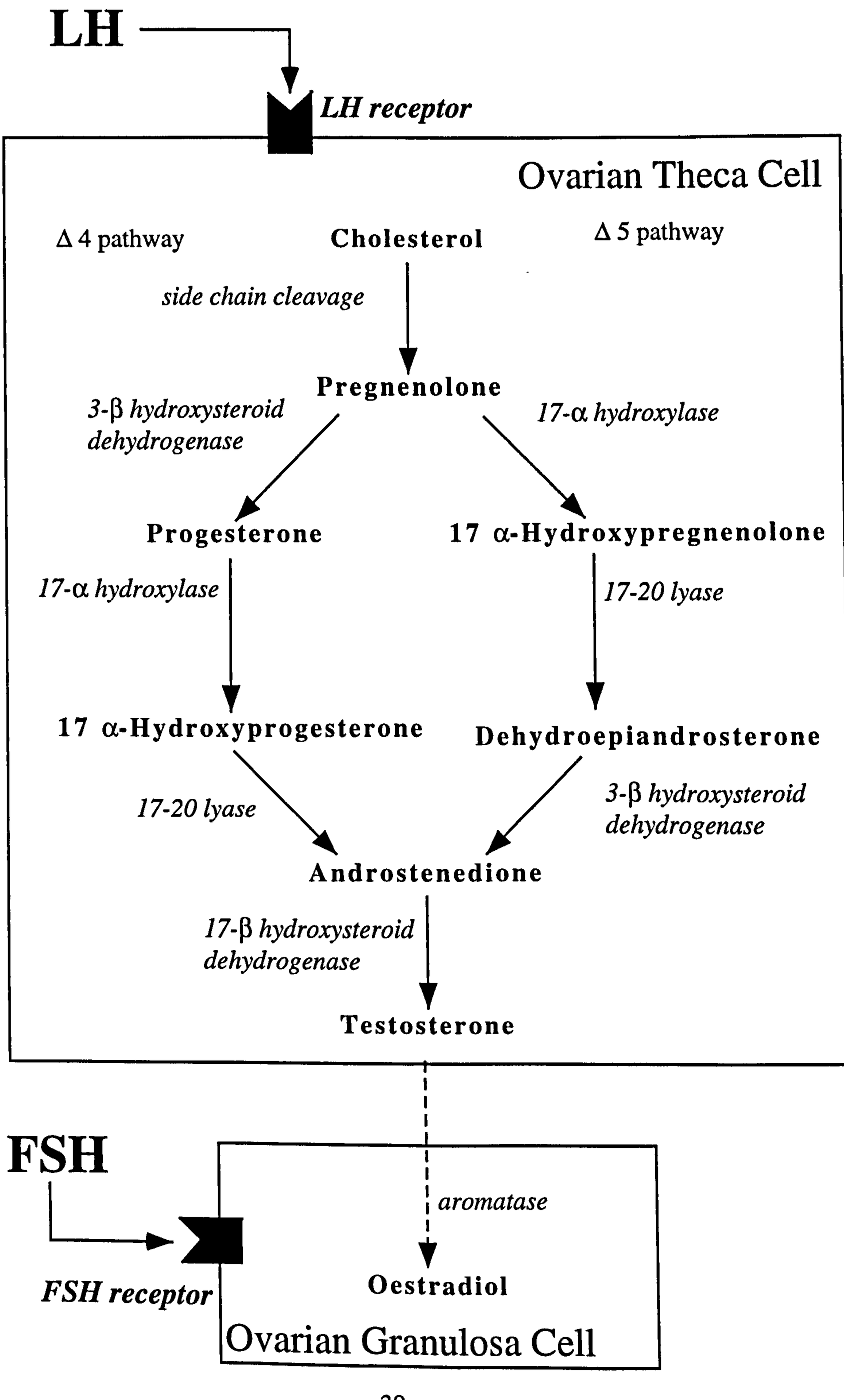
Androgens are secreted by normal women from the adrenals and the theca cell layer of the developing ovarian follicle (**section 1.2.1**). In normal women, 50% of androgens are secreted from the adrenal glands and 50% are derived from the ovaries. Hyperandrogenaemia is the most common feature of all women with polycystic ovaries (see **section 1.1.4**), regardless of clinical presentation. Circumstantial evidence that raised androgens may cause the polycystic ovarian morphology comes from the association of polycystic ovaries with other hyperandrogenaemic states (the so called "polycystic ovary-like syndromes" (refer to **section 1.1.3**)). Also a high prevalence of polycystic ovaries have been found in female to male transsexuals who have undergone exogenous androgen therapy (Futterweit and Deligdisch, 1986; Pache *et al.*, 1991; Spinder *et al.*, 1989). Interestingly, in this latter group the high-dose androgen treatment causes suppression of gonadotrophins, yet their ovaries are not suppressed, and are enlarged with increased numbers of follicles and theca-interstitial hyperplasia. Another study compared androgen receptor (AR) expression in ovaries of long-term androgen-treated female to male transsexuals, with ovaries derived from women with PCOS and those from normal controls. Evidence for androgen induced

up-regulation of AR expression was found in the PCOS and female transsexual patients (Chadha *et al.*, 1994). More recently a possible direct role has been demonstrated in the primate ovary (Vendola *et al.*, 1998). Normal cycling rhesus monkeys were treated with exogenous androgens for 3 to 10 days. Subsequent histological examination of their ovaries revealed polycystic ovary-like morphological changes.

1.4.2 Ovarian or adrenal source of hyperandrogenaemia ?

There has been much debate as to whether the adrenal or the ovary is the principal source of the excess androgen production in women with polycystic ovaries. Although hypersecretion of adrenal androgens is thought to contribute to the hyperandrogenism of some women with PCOS (since increased secretion of adrenal androgens and hyper-responsiveness of adrenal androgens to adrenocorticotrophic hormone (ACTH) has been seen in some women with PCOS (Rodin *et al.*, 1994)), the weight of evidence favours the ovary as the principal source (Franks, 1995; Gilling-Smith *et al.*, 1997; Nelson *et al.*, 1999). Clinical investigations have shown that pituitary-ovarian suppression by long-acting analogues of GnRH in women with PCOS results in a decline of serum androgens to the levels of women who have undergone oophorectomy (Chang *et al.*, 1983b; Couzinet *et al.*, 1986). Also catheterisation of ovarian and adrenal veins in women with idiopathic hirsutism and hirsute women with PCOS demonstrated an ovarian source of excess androgen (Kirschner *et al.*, 1976; Wajchenberg *et al.*, 1986). Furthermore, *in vitro* studies of isolated human theca cells (in both short and long-term cultures) have implicated a primary ovarian abnormality, by demonstrating significantly greater androgen production by cells originating from polycystic ovaries compared to those derived from normal ovaries (Gilling-Smith *et al.*, 1994; Gilling-Smith *et al.*, 1997; Nelson *et al.*, 1999). **Figure 1.4** illustrates the androgen biosynthesis / metabolism pathway in the ovary.

Figure 1. 4 Androgen biosynthesis/metabolism in the ovary.



1.5 GENETICS OF PCOS

Over the last 20 years, a small number of family studies of PCOS have been carried out, that have drawn attention to the presence of familial clustering of cases (Cooper *et al.*, 1968; Ferriman and Purdie, 1979; Givens, 1988; Hague *et al.*, 1988; Lunde *et al.*, 1989; Carey *et al.*, 1993; Norman *et al.*, 1996). Although this suggests a significant genetic component to the aetiology of the disorder, there is no agreement on the precise mode of inheritance. The main reasons why such studies have been difficult to perform have been highlighted by recent review articles (Simpson, 1992; Legro, 1995; Franks *et al.*, 1997; Legro *et al.*, 1998) and are as follows:

- A lack of consensus regarding the clinical phenotypes (**section 1.1.1**) has led to the use of various diagnostic criteria in the identification of probands and affected family members.

- Different methods have been used in the identification of affected family members. Some studies have used direct clinical observation (Givens, 1988; Hague *et al.*, 1988; Carey *et al.*, 1993; Norman *et al.*, 1996), some have done this by questionnaire alone (Ferriman and Purdie, 1979; Lunde *et al.*, 1989) and others by a combination of the two (Cooper *et al.*, 1968).

- Little attention has been paid to the characterisation of the male phenotype. Although most studies have suggested premature male pattern baldness (Ferriman and Purdie, 1979; Lunde *et al.*, 1989; Carey *et al.*, 1993; Norman *et al.*, 1996), other proposals have been increased "pilosity" (Cooper *et al.*, 1968; Lunde *et al.*, 1989) and oligozoospermia (Givens, 1988) and in one study, no attempt was made to characterise the men (Hague *et al.*, 1988).

- Difficulty in assigning affection status over more than one generation (since PCOS primarily affects women of reproductive age), make segregation analyses hard to perform.

1.5.1 Family studies of PCOS

Most studies have suggested an autosomal dominant mode of inheritance (Cooper *et al.*, 1968; Ferriman and Purdie, 1979; Lunde *et al.*, 1989; Carey *et al.*, 1993). Cooper *et al.* found that 9 of 19 sisters of probands (who had been identified by the presence of PCO on ovarian wedge resection) had oligomenorrhea. They therefore proposed an autosomal dominant mode of inheritance with reduced penetrance (Cooper *et al.*, 1968). Ferriman and Purdie identified 284 probands with hirsutism and enlarged ovaries and found a significantly higher prevalence of hirsutism, oligomenorrhea and infertility among first-degree relatives compared to controls. They suggested a modified dominant mode of inheritance, on the basis of a similar prevalence of symptoms between mothers and sisters (Ferriman and Purdie, 1979). They also found a high incidence of premature male-pattern baldness (MPB), defined as significant fronto-parietal hair loss before the age of 40, in first degree male relatives. Lunde and coworkers identified 132 probands on the basis of the presence of "multicystic ovaries" on ovarian wedge resection and two or more symptoms of oligomenorrhea, hirsutism, infertility and / or obesity. Information on family members was obtained by questionnaire. Overall, 58% of sisters were found to have PCOS-related symptoms. In addition, 19.7% of male first-degree relatives were reported to have early baldness or "excessive hairiness" compared to 6.5% of controls (Lunde*et al.*, 1989).

The study carried out at St Mary's by Carey and colleagues also provided evidence for an autosomal dominant mode of inheritance based on segregation ratio calculations (Carey *et al.*, 1993). This work focused on 10 well-characterised, multiple-affected families with polycystic ovaries. Women of reproductive age were assigned as affected on the basis of ultrasound evidence of polycystic ovaries and postmenopausal women were considered affected if they had a clear past history of symptoms (anovulatory infertility or menstrual irregularities and / or hirsutism). The prevalence of premature MPB in male family members was found to be 31%, which was significantly greater than controls (7%). Furthermore, by taking MPB as the

male phenotype, both male-to-female and female-to-male transmissions could be observed. For the purposes of this study the upper age limit for the onset of hair loss was taken as 30 years (rather than the conventional 40 years (Ferriman and Purdie, 1979; Lunde *et al.*, 1989)), to reduce the chances of false positives. This study differed from other earlier studies in that family members were primarily screened by direct interviews rather than by indirect evidence from questionnaires. Also, although diagnosis in the women was based on the polycystic ovarian morphology alone, greater than 90% of the women designated as affected on the basis of PCO morphology had at least one associated clinical or biochemical feature of PCOS. The segregation ratio was calculated to be 51%, which is consistent with an autosomal dominant model with high penetrance, suggesting a single gene effect. The more recent study carried out by Norman *et al.* suggest an agreement with an autosomal dominant mode of transmission. However, since only five pedigrees were studied a segregation ratio was not calculated (Norman *et al.*, 1996).

Other studies however, were not in agreement with an autosomal dominant mode of inheritance. One study proposed an X-linked dominant mode owing to an apparently stronger paternal transmission (Givens, 1988). A number of large multigenerational pedigrees (the largest of which had >150 members) were studied in great detail. These were ascertained from probands who had enlarged ovaries, hirsutism and oligomenorrhea. This was the first study to identify some of the metabolic abnormalities (such as diabetes mellitus, insulin resistance and dyslipidaemia) and the considerable phenotypic variability of PCOS within the same pedigree. In addition, oligozoospermia and elevated LH levels were identified in some of the male family members. Female members were assigned as affected on the basis of hirsutism and oligomenorrhea and were found to be affected through both maternal and paternal transmissions. However, the paternal transmission appeared stronger (87% compared to 47% maternal) and they suggested that this may reflect an X-linked dominant mode of inheritance (Givens, 1988). Finally, the study carried out by Hague and coworkers found that segregation ratios exceeded an autosomal dominant pattern

(Hague *et al.*, 1988). They identified 61 pedigrees from probands who presented to endocrine clinics with symptoms of menstrual disturbances and / or hyperandrogenaemia and had PCO on ultrasound scan. 45 of 52 sisters were found to have polycystic ovaries. The segregation ratio of 87% was considered to be too high for an autosomal Mendelian inheritance (Hague *et al.*, 1988).

1.5.2 Current hypothesis

Although segregation ratio calculations in most studies have suggested a dominant (rather than recessive) model, these types of studies cannot rule out other more complex modes of inheritance (e.g. genetic heterogeneity or poly/oligogenicity). In fact, there is a growing consensus that PCOS is more likely to be a complex trait (Simpson, 1992; Franks *et al.*, 1997; Legro *et al.*, 1998) that may result from the interaction of multiple genetic and environmental factors. This would explain the considerable symptomatic heterogeneity that is observed both in the general population as well as among sisters with polycystic ovaries. Our current hypothesis is that the apparent central disorder in women with polycystic ovaries is determined, primarily, by a small number of genes which are involved in follicular function and androgen synthesis/metabolism and that the heterogeneity of clinical and biochemical features can be explained by the interaction of these genes with environmental (notably nutritional) and other genetic factors. These would include determinants that modify the action and secretion of insulin.

At the start of this PhD project very few molecular genetic studies of PCOS had been performed. During the course of this project the results of studies carried out by other members of the group at St Mary's were published. These include the initial investigation of the gene *CYP17* which was carried out by Carey and coworkers (Carey *et al.*, 1994) (discussed later in **section 1.7.3**), and the evaluation of the insulin gene variable number tandem repeats (*VNTR*) conducted by Waterworth and colleagues (Waterworth *et al.*, 1997a) (reviewed below in **section 1.6**).

1.6 Insulin gene *VNTR* and PCOS

As mentioned earlier in **section 1.1.5**, there are characteristic abnormalities of insulin secretion and action in women with PCOS, particularly in those with anovulatory cycles. This raises the possibility that genes implicated in the action and secretion of insulin may have a role in the aetiology of PCOS and the development of anovulation. Since evidence for a disorder in pancreatic β-cell function has been demonstrated (Holte *et al.*, 1995), the insulin gene was investigated for a role in the aetiology of PCOS (Waterworth *et al.*, 1997a). Waterworth and coworkers evaluated the *VNTR* minisatellite located within the 5' regulatory region of the insulin gene on chromosome 11p15.5. This locus is thought to regulate the expression of the insulin gene (Bennett *et al.*, 1995; Kennedy *et al.*, 1995; Vafiadis *et al.*, 1996), and has been associated with susceptibility to type 1 diabetes (Bennett *et al.*, 1996), type 2 diabetes (Huxtable *et al.*, 2000) and hyperinsulinaemia related to central obesity (Weaver *et al.*, 1992). The *INS VNTR* has a bimodal distribution of alleles (amongst Europids), with class I alleles being short (average of 40 repeats) and class III being long (average of 157 repeats).

Initially linkage analysis was carried out using the GENEHUNTER program (Kruglyak *et al.*, 1996) (see **section 1.8.1**) and five polymorphic markers (including the *INS VNTR*) that spanned the insulin locus, in 17 of the PCOS/MPB pedigrees. Non-parametric linkage results provided evidence for excess allele sharing at the *INS VNTR* locus (NPL= 3.25, p = 0.002) and parametric analysis estimated that about 60% of the families may be linked to this locus (Waterworth *et al.*, 1997a). Waterworth *et al.* also looked for association between the *VNTR* and PCOS in two additional populations of Europid PCOS women. Odds ratios were calculated for *INS VNTR* genotypes both by using a case control approach (59 women with PCOS and 54 normal controls) and by using the affected family-based control (AFBAC) method (52 trios consisting of PCOS probands plus parents) (Thomson, 1995). The results of these two independent studies demonstrated association of the *INS VNTR* class III/III genotype with anovulatory PCOS, with odds ratios of 8.20 (p = 0.005) and 5.70 (p =

0.04) obtained for the case control and AFBAC studies respectively. Transmission disequilibrium test (TDT) analysis, another family based association method, was also carried out on the latter data set (Bennett *et al.*, 1997). TDT analysis revealed evidence for a parent-of-origin-effect; class III alleles were found to be transmitted significantly more often from fathers than from mothers ($p = 0.003$), suggesting a role for genomic imprinting.

The association of this gene with anovulatory PCOS corroborates well with the clinical observation that hyperinsulinaemia is predominantly a feature of women with PCO and anovulatory cycles (compared to equally hyperandrogenic women with PCO and regular menses). Thus, by using three different populations, evidence for linkage and association between the *INS VNTR* and PCOS was obtained. It was therefore concluded that the *INS VNTR* locus is a major susceptibility locus for PCOS (particularly anovulatory PCOS), which may contribute to the mechanism of hyperinsulinaemia and the high risk of NIDDM in PCOS. Further support for the existence of common pathogenic mechanisms underlying PCOS and type 2 diabetes comes from a recent study, in which a similar parent-of-origin effect (an excess of paternally-transmitted class III VNTR alleles to diabetic offspring) has been demonstrated (Huxtable, 2000).

1.7 CANDIDATE GENES FOR PCO/PCOS

Given the abundant clinical information on the biochemical phenotypes of women with polycystic ovaries, it has been possible to undertake a candidate gene approach using both linkage and association studies to investigate putative susceptibility loci (see later **sections 1.8.1** and **1.8.2**). An underlying role for androgens in the pathogenesis of PCOS has been proposed (**section 1.4.1**). Therefore, candidate genes that are key components of androgen biosynthesis and metabolism (refer to **figure 1.4**) have been selected for further investigation.

1.7.1 LH receptor

Androgen biosynthesis in ovarian theca cells is dependent on the appropriate stimulation by luteinizing hormone (LH), which exerts its gonadotrophic effect through the LH receptor (see **section 1.2.1, figure 1.4**). Hence, the LH receptor plays a crucial role in gonadal steroidogenesis. Hypersecretion of LH is observed in some women with PCOS, but this is unlikely to be the primary cause of ovarian dysfunction in PCOS (see **section 1.4.1**).

Several constitutively activating point mutations have been identified in the LH receptor gene (*LHR*) that lead to premature androgen secretion in familial male-limited precocious puberty (Shenker *et al.*, 1993; Laue *et al.*, 1995; Rosenthal *et al.*, 1996; Yano *et al.*, 1997). Most studies of familial precocious puberty in boys resulting from mutations in the *LHR* gene have focused their investigations only on the affected male members and although some have identified female carriers of the mutations, few have evaluated their effects on the pituitary-gonadal axis. Since studies of isolated human theca cells indicate that the putative defect(s) in steroidogenesis in polycystic ovaries occurs at or above the level of progesterone production (Gilling-Smith *et al.*, 1994), the *LHR* may be considered as a pertinent candidate for involvement in the aetiology of PCOS. The *LHR* gene has been cytogenetically mapped to chromosome 2p21 (Rousseau-Merck *et al.*, 1990).

1.7.2 *CYP11a*

It has been shown that theca cells from polycystic ovaries show a significant increase in both androstenedione and progesterone production *in vitro* when compared to normal theca (Gilling-Smith *et al.*, 1994). This suggests that the putative defect in steroidogenesis occurs at or above the level of progesterone biosynthesis (**figure 1.4**). The enzyme cholesterol side chain cleavage (cytochrome P450scc), which is the product of the gene *CYP11a* on chromosome 15q24, catalyzes the conversion of cholesterol to pregnenolone, the first, enzymatically rate-limiting step at the start of

the steroid hormone biosynthesis pathway (Stone and Hechter, 1955; Halkerston *et al.*, 1961). It may therefore be proposed that upregulation of this enzyme could lead to an increase in androgen production. Hence, in the work presented in this thesis, *CYP11a* is investigated as a candidate for involvement in the aetiology of PCOS.

1.7.3 CYP17

The gene *CYP17* codes for the enzyme cytochrome P450c17α which catalyses both the 17α-hydroxylase and the 17-20 lyase activities, and is a rate limiting step in androgen biosynthesis in the ovaries and the adrenals (**figure 1.4**). Based on evidence from several studies, which suggested that there may be an abnormal regulation of this enzyme in women presenting with PCOS (Barnes *et al.*, 1989; Rosenfield *et al.*, 1990; Gilling-Smith *et al.*, 1994), *CYP17* was the first gene to be investigated as a candidate for involvement in the aetiology of PCOS (Carey *et al.*, 1994). This work was carried out by other members of the PCOS group at St Mary's prior to the commencement of this thesis. Parametric linkage analysis, in twenty PCOS/MPB pedigrees (described in **section 2.1.1**), excluded the *CYP17* locus on chromosome 10q24.3 as a primary genetic defect. These data were later re-analysed by one of the authors of the original study (Waterworth, 1997b), using the non-parametric linkage approach of GENEHUNTER (see later, **section 1.8.1**). These results were in agreement with the original findings excluding evidence of excess allele sharing across the entire chromosomal region examined. Thus, it was concluded that *CYP17* was unlikely to be a major susceptibility locus involved in the development of PCOS/MPB.

Sequencing analysis of the *CYP17* promoter region, however, identified a nucleotide change, a T to C substitution at -34bp from the ATG start of translation site, that conferred an additional SP1-type promoter element (CCACC box). This base change created an *Msp*A1 restriction enzyme cleavage site allowing a simple screening assay to be set up. The frequency of this variant (A2) allele was investigated in a small,

case control data set comprising consecutively identified Europids with polycystic ovaries (n = 44) and normal controls (n = 24). Preliminary results showed an association of the A2 allele with PCOS (odds ratio =3.57, p = 0.03). It was therefore proposed that this extra promoter element may be involved in the regulation of expression of CYP17, resulting in an increased synthesis of androgens. In this thesis, this study is extended to include a larger case control data set (**section 2.1.2**).

1.7.4 *CYP19*

The steroidogenic enzyme aromatase (cytochrome P450 aromatase), catalyzes the conversion of the C_{19} steroids (androgens) to the C_{18} oestrogens (**figure 1.4**). Aromatase is encoded by the gene *CYP19* on chromosome 15q21.1. In humans, *CYP19* is expressed in the ovary, testis, placenta, adipose tissue and the brain, where tissue-specific expression is regulated by the use of tissue-specific promoters (Mahendroo *et al.*, 1991; Means *et al.*, 1991).

In the ovary, the development of a follicle to maturity is FSH dependent (see **section 1.2.1**). FSH acts on granulosa cells of the maturing follicles where it regulates the expression of a number of genes, crucial to mature granulosa cell follicular function, including *CYP19* (Richards *et al.*, 1987). Circulating levels of FSH however have been found to be within the normal midfollicular phase range in women with PCOS (Franks, 1989; Yen, 1980). Evidence for an altered regulation of aromatase in PCOS comes from the *in vitro* study of granulosa cells from polycystic and normal ovaries (Mason *et al.*, 1994). Mason and colleagues demonstrated that granulosa cells from anovulatory polycystic ovaries were hyperresponsiveness to FSH, displaying significantly greater oestradiol production, compared to those derived from both polycystic ovaries and normal ovaries.

Conversely, other evidence for a role for *CYP19* in the aetiology of hyperandrogenism and PCO comes from the description of a number of rare cases of total aromatase deficiency which are associated with hyperandrogenaemia (Shozu *et*

al., 1991; Harada *et al.*, 1992; Ito *et al.*, 1993; Conte *et al.*, 1994) and in some cases polycystic or multicystic ovaries (Conte *et al.*, 1994; Ito *et al.*, 1993).

Overall, these studies suggest that an altered regulation of this enzyme may be involved in PCOS.

1.7.5 *SRD5A1*

The enzyme 5 α-reductase converts testosterone to the more potent androgen dihydrotestosterone (DHT) (Wilson *et al.*, 1993). Male pattern hair growth is androgen dependent and may be caused by a number of different factors such as high circulating free androgens, enhanced peripheral conversion of testosterone to dihydrotestosterone (DHT) or an increased sensitivity of the hair follicles to normal androgen levels. DHT is the major androgen implicated in pathogenesis of benign prostatic hyperplasia, MPB, acne and idiopathic hirsutism.

5 α-reductase occurs in two isoforms (Russell and Wilson, 1994), type 1 and type 2, encoded by the genes *SRD5A1* and *SRD5A2*, located on chromosomes 5p15.5 and 2p23 respectively. The type 2 isozyme is located mainly in epididymis, seminal vesicles, prostate and genital skin. High level of type 1 isozyme activity have been found in the pilosebacious unit of skin as well as in testis, liver and ovary. The type 1 and type 2 isozymes can be distinguished biochemically by differences in both their pH optima (type 1 having a more basic pH optimum with a broad range of 6.0-8.5 and the type 2 having an acidic pH optimum with a narrow range of between 5.0-5.5) and their sensitivity to the 4-azasteroid inhibitor finasteride (17β-N-t-butyl-carbamoyl-4-aza-5α-androt-1-en-3-one), with the type 2 being strongly inhibited whereas the type 1 is poorly inhibited *in vitro*. Preliminary results from several studies in which finasteride has been used in the treatment of hirsutism, have shown positive results, indicating probable type 1 enzyme inhibition at therapeutic doses *in vivo* (Gormley, 1995; Castello *et al.*, 1996; Fasetti *et al.*, 1997). This further implicates 5 α-reductase in the aetiology of hirsutism. Therefore the gene *SRD5A1* may be considered as a

BIBL.

good candidate for investigation for a role in the development of hirsutism in women with PCOS.

1.8 EXPERIMENTAL STRATEGY

The aim of this thesis has been to contribute to the understanding of the molecular genetics of PCOS by identifying predisposing genes. The primary strategy used has been to carry out linkage analysis in the 23 PCOS pedigrees (refer to **section 2.1.1**) by taking a candidate gene approach. There are two reasons for assessing candidate genes as opposed to conducting a genome search. The first is a question of insufficient power - the relatively small size of the family collection (especially at the outset of this project) meant that a genome search was not a feasible option. The second reason relates to the vast amount of clinical information available that has allowed the selection of pertinent candidate genes for investigation. As discussed earlier in this chapter, three key genes involved in ovarian androgen biosynthesis/metabolism have been chosen for evaluation. The *CYP11a* and *CYP19* genes had previously been genetically mapped to the long arm of chromosome 15 and both have the added advantage of possessing an intragenic polymorphic microsatellite marker suitable for use in the linkage analysis. Precise genetic localisation also allowed the selection of highly polymorphic microsatellite markers (from published genetic maps such as the Généthon human genetic linkage map (Dib *et al.*, 1996)) that spanned the chromosomal regions flanking each gene.

The *LHR* gene, however, had only been cytogenetically mapped to chromosome 2p21, and its precise location relative to polymorphic markers had not been determined. So this was initially radiation hybrid mapped (using the Stanford G3 radiation hybrid panel) and subsequently appropriate microsatellite markers were chosen for linkage analysis.

Candidate genes were also assessed for association with PCO (and related symptoms and biochemical indices) using an additional population of unrelated cases with PCOS and matched normal controls. This separate resource provided another non-parametric approach for the investigation of candidate susceptibility loci and their role in the aetiology of PCOS-related symptoms in the population.

1.8.1 Linkage analysis in the pedigree data set

Definition of linkage

Consider the genotype of an individual at two loci A and B. For each locus one allele was inherited from the mother (m) (haplotype of the maternal gamete was $A_{(m)}B_{(m)}$) and the other allele from the father (f) (paternal gamete haplotype was $A_{(f)}B_{(f)}$). Thus, the genotype of the individual at these loci may be written as $A_{(f)}A_{(m)}$, $B_{(f)}B_{(m)}$. A gamete produced by this individual might contain two alleles from the same parent (i.e. $A_{(f)}B_{(f)}$ or $A_{(m)}B_{(m)}$), or it may contain one allele from each parental gamete (i.e. $A_{(f)}B_{(m)}$ or $A_{(m)}B_{(f)}$). In the former case, such gametes are defined as parental types or non-recombinants with respect to the two loci. In the latter situation, where the haplotype of the gamete is a new combination of alleles, different from either parental haplotye, such gametes are defined as non-parental types or recombinants. The recombination fraction (θ) between two loci is defined as the probability that a gamete is a recombinant (i.e. $A_{(f)}B_{(m)}$ or $A_{(m)}B_{(f)}$).

If two loci are on different chromosomes (or far apart on the same chromosome), they will segregate independently (i.e. obey Mendel's Law of independent assortment). Thus 50% of offspring will be recombinant and 50% will be non-recombinant for the parental haplotypes and the recombination fraction (θ) will be 0.5. However if the loci lie close together on the same chromosome, they will tend to be inherited together more than 50% of the time (i.e. the recombination fraction will be < 0.5) and are said to be linked. If the loci are very close together, then recombination will rarely separate them. The further apart they lie on a chromosome, the more likely it is

that they will be separated by a crossover. Therefore, the recombination fraction provides a measure of the genetic distance between two loci. This premise forms the fundamental basis of linkage analysis.

The unit of genetic distance is the Morgan, or more commonly the centimorgan (cM), where 1cM corresponds to 1 recombination per 100 meioses (i.e. 1% recombination fraction). The linear relationship between recombination fraction and map distance only holds for short distances. As the distance between two loci increases, then so does the chance of multiple crossovers occurring between them. All even numbers of multiple crossovers between two markers will not be detected and odd numbers will give the impression that only one crossover has taken place. The overall effect is to reduce the observed recombination fraction. Therefore mapping functions are used to convert recombination fractions to take account of double crossovers. The most widely used mapping function is the Kosambi function.

Parametric linkage analysis

The aim of classical linkage analysis is to extract all the available inheritance information from pedigrees exhibiting a particular heritable trait and to test for co-inheritance of chromosomal regions with that trait. This is done by specifically examining whether two loci (or a locus and the genetic trait) are linked, by estimating the recombination fraction (θ) between them, and testing whether this is significantly less than 0.5. Traditionally, parametric (model-dependent) methods have been employed and the most widely used approach to parametric linkage analysis is the LOD score method (Morton, 1955). The LOD (likelihood odds ratio) score, Z, is the logarithm of the odds that the loci are linked (at recombination fraction = θ) rather than unlinked (recombination fraction = 0.5):

$$Z(\theta) = \log_{10} \frac{L(\text{linkage at } \theta)}{L(\theta = 0.5)}$$

The imperfect structure of most human pedigrees (which are often small, with key individuals, such as grandparents, missing), together with the incomplete informativeness of DNA markers, means that often recombinants cannot be identified unambiguously. Therefore, human linkage analysis is dependent on computer programs to extract the linkage information and to generate the LOD scores. The most commonly used are the MLINK and LINKMAP programs of the LINKAGE package (Lathrop *et al.*, 1984b). LOD scores are calculated for a range of θ values, and the θ at which Z is greatest indicates the best estimate of the distance between the two loci. Since the overall probability of linkage in a data set of families is a product of the probabilities in each individual family, the LOD scores can be summed across the whole data set to generate the overall LOD score. Traditionally, for monogenic traits a LOD score of at least 3 is taken as evidence for linkage. This gives a 1000:1 odds in favour of linkage with a 5% chance of error. Linkage can be rejected if a LOD score of -2 or less is obtained and a Z score of between -2 and 3 is considered inconclusive.

Markers for linkage analysis

Between homologous pairs of chromosomes, differences may exist in the DNA sequence. These are often referred to as polymorphisms. Linkage analysis requires informative meioses. Therefore, highly informative polymorphic markers should be used. There are two measures of the informativeness of a polymorphic marker. The *heterozygosity* is the probability that a random individual will be heterozygous for any two alleles of the marker, and is derived using the following formula:

$$H = 1 - \sum_{i=1}^{n} p_i^2$$

The *polymorphism information content* (PIC) is defined as the probability that the genotype of a given offspring will allow unambiguous deduction of which allele was derived from which parent.

The PIC value for a marker is derived using:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where $\mathbf{p_i}$ is the frequency of the ith allele at the locus, $\mathbf{p_j}$ is the frequency of the jth allele at the locus and **n** is the total number of alleles. The heterozygosity and PIC are very comparable measures and both vary from 0 (never informative) to 1 (always informative). In general, a marker with many relatively common alleles will tend to be more informative. Traditionally, the markers of choice for linkage analysis are the microsatellites, which are simple di- / tri- / or tetranucleotide repeat polymorphisms. These are often highly informative and occur frequently throughout the genome and may even be located within or close to genes. Markers are selected from published genetic maps such as the Généthon human genetic linkage map (Dib *et al.*, 1996), where the order and location of numerous informative microsatellite markers (with an average spacing of < 1cM) is given.

Non-parametric linkage analysis

It is well established that, under the correct model of inheritance, the maximum-likelihood analysis (via LOD score) is the most powerful approach for establishing linkage (Lander and Schork, 1994). The use of parametric methods generally applies to the simple single gene disorders where segregation analysis can accurately estimate the disease parameters. However, for common complex disorders, in which the underlying genetic model is unknown (as in the case of PCOS), segregation analyses are often hard to perform and can even be misleading (Dizier *et al.*, 1993). Segregation analysis looks for an overall pattern of inheritance of a trait. However, for a complex trait that may involve a multilocus system it cannot necessarily determine the pattern of inheritance related to any one locus. Different loci may confer different inheritance modes at a particular locus (eg. some may be dominant and others recessive). Although LOD score methods have been shown to be fairly robust to inaccurate specification of certain parameters such as penetrance and gene

frequency (Clerget-Darpoux *et al.*, 1986), they are sensitive to misspecification of the dominance model and false exclusions can be obtained (Greenberg and Hodge, 1989b; Risch, 1990a; Risch, 1990b; Risch and Giuffra, 1992).

One way of dealing with this potential problem is to maximise the maximum LOD score with respect to genetic parameters (Clerget-Darpoux *et al.*, 1986; Greenberg, 1989a; Hodge and Elston, 1994). This means analysing data for linkage under several genetic models and using the highest maximum LOD score as the test statistic for linkage. The main concerns about this type of approach are as follows:

- Varying the genetic parameters or analysis model inevitably means that multiple tests are performed, which in turn increases the probability of type I error. Therefore, a correction for the multiple testing must be applied.

- The effect of the above correction, on the statistical power to detect linkage, is not known.

- The true inheritance model may not be adequately represented by any of the approximate models tested.

Therefore, for a common complex trait, where there is uncertainty regarding the mode of inheritance (as is the case with PCOS), there is much argument in favour of using a non-parametric approach that makes no assumptions about the underlying genetic model. In principle, rather than testing whether the inheritance pattern fits a specific model for a trait-causing gene, a non-parametric method tests whether the inheritance pattern (allele sharing) deviates from expectation under independent assortment.

Until relatively recently, non-parametric analysis was performed primarily by one of two methods, the affected sibpair analysis (Suarez, 1978) and the affected pedigree member (APM) method (Weeks and Lange, 1988). For both these methods, the observed allele sharing frequencies (for a particular marker) between affected relatives are compared with the expected allele sharing distribution for that relative pair.

Alleles shared identical by descent (IBD) have a common ancestral source, whereas alleles shared identical by state (IBS) simply match regardless of origin. Hence at a particular locus, siblings can share two, one or zero alleles IBD, 25%, 50% and 25% of the time respectively. The main problem with the sibpair approach is that much of the inheritance information within the pedigree structure is wasted (since only the affected sibpairs are considered), thus reducing the power of the analysis. Although the APM approach utilises more of the pedigree information it has several drawbacks. It is not a true linkage method as it only focuses on IBS allele sharing and makes no attempt to resolve whether alleles are actually IBD. APM is also very dependent on marker allele frequencies and the associated weighting functions used. This can make it prone to spurious positive results, for example if there is homozygosity for a rare allele in a single pedigree (in which parental genotypes are missing), it will interpret this as strong evidence for excess allele sharing rather than the rare chance event that it really represents. Another disadvantage of APM is that power is reduced since it involves comparisons between pairs of individuals at a time rather than sets of affected individuals. Moreover, it cannot be used to localise a particular locus relative to a map of markers since it lacks true multipoint formulation.

GENEHUNTER

In 1996 Kruglyak *et al.* released the GENEHUNTER program (Kruglyak *et al.*, 1996), which provided a unified multipoint approach to both non-parametric and parametric linkage analysis in pedigrees of moderate size. This has several advantages over previously used programs. The non-parametric linkage (NPL) analysis is a powerful multipoint IBD approach, which extracts all available inheritance information. GENEHUNTER uses the idea of the inheritance vector (Lander and Green, 1987). This specifies which of the distinct founder (individuals without parents present in the pedigree) alleles, at each locus along the linear map, are inherited by each nonfounder (individuals with parents available). In practise, genotyping data will only provide partial information about the inheritance in the

pedigree (due to missing key individuals and limitations on the map density and heterozygosity of markers). The partial inheritance information at each locus is then represented by a probability distribution over the possible inheritance vectors. In the absence of genotype data, all the possible inheritance vectors at each locus are assigned an equal probability, but as this information becomes available, the probability distribution is focused on a subset of the vectors. Full multipoint information is then extracted from the data set by considering inheritance distribution data at each locus and reconstructing the probability distribution at each point, conditional on the entire data set (all marker loci). Thus, all the loci tested are considered simultaneously to assess the extent of allele sharing IBD among affected pedigree members, and this is compared to the expected sharing under random segregation and the significance tested. The NPL score, which is a Z score (i.e. a measure of the number of standard deviations from the expected mean), and an associated p-value, are then generated at each locus. For imperfect data (small number of pedigrees with different pedigree structures), the distribution of the Z score is asymptotically normally distributed as the number of pedigrees is increased and as the information content approaches 100%. The significance level (p-value) assigned is conservative and becomes increasingly accurate as information content increases. This approach is referred to as the "perfect data approximation".

The multipoint parametric linkage results obtained by GENEHUNTER are comparable to LINKMAP, but have the added advantage of a faster running time, the capability to handle a larger number of markers, as well as the ability to generate LOD scores allowing for heterogeneity. Finally, GENEHUNTER also carries out information-content mapping (a measure of the extent of information extracted by the marker data at each point along the linear map) and maximum-likelihood reconstruction of marker haplotypes, which indicates the most likely position of crossovers and allows data error checking by identifying double crossovers.

Options for parametric analysis

Recently, Durner *et al.* have suggested that LOD score calculations based on two simple single locus (SL) models, one dominant and one recessive can be more powerful than non-parametric approaches in detecting linkage in complex diseases (Durner *et al.*, 1999). The basis of their argument is that the critical factor in LOD score analysis is the mode of inheritance at the linked locus and not that of the disease or trait *per se*. Thus for even a complex multilocus disorder, at each given locus either one or both alleles contribute to trait susceptibility, i.e. will confer either a dominant or recessive inheritance at that specific locus.

Simulation studies were carried out based on a variety of two-locus models with one gene dominant and the other recessive, under three different models (epistasis, heterogeneity and another more complex model). The performance of two non-parametric methods (affected sibpair (ASP) and NPL methods) and the two SL parametric methods (one dominant, one recessive) were then compared with a parametric analysis under the correct two-locus model and an analysis based on a single major gene model derived by the segregation analysis program POINTER (Lalouel and Yee, 1980). Their results demonstrated that the "two SL" approach was almost as powerful as the analysis under the correct model. It was also as powerful as an analysis using POINTER (when the correct locus model parameters had been derived). They further demonstrated that this approach was more sensitive than both the non-parametric methods. However, the NPL statistic was shown to consistently outperform the ASP method.

Although the NPL method had less power than the two SL approach, it was shown that, for all but one model, the NPL method had greater than 80% (often >90%) power to detect linkage at a significance level of $p < 0.0001$. Furthermore this method was more powerful than the LOD-score analysis based on an inaccurate model derived by segregation analysis (using POINTER).

The only model for which the NPL score was unable to detect linkage was one that assumed a two locus heterogeneity model, with the dominant locus (of low gene frequency) being tested and the second locus being a high gene frequency recessive locus that accounted for disease in two thirds of the pedigrees. It is noteworthy that, for this particular model, all the methods tested faired badly, with the correct model only having just over 20% power to detect linkage. This highlights the problems associated with the detection of genetic heterogeneity for loci of smaller effect in common complex diseases.

Therefore, there is evidence that both the two SL methods and the non-parametric NPL method seem to be sensitive and valid approaches to the analysis of complex disorders. Since GENEHUNTER is capable of performing both parametric and non-parametric analyses, this program was used to carry out linkage analyses in the PCOS/MPB pedigrees. For each locus to be tested, a non-parametric and two SL parametric analyses (one dominant and one recessive) were carried out.

1.8.2 Case control association studies

Population association studies offer a complementary, non-parametric approach to assessing candidate genes for a role in disease susceptibility (Nothen *et al.*, 1993; Owen and McGuffin, 1993). The main difference between linkage and association is that linkage is a relationship between loci, whereas association is a relationship between alleles.

Disease-marker associations may be found by comparing the allele (or genotype) frequencies of a candidate gene associated marker in a series of unrelated patients (e.g. women with PCOS) and ethnically matched normal controls. The significance of any difference observed may then be evaluated statistically.

Allelic association may be observed in a variety of circumstances, such as:

- If the marker allele being tested directly causes susceptibility to disease: in this case, the same marker allele is expected to be associated with disease in any population studied (unless different populations have different causes of the disease).

- If there is linkage disequilibrium (LD) (see **section 2.4.9**) between the marker allele being tested and the tightly-linked disease-causing mutation. This is the non-random association of particular alleles at two linked loci. This association may, for example, be detected if there is a founder-effect (i.e. most disease-bearing chromosomes in the population are descendent from one or a few ancestral chromosomes). In this case the particular marker allele associated with disease may be different in different populations.

- If there is LD arising due to other mechanisms such as random genetic drift, selection and admixture between two ethnically distinct populations (**see section 2.4.5**).

Spurious allelic association may be obtained if there are differences in population substructure (e.g. population stratification may occur if there are ethnic differences between the cases and non-related controls) (see **section 2.4.5** for limitations of case control studies).

2.0 SUBJECTS, MATERIALS AND METHODS

2.1 SUBJECTS

In this study, two types of data sets, together with two methodological approaches, have been used for the investigation of candidate genes in the aetiology of PCOS. A PCOS/MPB family data set used for linkage analysis (**section 2.1.1**) and a case control data set used for association studies (**section 2.1.2**).

For all the women (except the postmenopausal and pre-menarchal (see **section 2.1.1** below)), ovarian morphology was defined by pelvic ultrasonography using an Acuson 128 ultrasound machine. A polycystic ovary (PCO) has been defined as one which contains 10 or more cysts 2-8mm in diameter arranged peripherally in the ovary with an increase in the ovarian stroma (Adams *et al.*, 1985; Adams *et al.*, 1986). Ovarian volume was also calculated from the measurements obtained from the ultrasound scan. Women were assigned as hirsute if they obtained a score of at least 8 using the Ferriman-Gallwey index (Ferriman and Gallwey, 1961) and non-hirsute if they scored 5 or less. The presence or absence of acne was also assessed for the women. Women with oligomenorrhea (intermenstrual interval of > six weeks) and amenorrhea (intermenstrual interval > 6 months) were designated anovulatory. In the pedigree data set, assignment of affected status was based on the presence of a polycystic ovarian morphology. In the case control data set, women with polycystic ovaries were further defined as being either symptomatic (designated "PCOS") (i.e. in addition to having polycystic ovaries they presented with either symptoms of anovulation and / or hirsutism), or asymptomatic ("aPCO") if they were non-hirsute and had regular menstrual cycles.

2.1.1 Pedigree data set

Twenty three PCOS/MPB pedigrees (159 members) were identified from probands presenting at the Reproductive Medicine and Endocrinology clinics at St. Mary's Hospital. Probands had presented with menstrual disturbances and or hirsutism and all had bilateral polycystic ovaries on ultrasound scan. Other causes of anovulation and hirsutism, such as Cushing's syndrome (excluded on the basis of other distinctive associated clinical features) and late-onset congenital adrenal hyperplasia (CAH), were excluded. CAH was excluded by carrying out, in the first instance, a single unstimulated measurement of 17-hydroxyprogesterone, if this was found to be abnormal (elevated) then a stimulated 17-hydroxyprogesterone test (short synachten test) was carried out to confirm the diagnosis of CAH.

Diagnostic criteria

Assignment of affection status in women of reproductive age was based solely on the presence of bilateral polycystic ovaries on ultrasound scan (refer to **section 1.1.3**). Women with normal ovarian morphology were assigned as normal. The only exception to this strict diagnostic scheme was the assignment of affection status in postmenopausal women, in whom ovarian imaging is not informative given reduced ovarian activity. In such women, individuals were assigned as affected on the basis of a clear history of previous menstrual dysfunction and/or hirsutism: by this criteria, seven post menopausal women were assigned as affected. Pre-menarchal women and post-menopausal women, with a negative history, were assigned as unknown. There are 64 affected women (probands and relatives) in the pedigree data set, and over 80% of the affected female relatives are symptomatic.

Men were assigned as affected if they demonstrated significant premature male-pattern baldness (MPB), defined as greater than a revised Hamilton IIa score (Lesko *et al.*, 1993), before the age of forty. If MPB was found to be present, age of onset was estimated using both historical photographic evidence and subjective assessment.

Our previous definition of onset before the age of 30 years (Carey *et al.*, 1993), has not been used because during the collection of further families we have observed that within certain families the age of onset of hair loss is between 30 to 40 years. This observation is generally conserved within a single pedigree (eg. in pedigrees 17 and 23, all the affected men had an onset of hair loss between the ages of 30 to 40 years). Thus to avoid the loss of power in those pedigrees, we have used the Ferriman and Purdie definition of MPB (Ferriman and Purdie, 1979), which is the onset of significant fronto-parietal hair loss before the age of 40. Men over the age of forty without signs of hair loss were assigned as unaffected and men under the age of forty without hair loss were assigned as unknown. There are a total of 26 men assigned as affected, 10 of whom had an onset of MPB between the ages of 30-40 years.

All family members underwent a full screening procedure. Measurements of height, weight and waist/hip ratios were taken. Blood samples were taken from all screened members of the pedigrees and from probands and normal controls in the case control data set (see below, **section 2.1.2**) for DNA analysis as well as for the measurement of various serum indices. Members of the pedigree study had a 12 hour fasting blood sample taken for an extensive range of hormone and lipid assays: Follicle stimulating hormone (FSH) and luteinizing hormone (LH), total testosterone, 17-hydroxyprogesterone, androstenedione, DHEAS, total insulin, specific insulin, split-proinsulin, and proinsulin, C-peptide, glucose, cholesterol, Tri-G, HDL-C, HDL-3C, HDL-2C, LDL-C, apolipoprotein A1, apolipoprotein B and lipoprotein (a).

Pedigree data set size and structure

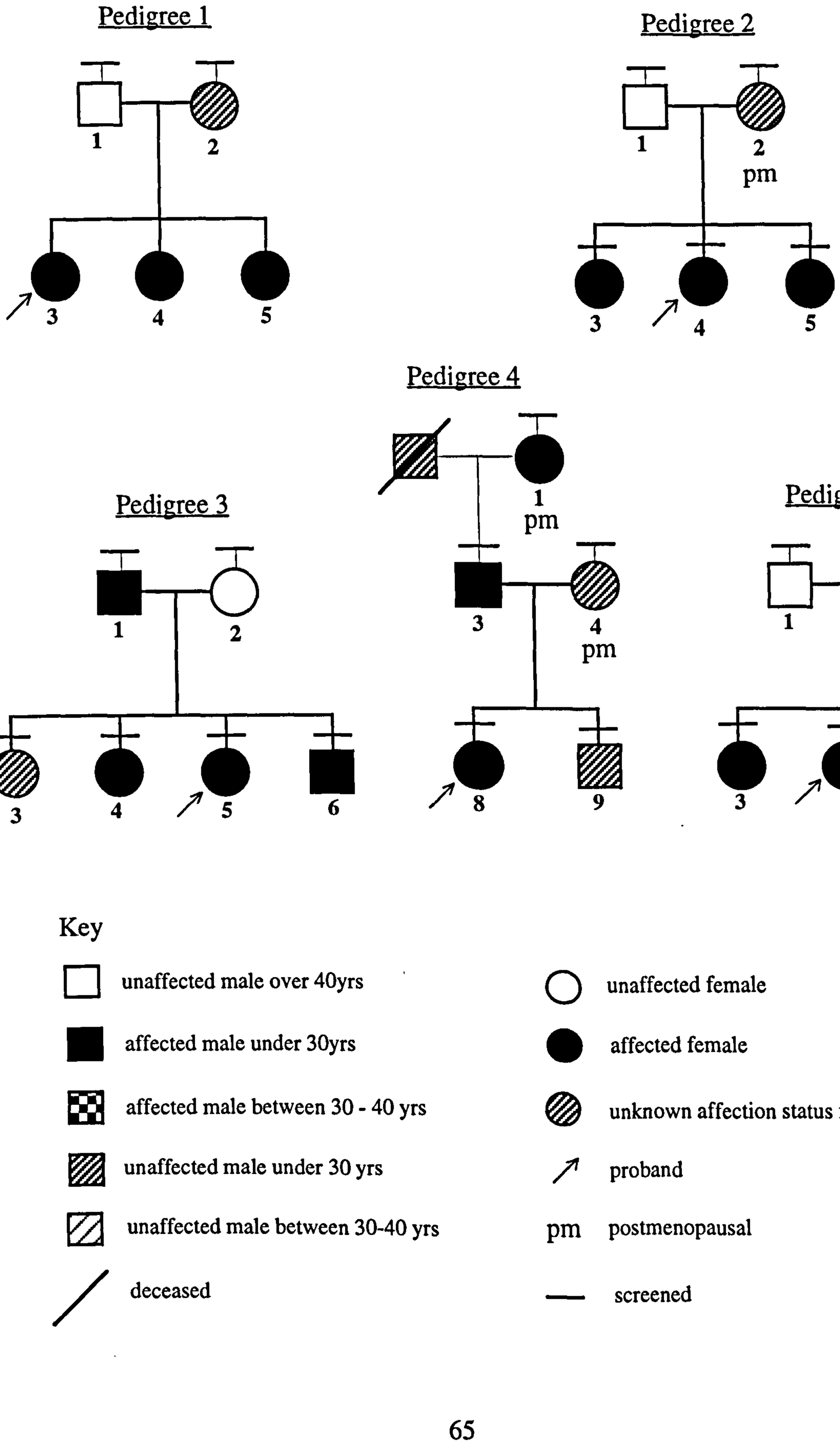
Twenty of these pedigrees are the same as those collected and studied by Carey *et al.* (Carey *et al.*, 1993; Carey *et al.*, 1994). The three further families, ascertained in a similar fashion, were added during the course of this study. Eighteen of the pedigrees were Europid, two were Iranian, two Indian and one Afro-Caribbean. Pedigree sizes ranged from 4 to 20 members. The size and structure of the individual pedigrees is summarised in **table 2.1.1** and illustrated in **figure 2.1.**

Table 2.1.1 Summary of affection status in members of the pedigrees.

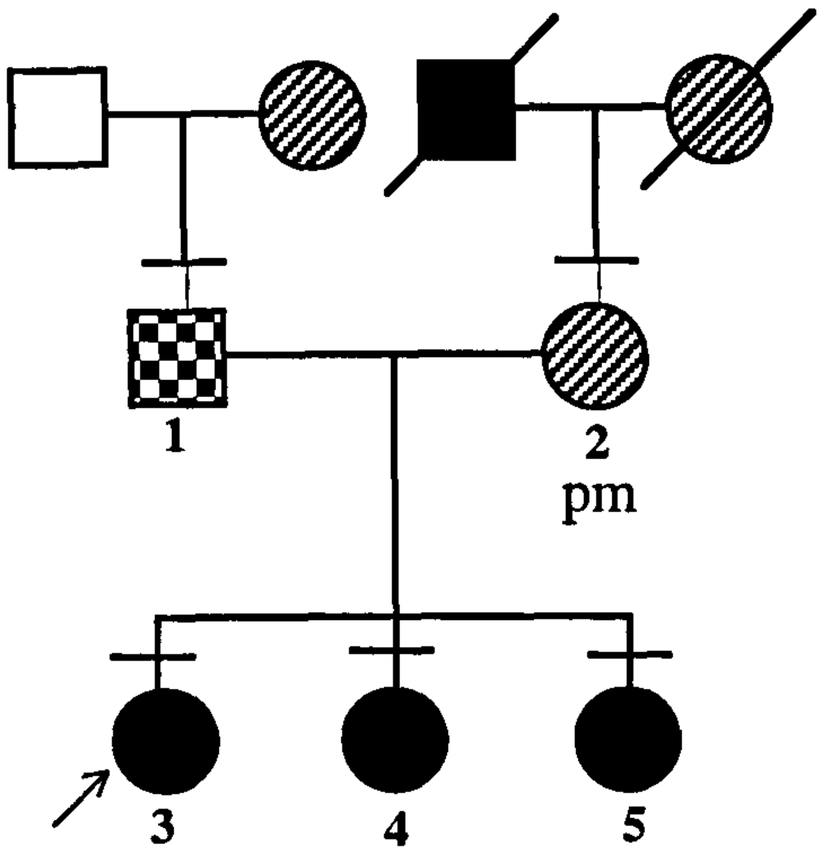
Ped. no.	Affected female	Unknown female	Unaffected female	Affected male	Unknown male	Unaffected male	Total no./ped.	Ethnic origin
1	3	1	0	0	0	1	5	Iran.
2	3	1	0	0	0	1	5	Eur.
3	2	1	1	2	0	0	6	Eur.
4	2*	1	0	1	2	0	6	Eur.
5	4*	0	0	0	0	1	5	Eur.
6	3	1	0	1	0	0	5	Eur.
7	3*	0	2	1	0	1	7	Eur.
8	1	0	2	1	0	0	4	Af-car
9	3	1	0	1	0	1	6	Ind.
10	1	1	0	1	1	2	6	Eur.
11	4*	2	0	1	0	1	8	Eur.
12	2	2	0	0	2	0	6	Eur.
13	2	1	0	1	2	0	6	Iran.
14	4	0	1	1	0	0	6	Eur.
16	2	1	0	2	0	0	5	Eur.
17	7	4	3	4	2	0	20	Ind.
18	1	1	1	1	1	0	5	Eur.
19	1	1	0	3	0	0	5	Eur.
20	6*	1	2	1	0	1	10	Eur.
22	3	1	0	1	0	1	6	Eur.
23	1	3	3	3	0	2	12	Eur.
24	2	2	0	1	1	2	8	Eur.
25	4	0	1	1	0	1	7	Eur.
TOTAL	64	26	16	28	11	15	159	

Refer to diagnostic criteria above for the assignment of affected status. For each pedigree, the number of "affected females" (PCO) including the proband, "unaffected females" (normal ovaries) and "unknown females" (ovaries not scanned or postmenopausal with no history of symptoms), has been given together with information on the number of "affected males" (MPB), "unaffected males" (no hairloss before the age of 40 years) and "unknown males" (before the age of 40 without hairloss or hairloss status unknown). The total number of individuals in each pedigree ("total no. / ped.") and the total number for each category for the total data set ("TOTAL"), is also given. * These pedigrees include the assignment of a postmenopausal women (pm) as affected. In pedigrees 4,5,7 and 11 there is only one pm affected women and in pedigree 20 there are 3. Ethnicity of each pedigree is also given.

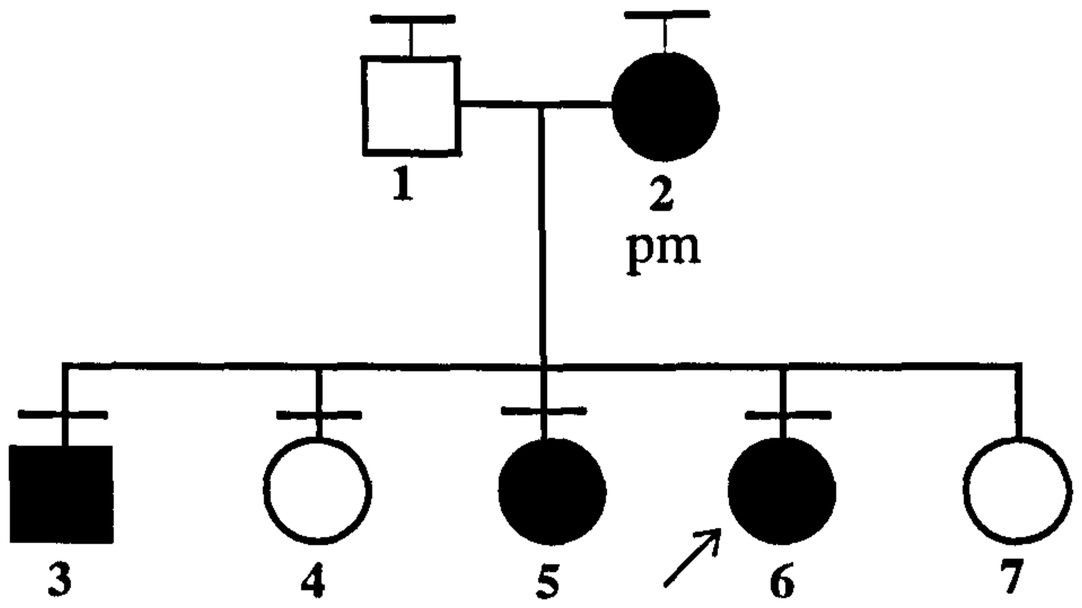
Figure 2.1 Pedigree trees



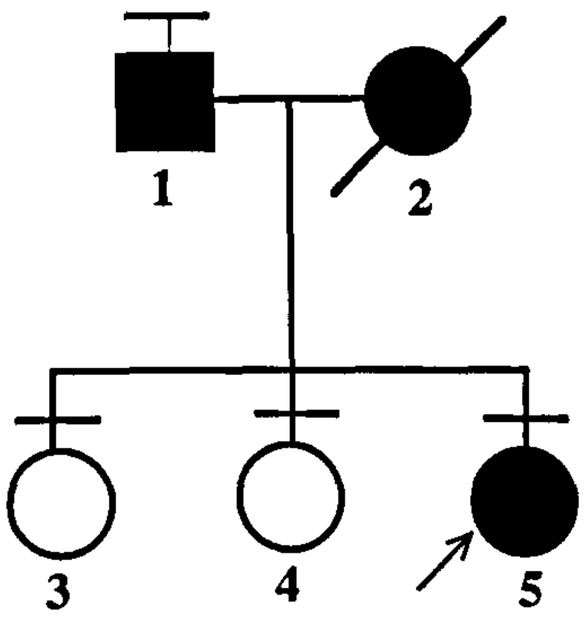
Pedigree 6



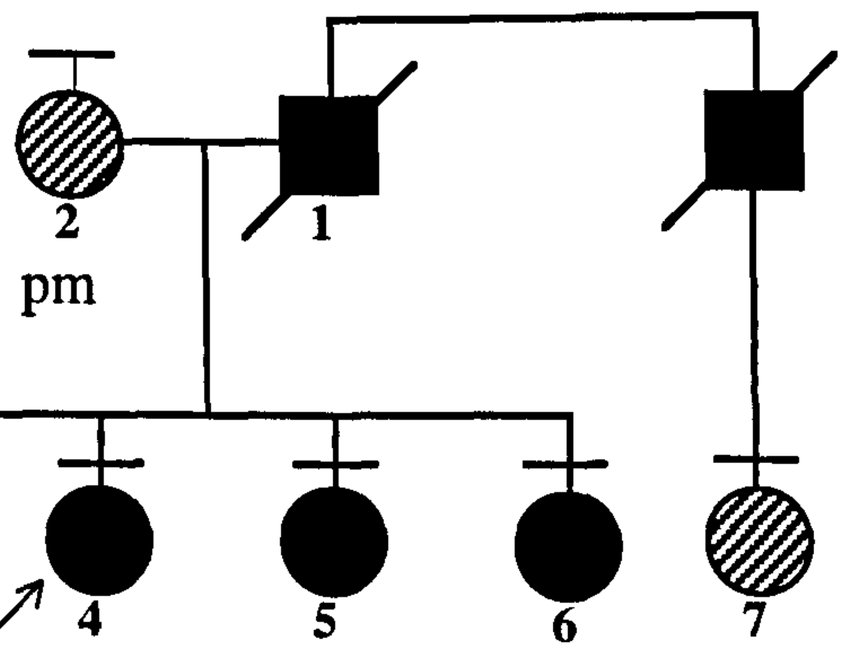
Pedigree 7



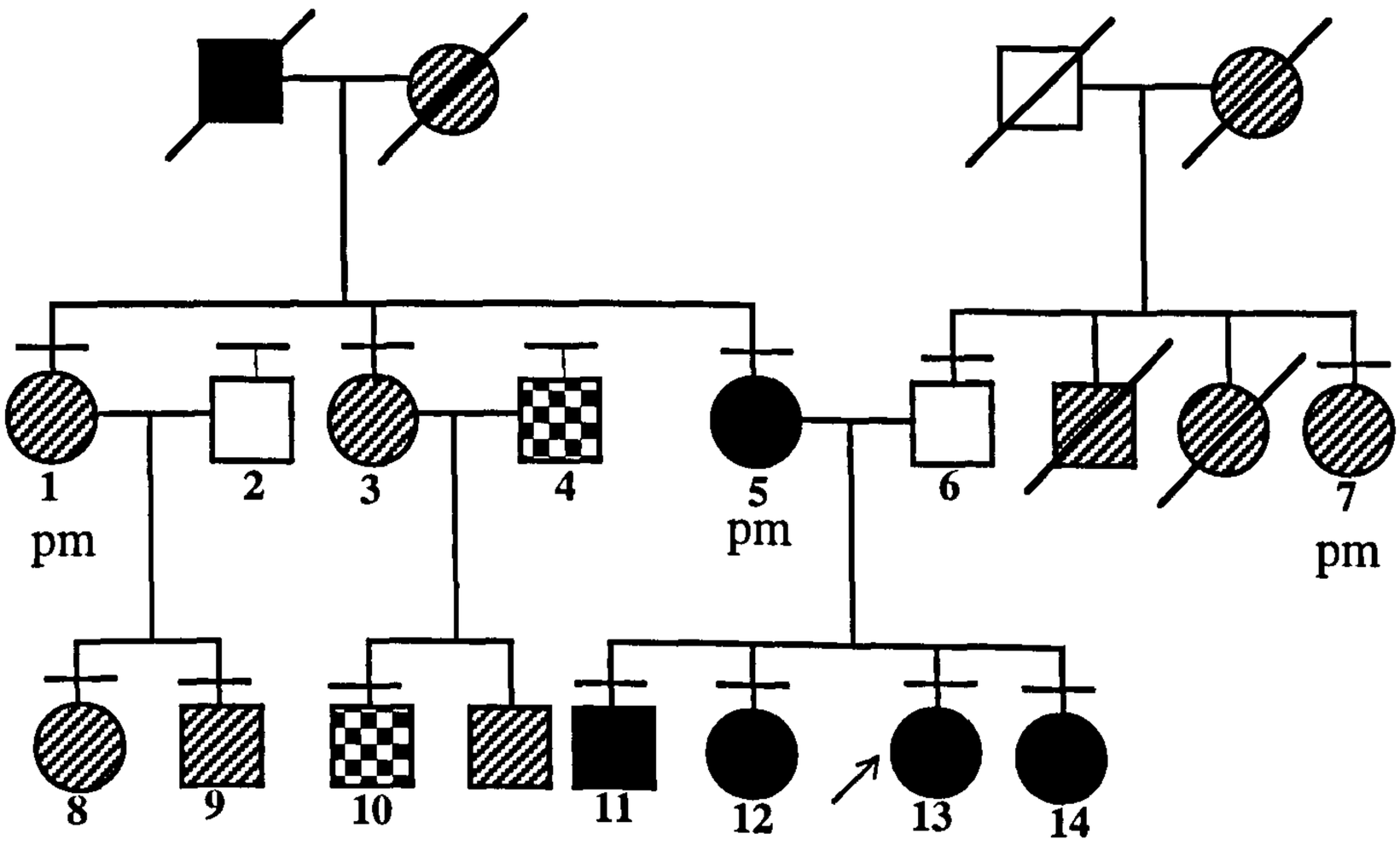
Pedigree 8



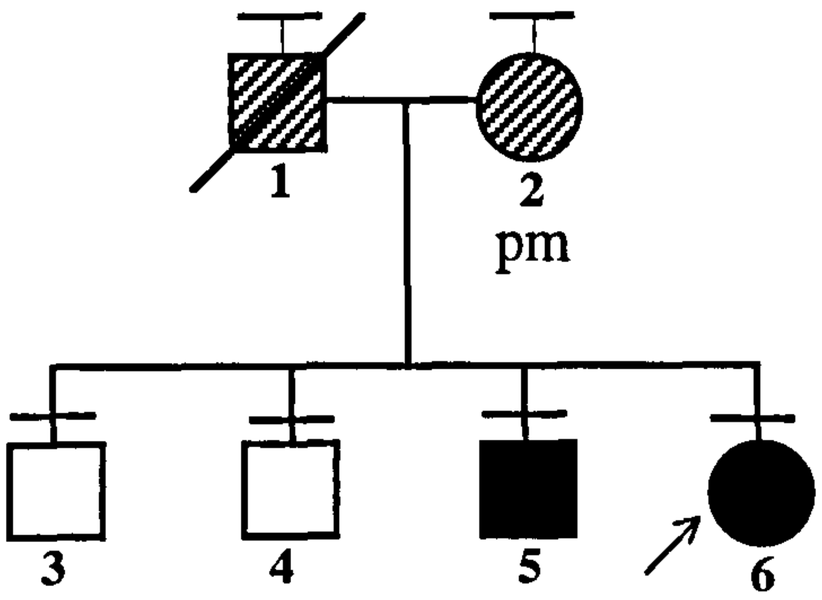
Pedigree 9



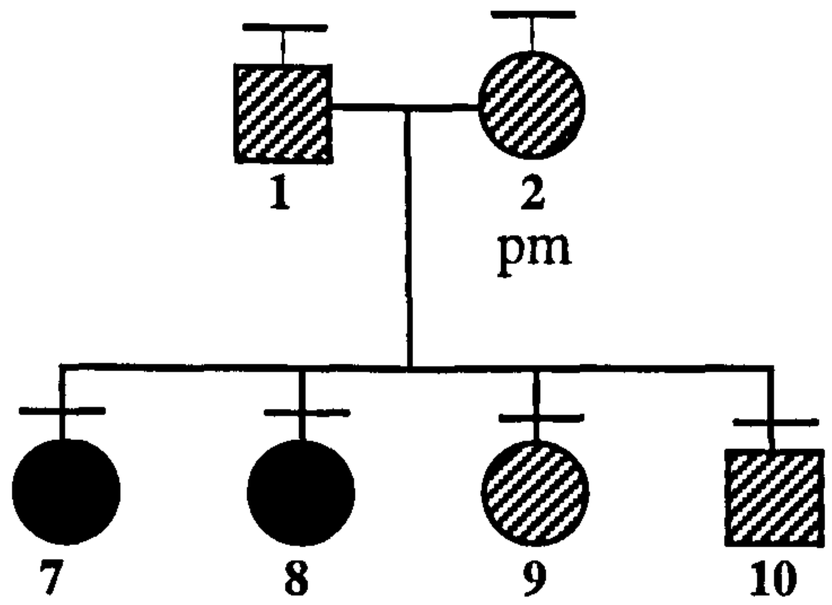
Pedigree 11



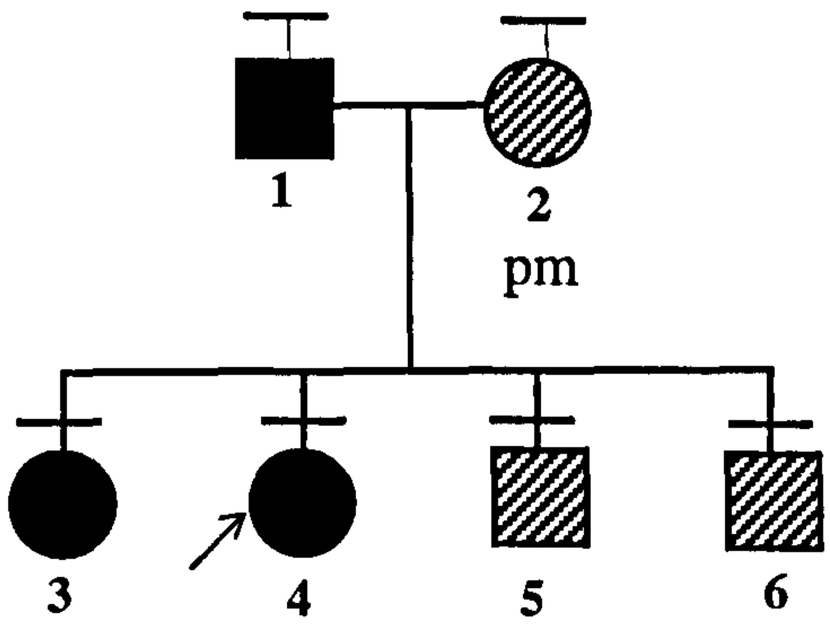
Pedigree 10



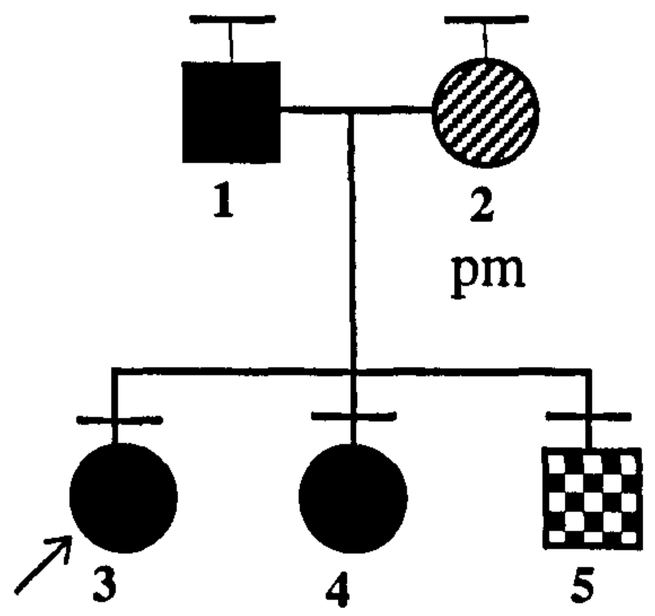
Pedigree 12



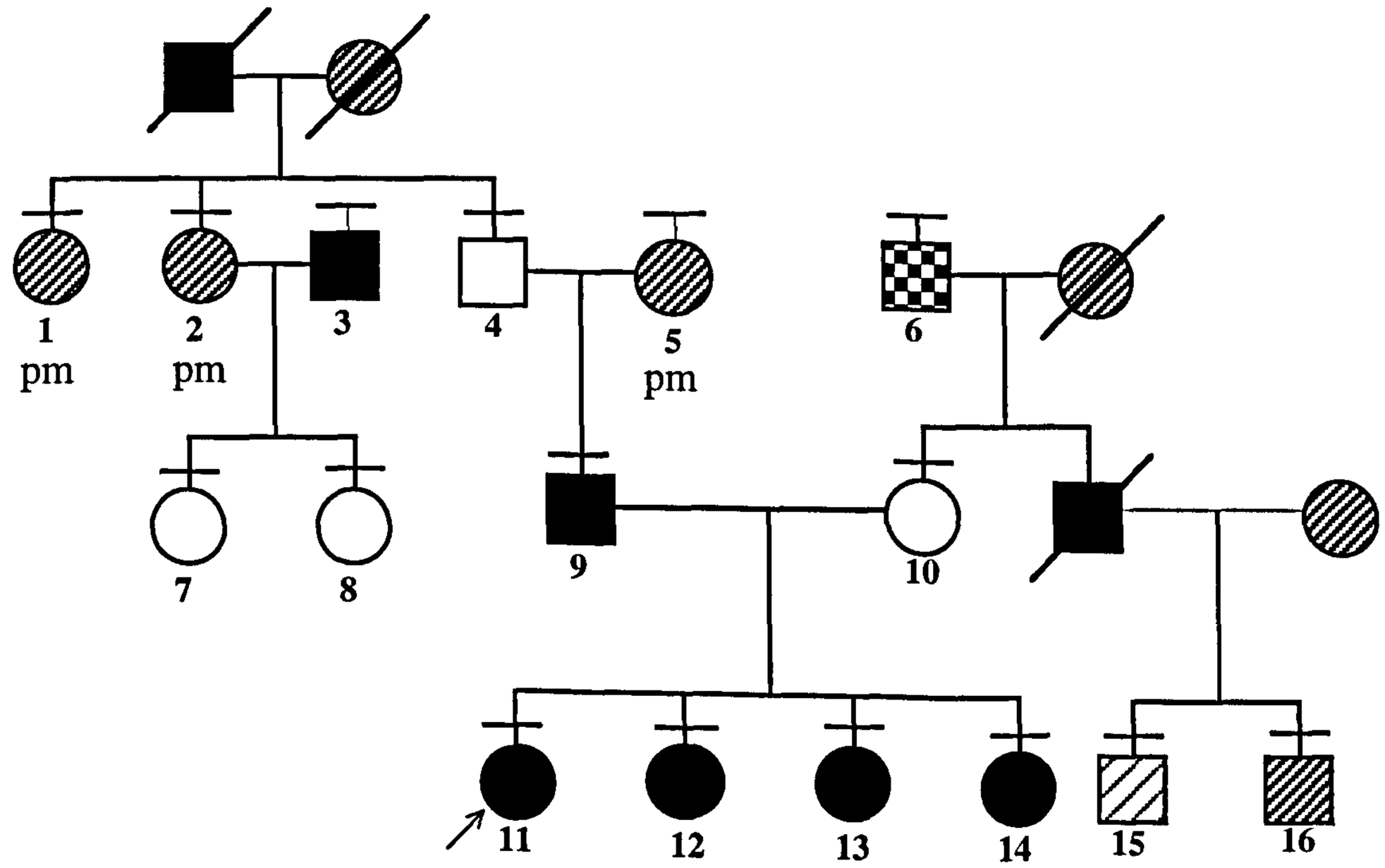
Pedigree 13



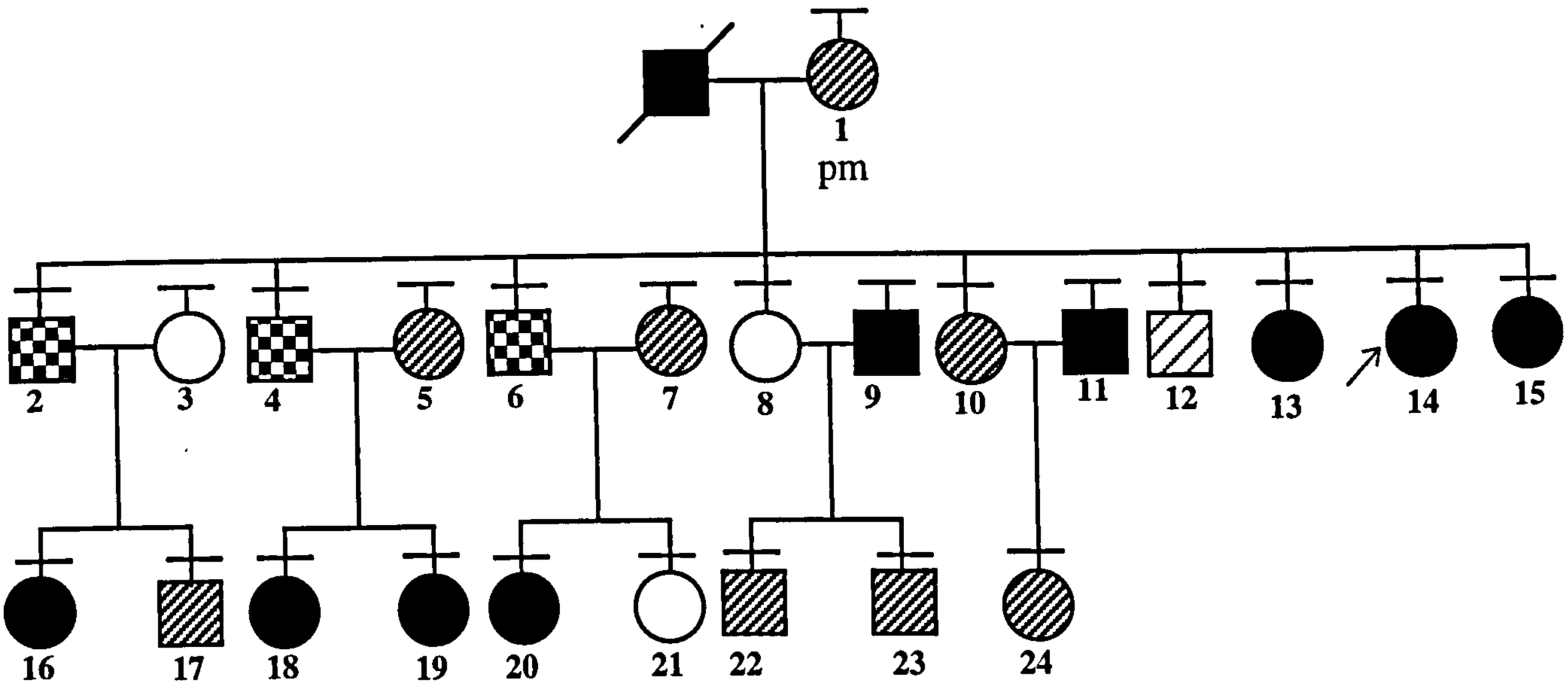
Pedigree 16



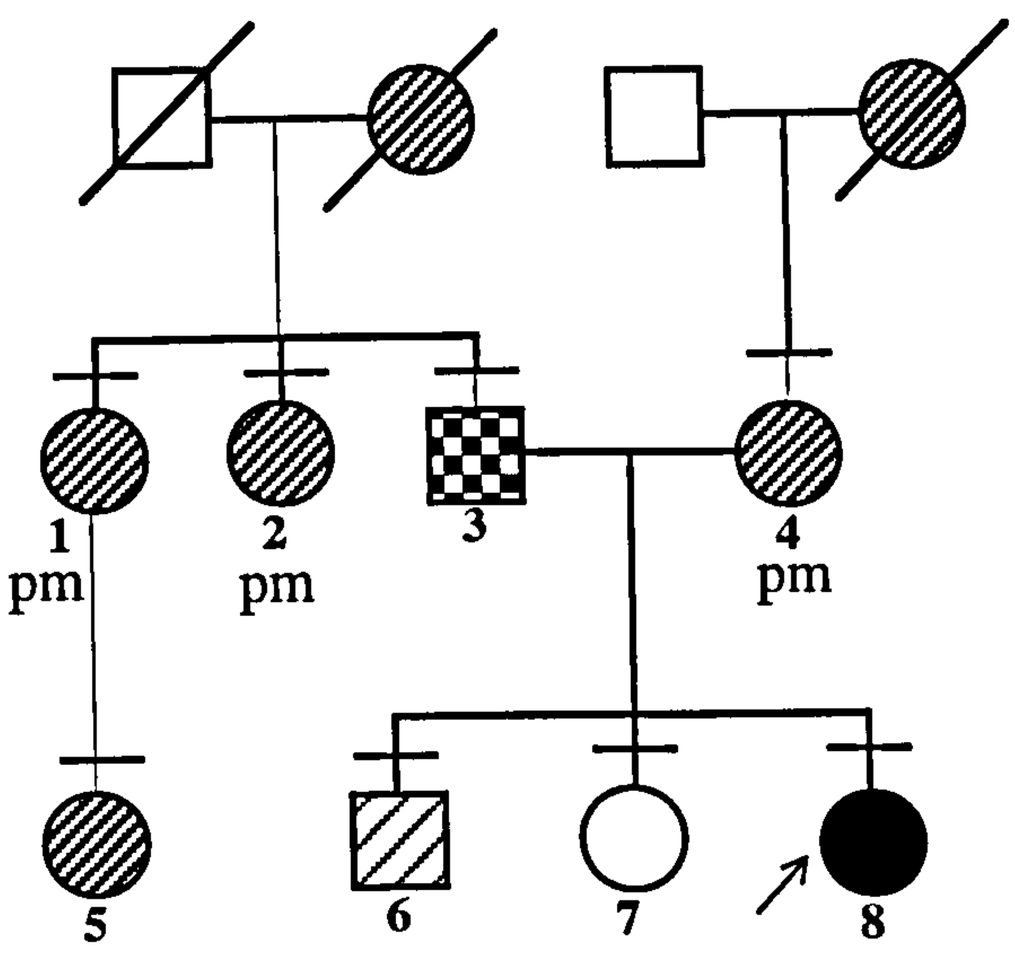
Pedigree 14



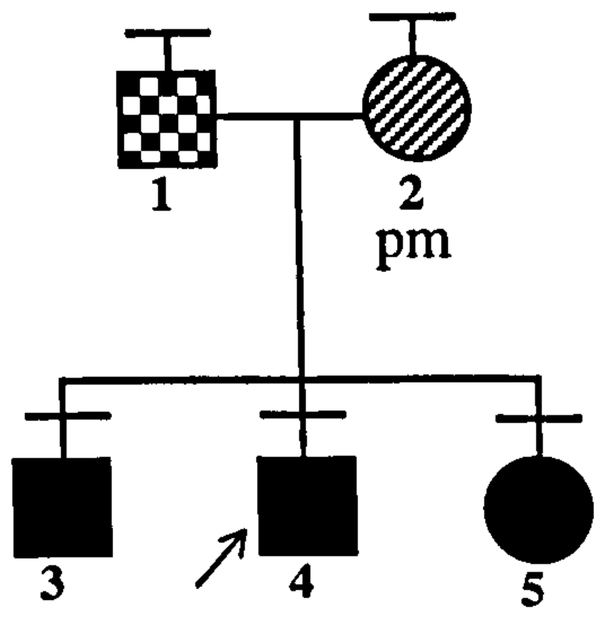
Pedigree 17



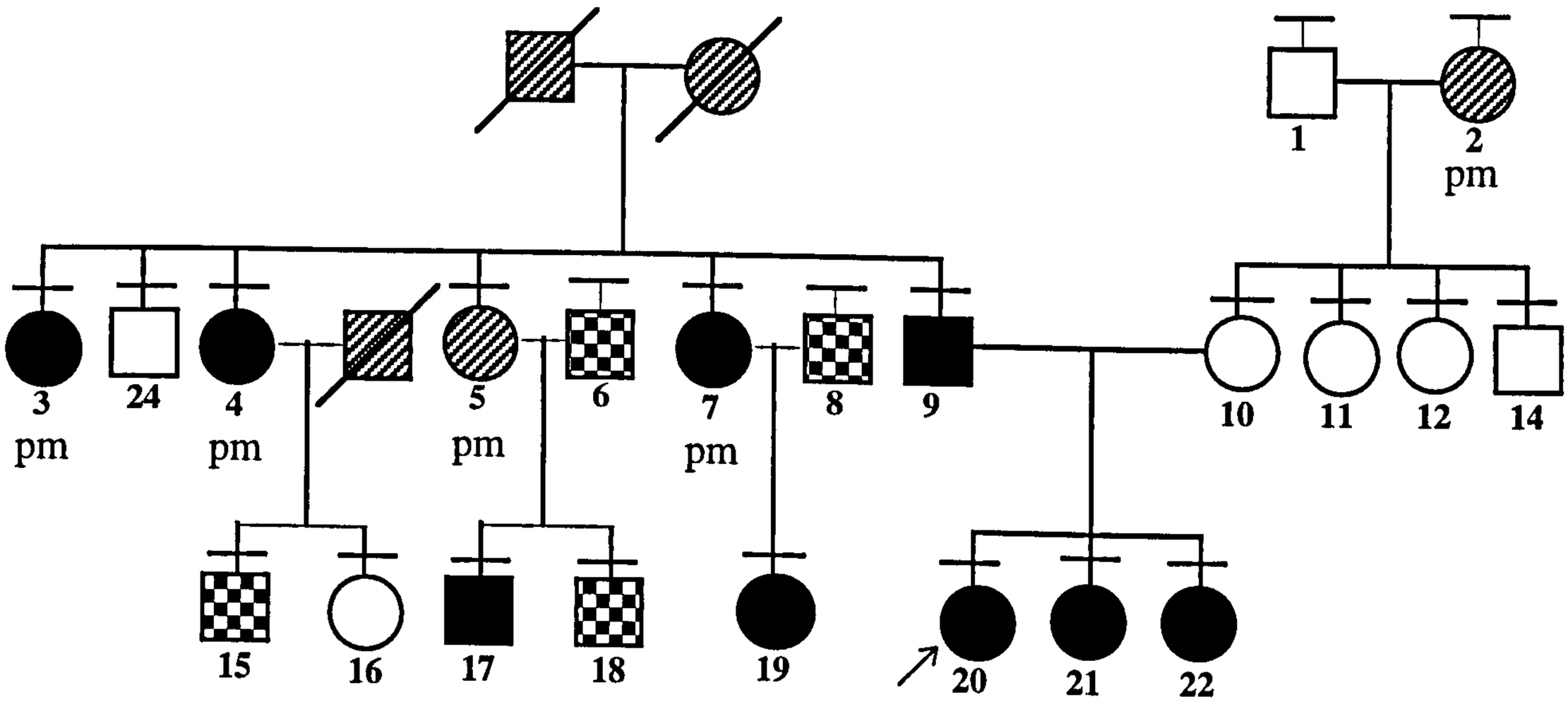
Pedigree 18



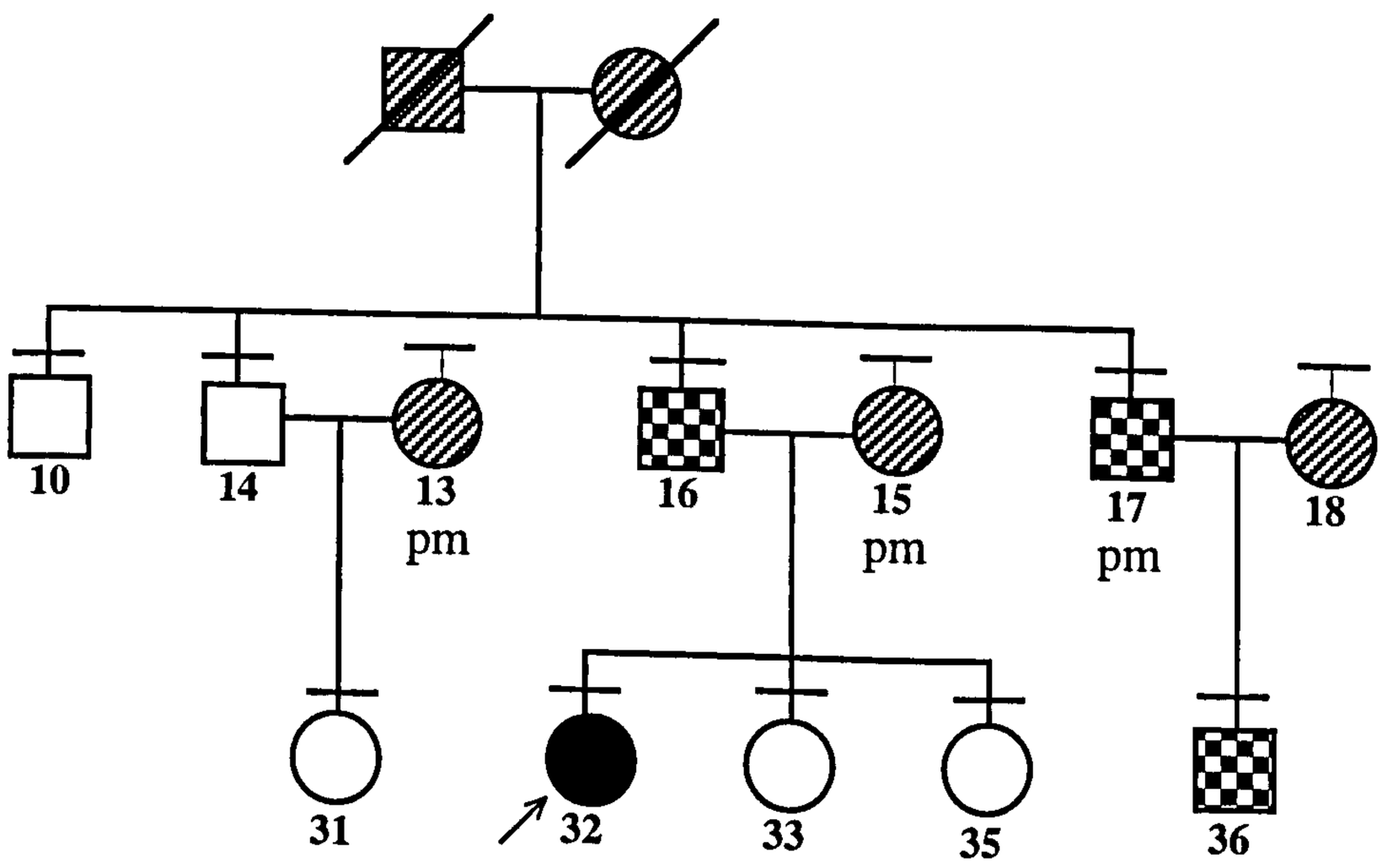
Pedigree 19



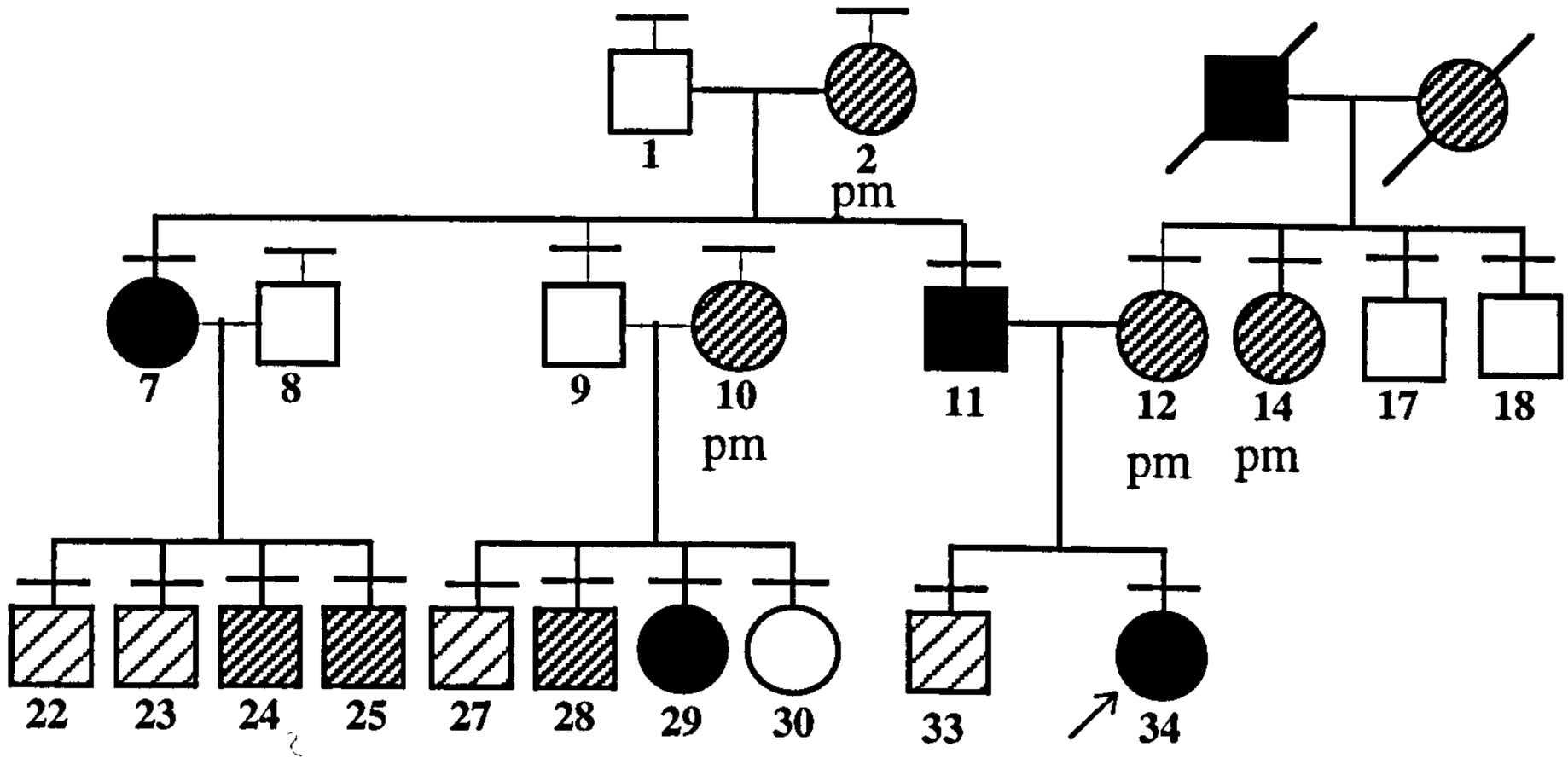
Pedigree 20



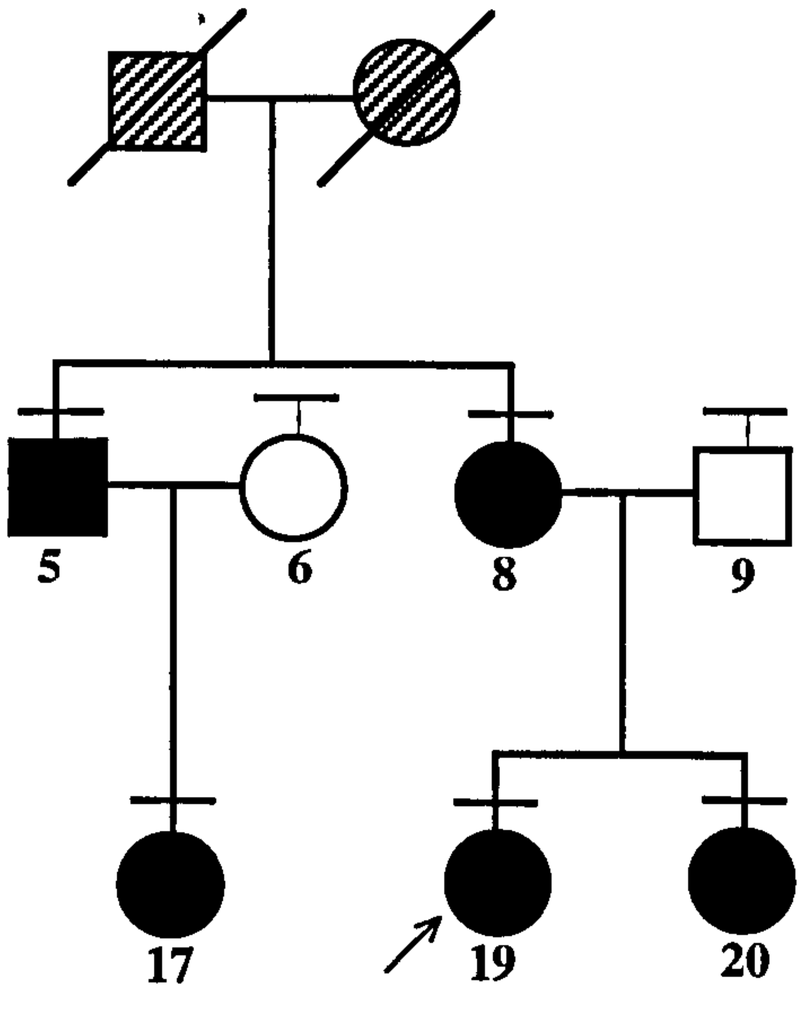
Pedigree 23



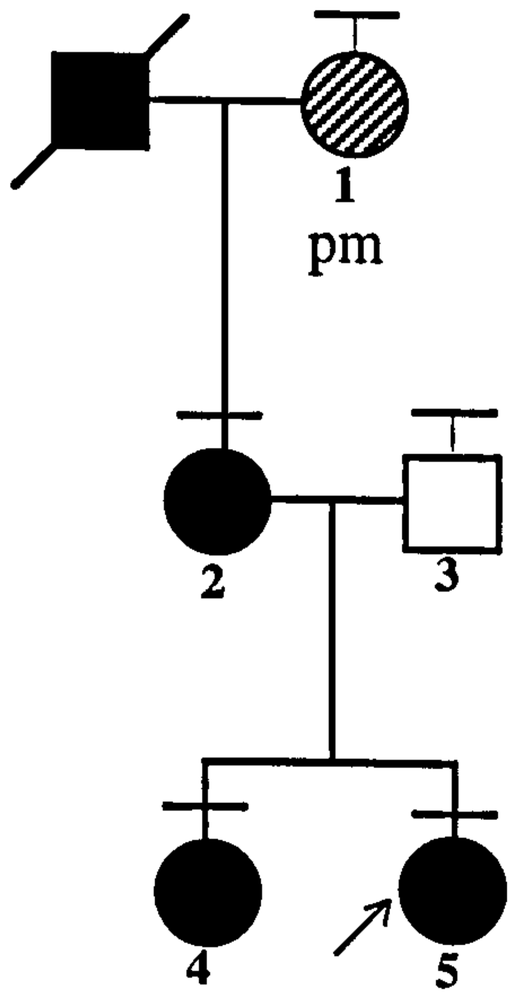
Pedigree 24



Pedigree 25



Pedigree 22



2.1.2 Case control data set

Women in the case control study were all Europid and were collected consecutively from two centres: the Reproductive Medicine and Endocrinology clinics at St. Mary's Hospital, and from the Endocrinology clinics at the Middlesex Hospital, London. This data set consists of 150 probands who presented with anovulation and/ or hirsutism and who were found to have polycystic ovaries on ultrasound scan, designated "PCOS". In addition, 60 non-hirsute women having regular menstrual cycles and polycystic ovaries (referred to as "asymptomatic PCO" or "aPCO") were collected together with 67 control subjects with regular menstrual cycles and normal ovarian morphology on ultrasound scan ("normal controls"). The latter two groups were all collected from St Mary's Hospital and were either hospital staff volunteers or patients with non-PCOS related presentations such as tubal / or male-related infertility. Of the total 277 women in this data set, only 204 had been definitely assessed for the presence or absence of hirsutism. Of these, a total of 78 were assigned as hirsute (71 with polycystic ovaries and 7 with normal ovaries) and 126 were non-hirsute (97 with polycystic ovaries and 29 without). Measurements of height, weight, total testosterone and day eight gonadotrophins were available in the majority of cases. **Table 2.1.2** gives a summary of the structure of this data set.

Table 2.1.2 Summary of symptomatology in the case control data set.

	ovulatory / hirsute	ovulatory / non-hirsute	ovulatory / ?	anov. / hirsute	anov. / non-hirsute	anov. / ?	Total
PCOS	14	-	-	57	59	20	150
aPCO		38	22	-	-	-	60
normal controls	7	29	31	-	-	-	67
TOTAL	21	67	53	57	59	20	277

For each of the groups ("PCOS", "aPCO" and "normal controls") a breakdown of symptomatology is given, such that: individuals with "ovulatory / hirsute" have regular ovulatory cycles and are hirsute, those with "ovulatory / non-hirsute" have regular ovulatory cycles and are non-hirsute and those with "ovulatory / ?" have regular ovulatory cycles but there is no information on hirsutism status. Anovulatory women are classified in a similar way as either "anov. / hirsute", "anov. /non-hirsute" or "anov./?".

2.2 MATERIALS

2.2.1 Markers for linkage studies

Chromosome 15

The genetic and physical maps of chromosome 15 are shown in **figure 2.2.1,** indicating the position and order of the microsatellite markers used in the chromosome 15 linkage analysis. A summary of the marker information for these markers is also given in **table 2.2.1.**

Figure 2.2.1 Physical and genetic maps of chromosome 15.

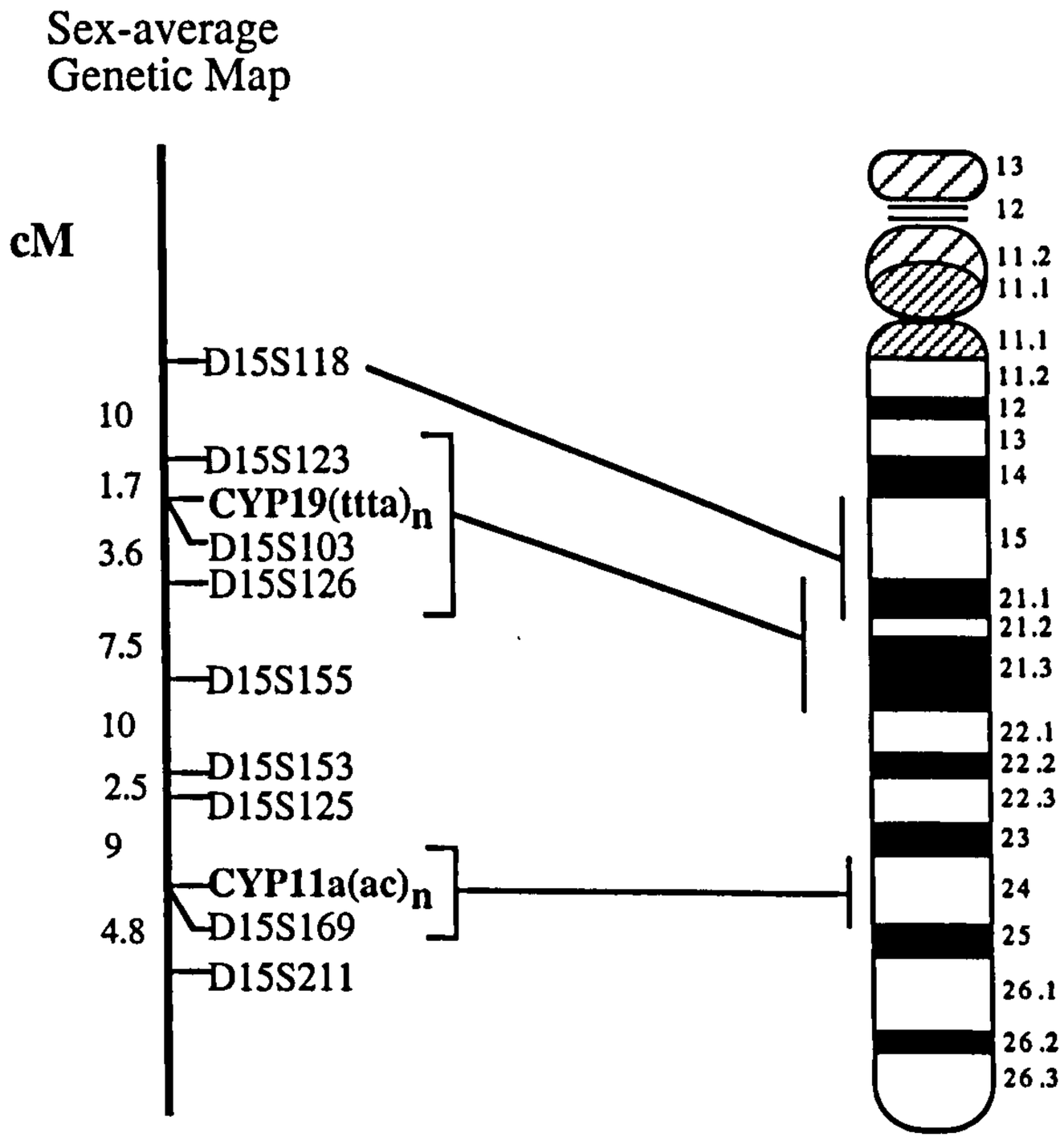


Figure 2.2.1 Chromosomal order and interval distances were obtained from the chromosome 15 genetic map at: http://cedar.genetics.soton.ac.uk/pub/chr15. All D15S markers are Généthon microsatellite markers (Gyapay *et al.*, 1994), *except markers D15S169 and D15S103* (Beckmann *et al.*, 1993). *For each of the above two loci an intragenic microsatellite has also been used in the linkage analysis, CYP19 (ttta)$_n$* (Polymeropoulos *et al.*, 1991) *and CYP11a (ac)$_n$* (Beckmann *et al.*, 1993).

Table 2.2.1 Chromosome 15 microsatellite marker information.

Marker Name	Locus Symbol	Primers	Heterozygosity	no. Alleles
D15S118	AFM112xa1	TCAAAGACCCATATCAACA* GTGCTGAAAAGCGACACTA	0.81	6
D15S123	AFM205ye1	AGCTGAACCCAATGGACT* TTTCATGCCACCAACAAA	0.81	8
CYP19(ttta)$_n$		GCAGGTACTTAGTTAGCTC[a] TTACAGTGAGCCAAGGTCT	0.72	8
D15S103	G113	GTTCAAAAGGCTCATCATGTAGC* TTCCTGAGTCCTTATAGCTTCCA	0.73	12
D15S126	AFM218yf12	GTGAGCCAAGATGGCACTC* GCCAGCAATAATGGGAAGT	0.81	11
D15S155	AFM211zc1	TTTTCTAGGCAGGTAGTCCCA* GATTTCCATAGCACACATTTGAGT	0.73	7
D15S153	AFM205ye3	AGTACCTGAAAGGGTGGG* GATCAGTGTAGGCTCCAAA	0.79	12
D15S125	AFM214xd10	TTCCACACATGACCGC* CCCCTGAAGACCGTGA	0.85	7
CYP11a(ac)$_n$		GGTCAAGTGGCTGTGTAAGGA* TTTAAATGGGTGAATGTATGGTG	0.63	7
D15S169	Utsw1591	CAGGAGAGAGCCTTGGAT* GAGACATCTCTTCTGAAAGCTC	0.74	9
D15S211	AFM323yd9	AAGCAGGTGGAATCCTTG* AAAAGCCCCAGGTAGGG	0.94	22

"Marker Name" is the generic name for the particular locus. "Locus Symbol" is the nomenclature used by the particular genome organisation, such that all Généthon markers have "AFM..." locus symbol, G113 and Utsw1591 are HUGO locus symbols. Heterozygosity is the observed frequency of heterozygotes (usually ascertained using CEPH pedigrees) and is a measure of the informativeness of the marker (1 being the maximum). No. of alleles is the number observed in the control population (eg. CEPH). All markers are dinucleotide(ac)$_n$ microsatellite markers (CA-strand primer), except CYP19(ttta)$_n$ which is a tetranucleotide (ttta)n microsatellite markers ([a] TTTA-strand primer).*

Chromosome 2

Figure 2.2.2 Physical and genetic maps of chromosome 2.

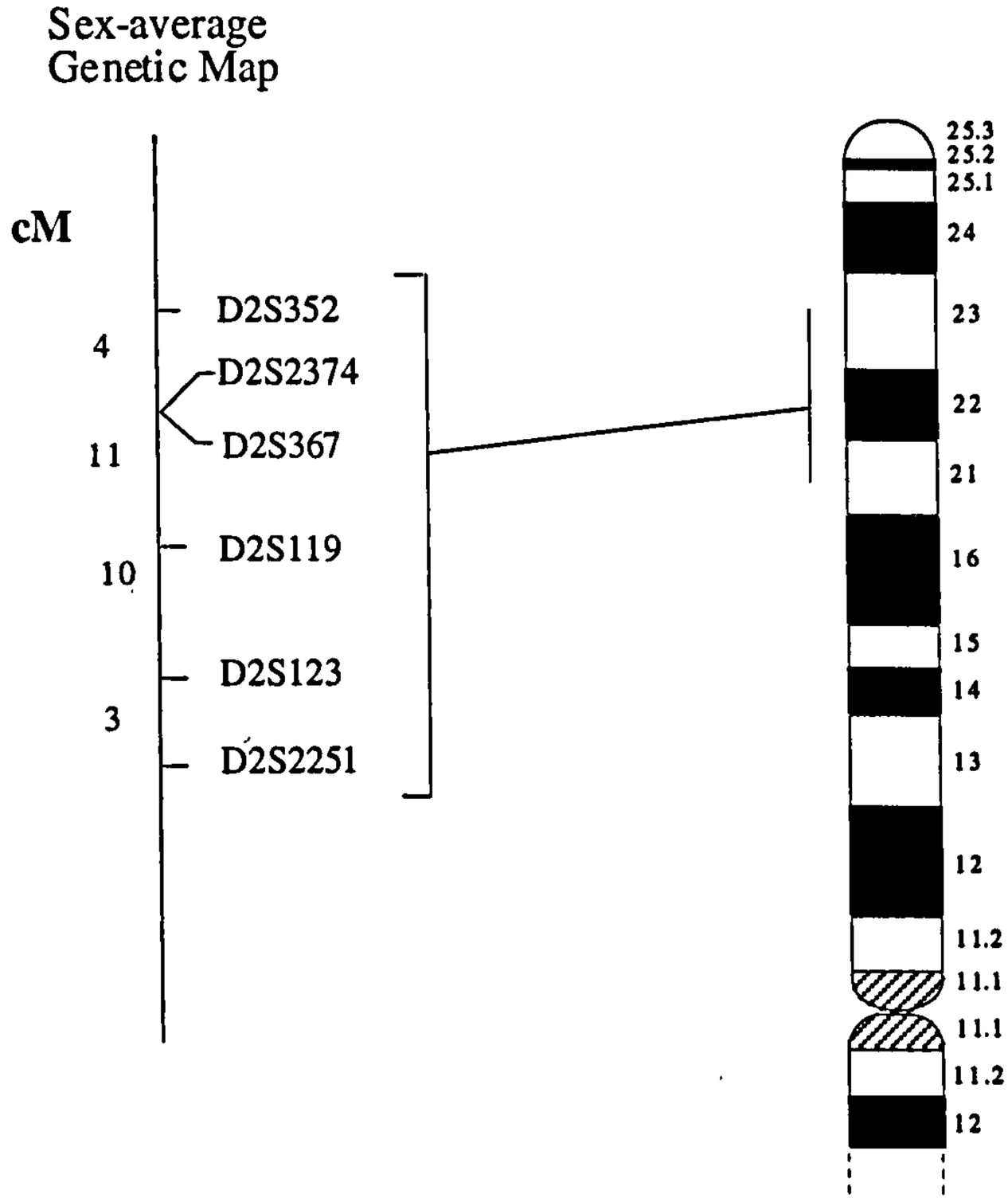


Figure 2.2.2 All markers are Généthon microsatellite markers and the Chromosomal order and interval distances were obtained from the 1996 Généthon map (Gyapay *et al.*, 1994).

Table 2.2.2 Chromosome 2 microsatellite marker information.

Marker Name	Locus Symbol	Primers	Heterozygosity	no. Allel
D2S352	AFM296	GCAAAGTCGTTCTCAGGTG* CTACAGGGCTTCAGCATCC	0.83	13
D2S367	AFM303ze1	TTCTTTGGTCTAAGGGTCAC* AGCTTCTTGTTCACAGGTGT	0.86	13
D2S2374	AFMa066ze9	TCCTATGATAATTGTTAAGGTTTG* AGAATGGCACCCCTAA	0.85	13
D2S119	AFM077yb7	CTTGGGGAACAGAGGTCATT* GAGAATCCCTCAATTTCTTTGGA	0.81	10
D2S123	AFM093xh3	AAACAGGATGCCTGCCTTTA* GGACTTTCCACCTATGGGAC	0.77	6
D2S2251	AFM205te7	AATGGATCTTTACGAAAAGG* AAACTACCCCTGTTGAGTGA	0.78	7

Refer to legend of Table 2.2.1 for abbreviations. All the above markers are dinucleotide (AC)n repeat markers (CA-strand primer).*

2.2.2 Chemicals and reagents

Chemicals

Most chemicals were AnalaR grade and were obtained from British Drug House (BDH), U.K. Others, such as, ethidium bromide (EtBr), sodium dodecyl sulphate (SDS), mineral oil, NaCl, xylene cyanol and bromophenol blue were from Sigma Chemical Company, U.S.A and ammonium persulphate was from Kodak, U.S.A.

Nucleic Acids and oligonucleotides

Molecular weight size standards were from Gibco BRL, U.K. or Promega, U.K. Oligonucleotides (used as PCR primers), were synthesized, either by Pharmacia, U.K. or by Genosys, U.S.A. 2'-deoxyribonucleoside 5'-triphosphates (dNTPs) were purchased from Pharmacia. The M13 sequencing ladder was supplied as part of the Sequenase version 2.0 kit, by United States Biochemical Corporation (USB), U.S.A.

Reagents for sequencing

Dynabeads M-280 streptavidin coated beads were supplied by Dynal, Norway. The Wizard PCR purification kit was purchased from Promega, U.S.A. The Sequenase version 2.0, used for both the *CYP11a* promoter sequencing and to produce the M13 sequencing ladder, was supplied by USB.

Enzymes

Restriction enzymes were either from Boeringer Mannheim, Germany, or New England Biolabs (NEB), U.S.A. AmpliTaq DNA polymerase was from Perkin-Elmer Cetus, U.S.A and Biotaq polymerase or Bioline, U.K. Proteinase K was supplied by BDH.

Stanford G3 radiation hybrid panel

The Stanford G3 RH panel was supplied by Research genetics, U.S.A.

Autoradiography and photography

X-ray film was Hyperfilm from Amersham and Polaroid film was from Polaroid, UK.

Radioisotopes

[α^{33}P]-dATP (1000Ci/mMol) was purchased from Amersham International, U.K..

Electrophoresis reagents

Agarose was purchased from Gibco BRL, NuSieve™ agarose was from FMC BioProducts, U.S.A and polyacryamide was Sequagel sequencing system from National diagnostics, U.K. The S2 vertical sequencing gel tanks and the horizontal agarose gel tanks were purchased from Gibco BRL.

2.2.3 Solutions

Agarose gel loading buffer

0.25% bromophenol blue

0.25% xylene cyanol FF

30% glycerol in water

Formamide loading buffer

99% deionized formamide

5mM EDTA

0.1% xylene cyanol

0.1% bromophenol blue

***Lysis buffer*:**

10 ml Tris-HCL pH8

5ml 1M $MgCl_2$

10 ml Triton-X 100

Make up to 1L with distilled water and store at 4°C. Add 109.5g of sucrose to this solution just before use.

Polyacrylamide gel solution:

46ml diluent

17ml concentrate

7ml 10 × TBE

400µl ammonium persulphate

50µl Temed (N,N,N'-Tetramethylethylenediamine)

Proteinase K solution:

1% SDS

2mM EDTA

2mg/ml final of Proteinase K

T20E5:

20mM Tris-HCL

5mM EDTA

TAE:

40mM Tris-Acetate

1mM EDTA pH 8.0

TBE:

89mM Tris-Borate

2mM EDTA pH 8.0

TE:

10mM Tris HCl pH 7.6

1mM EDTA pH 8.0

2.3 METHODS

2.3.1 Extraction of high molecular weight DNA from human nucleated cells

Blood collected into EDTA tubes was gently inverted to avoid clotting and was either used immediately for DNA extraction or was stored in the -20°C freezer. Between 5ml to 10 ml of blood (fresh, or thawed from frozen) was decanted into 50 ml Nunc tubes and ice cold lysis buffer was added to the 30 ml mark. The tube was inverted to mix and then spun at 2000g for 10 minutes. The supernatant was poured off and the pellet resuspended in 10ml lysis buffer. After a 5 minutes incubation on ice, the tubes were spun at 2000g for 5 minutes. The supernatant was poured off to reveal a whitish pellet. The following solutions were added to the pellet and mixed by inversion after adding each solution: 3ml T20E5, 200µl 10% SDS and 500µl Proteinase K solution. The samples were incubated overnight at 47°C. 1ml of saturated NaCl solution was added to each sample and mixed vigorously for 15 seconds. They were then spun at 2500g for 30 minutes at 4°C. Much of the protein precipitates out and a white pellet should be visible. The supernatant was carefully removed to a fresh Nunc tube and the DNA precipitated using 2 volumes of ethanol at room temperature. After mixing, the DNA was visible and was gently spooled using an inoculation loop. After a wash in 70% ethanol, the pellet was air dried for a few minutes and dissolved in 1ml\water or T.E. The DNA was store at 4°C.

2.3.2 DNA quantification

The concentration of extracted DNA was measured by spectrophotometry. The DNA sample was diluted in water by a factor of 1:40. The optical density (OD) was then measured at a wavelength of 260nm (OD_{260}). This measures the concentration of DNA, where an OD_{260} of 1 corresponds to 50µg/ml of double-stranded DNA. DNA concentration was obtained by multiplying $OD_{260} \times 40$ (dilution factor) $\times$ 50µg/ml. The OD_{280} was also obtained (this is a measure of protein content) and an estimate of

the purity of the sample was made by the OD_{260}/OD_{280} ratio, where a value of >1.8 was considered acceptable.

2.3.3 Polymerase chain reaction (PCR)

All polymerase chain reactions of genomic DNA were carried out in a Hybaid OmniGene thermocycler.

Microsatellite markers

For all the microsatellite markers (all chromosome 15 and chromosome 2 markers (**section 2.2.1**) and the *CYP11a* (tttta)n marker (**section 2.3.11**)) the following standard conditions were used:

The reaction mixture consisted of (total volume 12.5µl):

1µl forward primer (25 pmoles)

1µl reverse primer (25 pmoles)

1.25 µl 10 × Parr Excellence reaction buffer

[50mM KCL, 10mM Tris-HCL (pH9.0) 0.1% Triton X-100, 15mM $MgCl_2$]

1.25 µl dNTPs [stock solution of 2mM dTTP/dGTP/dCTP and 250 µM dATP]

5.9 µl filtered autoclaved distilled water

0.1 µl [α-^{33}P] dATP

1 µl genomic DNA (25ng/µl working solution)

1 unit AmpliTaq DNA Polymerase

The above mixture was placed into a small (0.6ml) PCR Eppendorf tube and overlaid with one drop of mineral oil, before placing into the thermal cycler for PCR amplification.

PCR thermal cycling conditions were:

An initial denaturation step of 5 minutes at 94°C followed by 35 cycles of:

- denaturation at 94°C (45 seconds)

- annealing at 54°C (45 seconds)

- extension at 72°C (1 minute)

A final extension step of 10 minutes at 72°C.

The only variation to these PCR reaction conditions was in the annealing temperature required for three of the markers. A temperature of 52°C was used for marker D15S123 and 63°C was used for both D15S103 and *CYP11a* (tttta)n.

After PCR amplification, 3µl of formamide loading buffer were added to 3µl of each of the PCR products. These were denatured for 5 minutes at 90°C and then electrophoresed through a 6% polyacrylamide denaturing sequencing gel (**section 2.3.4**).

RFLP markers and LHR 1a/1b PCR

PCR conditions for the *LHR* 1a/1b primer set used in the RH panel screening (**section 2.3.10**), *CYP17* and *SRD5A1* RFLP assays (**sections 2.3.13**), were carried out in a total volume of 25µl and were set up as follows:

1.25 µl forward primer (30 pmoles)

1.25µl reverse primer (30 pmoles)

2.5 µl 10 × Parr Excellence reaction buffer

1.25 µl dNTPs [stock solution of 2mM dTTP/dGTP/dCTP/dATP]

5.9 µl filtered autoclaved distilled water

2 µl genomic DNA (50ng/µl working solution)

1 unit AmpliTaq DNA Polymerase

The above conditions were used for *LHR* 1a/1b primer set and the *CYP17* RFLP marker amplification. For the *SRD5A1 Nsp I* assay, a different PCR reaction buffer was used (50mM KCl, 10mM Tris-HCl (pH8.3), 15mM $MgCl_2$, 0.01% gelatin) and, for the *SRD5A1 Hinf I* assay, the Bioline PCR reaction buffer (0.5mM $MgCl_2$ final concentration) and Bioline Taq polymerase were used.

The PCR cycling conditions were the same as for the microsatellite marker PCR amplifications given above, except that for the *SRD5A1 Nsp I* and *Hinf I* markers an annealing temperature of 68°C was used and the time for the extension step of the cycles was increased from 1 minute to 3 minutes.

PCR conditions for S1, S2 and S3 (sequencing of CYP11a promoter)

For each PCR reaction, the100µl amplification mixture contained:

5 µl forward primer (125 pmoles)

5 µl reverse primer (125 pmoles)

10 µl 10 × Gibco BRL PCR reaction buffer

[50mM KCL, 10mMTris-HCl (pH 9), 0.1% Triton X-100] and 10mM $MgCl_2$ (for the S2 fragment amplification) or 15mM $MgCl_2$ (for the S1 and S3 fragments)]

5 µl dNTPs [stock solution of 2mM dTTP/dGTP/dCTP/dATP]

68.5 µl filtered autoclaved distilled water

4 µl genomic DNA (100ng/µl working solution)

2.5 unit AmpliTaq DNA Polymerase

The reaction conditions were: 35 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 60°C (for S1 and S2), 63°C (for S3) and 1 minute extension at 72°C. An initial denaturation step of 5 minutes at 94°C and a final extension of 10 minutes at 72°C were employed.

2.3.4 Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis was used for both the microsatellite marker typing as well as for the sequencing analysis. 70ml of a 6% denaturing polyacrylamide gel was prepared and poured by capillary action between two large, glass, sequencing plates. The plates were separated by two vinyl spacers, which allowed a gel thickness of 0.4mm. Prior to assembly, both plates were washed thoroughly and one plate was

coated with dimethyldichlorosilane solution to prevent the polymerised gel from sticking to both plates. Immediately after pouring, a well former (of approximately 50 wells, each with a capacity of 4µl) was inserted into the top of the gel. The gel was then allowed to polymerise (about 1 hour), after which the well former was gently removed and the plates were assembled to a vertical sequencing electrophoresis tank. The gel was pre-run in 1 × TBE buffer at 70W for at least 30 minutes before samples were loaded. For microsatellite marker analysis, 3µl of formamide loading buffer were added to 3µl of PCR product. This was denatured for 5 minutes at 90°C and then 4µl were loaded onto the gel. For the sequencing analysis, 4µl of the final sequencing reaction mixture for each sample was heat denatured as above, and loaded onto the gel. Electrophoresis was carried out at 70W for 2-4 hours depending on the desired size separation (about 2 hours for 100-200bp range, and 3-4 hours for alleles ranging from 250-400bp). To determine microsatellite marker allele sizes, migration of the PCR fragments were compared with an M13 sequencing ladder. After electrophoresis the gel was vacuum dried at 80°C for 30 minutes before being autoradiographed by exposure to X-ray film overnight at room temperature.

2.3.5 Agarose gel electrophoresis

This was used for the analysis of the RFLP polymorphisms (**sections 2.3.12** and **2.3.13**) and for the gel purification of PCR products (**section 2.3.7**). An appropriate amount of powdered agarose was added to 1 × TAE buffer and dissolved by boiling. This was allowed to cool to about 50°C before adding ethidium bromide solution (to a final concentration of 1µg/ml) and then pouring into the appropriate gel former and comb arrangement. Once set, the gel was submerged in 1 × TAE buffer, the comb was removed and samples were loaded into the wells and run at a voltage of between 50-120V (depending on the size of the gel and the desired fragment separation). Separated fragments were visualised and photographed (using a Polaroid Land

camera) by placing the gel over a UV transilluminator. Agarose gel electrophoresis may be used for the separation of DNA fragments of between 100bp-10kb.

2.3.6 Mutation screening of *CYP11a* promoter region

Three overlapping PCR fragments were generated using primers designed from the published sequence of the 5' region of *CYP11a* (**Figure 2.3.6**), spanning 1.85kb of sequence upstream of the ATG start of translation site. The three fragments were designated S1, S2 and S3 (**Table 2.3.6**).

Table 2.3.6 Summary of information on the PCR fragments S1, S2 and S3.

PCR Fragment	Primer-set sequence	Size of product (bp)	Distance from ATG codon
S1	CTTCTGAGGGAGGAATGTGG* GGTTGCTTAGAGCTAGGGGG	700bp	-1174 to -1874
S2	TCCTCTAATTCCTCTATCCCCC* GAGAGTTTCATTCTTGTTGCCC	660bp	-551 to -1211
S3	GGTGAAACTGTGCCATTGC* CTGTACCTGCTCCACTTCAGC	581bp	-21 to -602

***- forward strand**

In order to obtain single-stranded DNA for sequencing in both the forward and reverse directions, two PCR reactions (one with only the forward strand primer biotinylated and the other with the reverse strand primer biotinylated) were carried out for each of the three fragments (**section 2.3.3**). Each PCR product was initially purified using the Promega Wizard PCR purification kit (**section 2.3.7**). Biotinylated single-stranded DNA was obtained using the Dynabeads M-280 Streptavidin system (**section 2.3.8**). Single-stranded product was then sequenced using the Sequenase version 2.0 DNA sequencing kit (**section 2.3.9**).

Figure 2.3.6 Sequence of 5' region of human *CYP11a*

```
-1838  ttcttccatg ccacatcgat tagggctcct tctgagggag gaatgtgggg ctgcgtagaa caatgggatt gactttaagt cagaaagtta taaatgtcac
-1738  ctcagtgctg agacccttgg aggaaaaact agtccttgga agactgcttt tcttgtggaa gctcatcacc ctgccgctgc tcgtgagaca ctgccttcct
-1638  tggctgatgt cattccaggc tcaaggtcat catggaggca aaacaggctt tctcatactc tctttatcag aaggttcatg actgatgagg tagtggtcac
-1538  tccagcggga agagcaacaa ccactcttga taagtacttt ttttttttct tctaaaaact gttgctctaa atttgttgaa agtggttcaa cagtattgga
-1438  gtctggggtc aagtggctgt gtaaggacaa cttttgccat tgtgggtgat ctatgggctg acacatacaa cagaagaggc caggaggatg tcactcgtgt
-1338  gtgtgtgtgt gtgtgtgtgt gtgttggtaa cagctctatt gagatataat tcacacacca tacattcacc catttaaagt atgcaattca atggctttta
-1238  gtgtatagag agttgttcaa taattaccac atatctttag aatattttca ttttcatcat cctaagaatc cctacacact ttatgtccca tcctctaatt
-1138  cctctatccc cctagctcta agcaaccacc agtctatttt ctgtctgtta gatgtgtgtt ttaaatgcct tatttgccag aaaataagat tttggggaaa
-1038  aaaagcaaga agtaccttcc aatgataagt ataaggtata gtgtagattg ctgtcatgaa atgctctaca tggcacatgt atgttttctg acctacacat
 -938  ttttaccaag tgggctgtaa ttttgttgaa atgacttgaa aactctacag gtgactgacc cttgttgttt gggaaagtgg taggtgcagg gtgatggggg
 -838  gtgggtgtgg gggctgacct gggctggaag gatgggtctg gggatatgat atgatgggag ggggtatgga agcagctctg aggggtttgc agcacagatc
 -738  aagtgtgtgg acagggaaga gctgacatcc tgagttccgg atggcaacca gatttgccaa ggtcttagag tgtgtccaga gtggagcctg accacagacc
 -638  tcagctcaag ggacccagag cccctctgag tcagctgtac tgaattacag ccccaaatct gggtcaactg gggagagacg acgaggatta gggttccaag
 -538  gtgaaactgt gccattgcgc tccagcctgg gcaacaagaa tgaaactctc ttaaaataaa ataaaataaa atagcctaag gatgcatttc tcagaactta
 -438  tccctgttgt tcaatgatgt gtgtctatac agtggggcca taactaagac gtatgttgcc caagctggca agatagctct gaccttctct tgggcccctc
 -338  atttccccca aacacaggtt gtctgcagtc ttgaccaatg gctgccaggg catggactcc gctgcagggg ccagtgggag gccccagctc aggcaaaagc
 -238  acaggcagat atttcaggag tctgctaggg ctggcactga gggcagagac agaggggtct ccctgtcctt tggagaacct cacgctgcag aaattccaga
 -138  ctgaaccttg ataccgagta ggggaggagc tgtctgcggg tttgagcctg cagcaggagg aaggacgtga acatttatc agcttctggt atggccttga
  -38  gctggtagtt ataatcttgg ccctggtggc ccagggctac agtcatccta gcagtccccg ctgaagtgga gcaggtacag tcacagctgt ggggacagca
  +62  atg
```

The arrowhead marks the site of transcription initiation and the "atg" codon at the start of translation has been underlined. The CAAT and TATA motifs are shown in bold font. There are five putative SF-1 sites and these have been boxed in red (at positions -46, -359, -543, -682 and -1618) (Morohashi et al. 1992, Takayama et al. 1994). The SP-1 like promoter element at position -118 has been boxed in green (Moore et al. 1990). There is a putative CRE element at position -1635 (Chung et al. 1989) and another CRE-like Ap-2 element at position -95 (Moore et al. 1992), both are boxed in turquoise. The two microsatellite repeat sequences present in this region have been underlined in blue.

2.3.7 Purification of PCR product

The Wizard PCR purification kit (Promega) was used to purify the S1, S2 and S3 fragments in preparation for sequencing. After PCR amplification, DNA fragments were separated by agarose gel electrophoresis (**section 2.3.5**) and the appropriate band excised (using a razor blade) from the gel under UV illumination. The gel slice was weighed and then placed in an Eppendorf tube. 1ml (per 0.3 g of gel) of DNA binding resin was added and this was incubated at 65°C for 5 minutes. A 2ml syringe (with the plunger removed) was attached to a Wizard minicolumn and the melted agarose/DNA/resin mixture was pushed through this using the plunger. The PCR product, which remains in the column, was washed with 2ml 80% isopropanol. The spin column was centrifuged for 20 seconds to remove all traces of isopropanol. The DNA was finally eluted from the column by adding 50µl of distilled water and centrifuging for 20 seconds. The DNA was now ready for strand separation (**section 2.3.8**) in preparation for sequencing (**section 2.3.9**).

2.3.8 Preparation of single-stranded DNA

This was used to obtain single-stranded DNA for sequencing. The S1, S2 and S3 fragments (**section 2.3.6**) were PCR amplified (**section 2.3.3**) in duplicate, with the forward primer biotinylated in one PCR reaction and the reverse primer biotinylated in the second reaction. This allowed both forward and reverse single-stranded DNA template to be obtained so that sequencing reactions could be carried out in both directions.

For each sample, 15µl of magnetic streptavidin coated beads were pipetted into an eppendorf tube and washed 3 x with TE/1M NaCl (to remove traces of sodium azide preservative). The 50µl gel-purified (**section 2.3.7**) PCR product was then added to the beads. The mixture was incubated at room temperature for 30 minutes to allow

the biotin-labelled PCR product to bind the streptavidin-coated beads. The tube containing this mixture were then placed in a magnetic rack (Dynal MPC-E) which collected the beads (together with the bound DNA) to one side and allowed the bound DNA to be washed with 30µl TE/NaCl. The tube was then removed from the rack and the DNA denatured by incubating with 20µl 0.15M NaOH at 28°C for 5 minutes. After denaturation, the single-stranded DNA was washed three times with 40µl water (using the magnetic rack), and finally resuspended in 10µl water ready for sequencing.

2.3.9 Sequencing analysis

This was used to screen the three S1, S2 and S3 fragments, from the *CYP11a* promoter region (**section 2.3.6**), for mutations. Single-stranded DNA (**section 2.3.8**), was sequenced by the dideoxy chain termination method using the Sequenase version 2.0 DNA sequencing kit (USB Corporation, USA). This method was also used to prepare M13 sequencing ladder, which was used as a sizing standard for microsatellite marker typing (**section 2.3.4**). Each of the fragments S1, S2 and S3 were sequenced in both directions using the nonbiotinylated forward or reverse primers used in the PCR amplification step.

The first step of the DNA sequencing reaction was to anneal the primer to the single-stranded template. This was set up in the following way (total volume10µl):

7µl	resuspended single-stranded template DNA (from **section 2.3.8**) (or 5µl (1µg) M13mp18 ss DNA provided by the kit)
2µl	5 × reaction buffer (200mMTris-HCl, 100mM $MgCl_2$, 250mM NaCl)
1µl	primer (25pmol) (or M13 universal sequencing primer provided)

This was incubated at 65°C for 2 minutes, and then slowly allowed to cool (over a period of 30 minutes) before placing on ice. While the annealing step was being carried out, four termination tubes (1.5ml Eppendorfs) were prepared for each DNA template to be sequenced. These were labelled "A", "T", "C" and "G" and to each

2.5µl of the corresponding termination mix (ddATP, ddTTP, ddCTP and ddGTP respectively) was added. These tubes were prewarmed to 37°C for about 2 minutes before the sequencing reaction was started. A 1:5 dilution of the labelling mix (7.5µM each of dTTP, dCTP and dGTP) and a 1:8 dilution of the Sequenase polymerase were prepared. The sequencing reaction was then carried out as follows:

10µl	annealed DNA mixture
2µl	diluted labelling mix
1µl	DTT (dithiothreitol, 0.1M)
0.5µl	[α-^{33}P] dATP
2µl	diluted Sequenase polymerase

The reaction mixture was incubated for 5 minutes at room temperature before transferring 3.5µl into each of the four labelled, prewarmed termination tubes and incubating at 37°C for a further 5 minutes. Finally, the reaction was stopped by adding 4µl of stop solution.

For each template sequenced, the four termination reactions were heated to 90°C before loading onto a 6% polyacrylamide sequencing gel (**section 2.3.4**). After gel electrophoresis, the gel was dried and autoradiographed.

2.3.10 Stanford G3 radiation hybrid panel screening

The Stanford G3 radiation hybrid panel (Research Genetics, Huntsville, AL, U.S.A.) is a medium resolution panel of 83 radiation hybrid (RH) clones of the whole human genome. It was created at the Stanford Human Genome Center by exposing a human lymphoblastoid cell line to 10,000 rad of x-rays, and subsequently fusing this with nonirradiated hamster recipient cells (Stewart *et al.*, 1997). Irradiation causes each chromosome to break into thousands of fragments whose size depends on the dose of irradiation. Irradiated cells are rescued by fusing with nonirradiated thymidine kinase-deficient hamster cells, which allows the selection of somatic cell hybrids.

Eighty-three independent somatic cell hybrid clones were isolated, in which human chromosomal fragments were inserted into the hamster chromosomes. In this manner, each RH clone retains a percentage of the human genome in fragments. The frequency of irradiation induced breakage between two markers is used as a measure of distance. The marker order may therefore be determined in a manner that in analogous to meiotic linkage mapping (Cox *et al.*, 1990). Distance is measured in centirays (cR), and for this RH panel $1cR_{10,000}$ ≈25Kb. Therefore, the G3 RH panel provides a mapping reagent for ordering sequence-tagged sites (STSs) and for determining the distance between them in the human genome. Over 10,500 STSs have been mapped, with an average retention frequency of 16%. These include random genomic DNA sequences, expressed sequences as well as previously mapped genetic (microsatellite) markers.

The *LHR* gene had previously been cytogenetically mapped to chromosome 2p21 (Rousseau-Merck *et al.*, 1990): however its precise genetic location relative to known polymorphic microsatellite markers had not been established. Therefore a PCR based sequence tagged site, together with the Stanford G3 radiation hybrid panel was used to map the *LHR* gene.

This was carried out by using a primer pair (CGATTTCACCTGCATGGC (1a) and CCCGACGTTTACAGCAGCC (1b)) which amplified a 212bp product from part of exon 11 of the *LHR* gene (Kremer *et al.*, 1995). This was optimised for mapping in the Stanford G3 Radiation Hybrid Panel. A hamster only and a human only control samples were also provided and were used as negative and positive controls respectively in the PCR optimising step. The PCR conditions were as described in **section 2.3.3**. Subsequently, the G3 RH panel was screened in duplicate using the *LHR* 1a/1b primer set.

The results of the G3 RH panel screening were scored as a vector, starting from hybrid 1 to hybrid 83, where "0" represents no positive, "1" represents a clearly positive score and "R" represents an ambiguous score. The raw scoring data was

submitted, together with information on expected chromosome assignment (i.e. chromosome 2), for analysis using the radiation hybrid mapping web server at the Stanford Human Genome Center (www-shgc.stanford.edu/Mapping/rh). A two-point linkage analysis was carried out using the RH2PT routine from RHMAP Version 2.0. The results of the Stanford G3 RH panel screening are reported as the SHGC framework marker which best links with the submitted test marker with a LOD score of 6 or greater (and/ or the best LOD score of 3 or greater if it matches the chromosomal assignment).

2.3.11 *CYP11a*(ttttta)$_n$ genotyping

The *CYP11a* (tttta)n polymorphism is present within the 581bp S3 PCR fragment (**section 2.3.6**). In order to reduce the size of the PCR product and therefore allow easy screening of this microsatellite, a new reverse primer (GTTTGGGGGAAATGAGGGGC), corresponded to the region -324 to -344 from the start of transcription, was designed from the published sequence (**figure 2.3.6**). This was used together with the S3 forward primer (**table 2.3.6**) to genotype all members of the case control study for the *CYP11a* (tttta)$_n$ microsatellite marker. Following PCR amplification (**section 2.3.3**), the radiolabeled samples were separation on a 6% denaturing polyacrylamide gel (**section 2.3.4**) and analysed after overnight exposure of the dried gel to X-ray film.

2.3.12 CYP17 assay

A PCR based RFLP assay had previously been designed by another member of the group, Dawn Waterworth. This allowed easy screening for the single base change polymorphism identified at position -34bp from the ATG codon in the *CYP17* promoter region (Carey *et al.*, 1994). Using the following primer set : CATTCGCACTCTGGAGTC (Forward) and AGGCTCTTGGGGTAACTTG (Reverse) a

459bp fragment corresponding to part of exon 1 and the adjacent 5' region of the gene was amplified using the standard PCR method (**section 2.3.3**) in a total volume of 25µl. The restriction enzyme *MspAI* was used to detect the presence of the base change (only cutting if the 'C' was present (designated the A2 allele) instead of the 'T' (designated the A1 allele)). If the A2 allele was present then the 459bp PCR product was cut into two fragments of 335bp and 124bp. This enzyme did not cut anywhere else in the PCR fragment. 3µl buffer (+BSA) and 0.2-0.5 units *MspAI* were added to the 17µl of the PCR reaction to a final volume of 30µl and incubated overnight at 37°C. This was separated on a 3% NuSieve™ agarose gel (**section 2.3.5**). Products were visualised by ethidium bromide staining.

2.3.13 *SRD5A1* RFLP assays

Two intragenic single base change RFLPs had previously been identified in the *SRD5A1* gene (Jenkins *et al.*, 1991). These were the *Hinf I* and the *Nsp I* RFLPs present in exons 1 and 2 respectively. Initially, the published PCR primers were ordered for both of these RFLPs. However, after a number of unsuccessful attempts at optimising the PCR reactions, both sets of PCR primers were re-designed based on the published sequence of the gene. Relatively long PCR primers were designed as this increases the specificity of the primers (by increasing the required PCR annealing temperature).

SRD5A1 NspI assay

The following primer set was designed from the published sequence of the gene:

CAAGAAAGTAAGATTTAAAACCCAAATCATTTAAGATAGG (Forward)

and AACCACTGGGACATCAATATAAAAGCAATGATGTGAACAAGGC (Reverse) a 337bp fragment (spanning the entire exon 2 region) was PCR amplified (**section 2.3.3**) in a total volume of 25µl. The restriction enzyme *Nsp I* was used to detect the presence of the base change (only cutting if the A is present (designated the "+" allele) and not cutting if the G is present (designated the "-" allele)). In the presence of the "+"

allele, the 337bp PCR product was cut into two fragments of 155bp and 182bp. This enzyme did not cut anywhere else in the PCR fragment. 3µl buffer and 0.2-0.5 units *Nsp I* were added to the 17µl of the PCR reaction to a final volume of 30µl and incubated overnight at 37°C. This was separated on a 3% NuSieve™ agarose gel. Products were visualised by ethidium bromide staining.

SRD5A1 Hinf I assay

New primers were re-designed as above:

TGCCCGCCGCGGCCTCTGGGGCATGGAGCACGCTGCCCAGCCCTG (Forward)

and GGCCCGCGCCGGCACTCGGAGCCTGTGGCTGGGCA (Reverse), this was used to amplify a 254bp PCR fragment (covering the whole of exon 1) (**section 2.3.3**) in a total volume of 25µl. The restriction enzyme *Hinf I* was used to detect the presence of the base change (only cutting if the G is present (designated the "+" allele) and not cutting if the C is present (designated the "-" allele)). In the presence of the "+" allele, the 254bp product was cut into two fragments of 138bp and 116bp. This enzyme did not cut anywhere else in the PCR fragment. The digest and agarose gel electrophoresis, were carried out as for the *Nsp I* assay.

2.4 COMPUTER PROGRAMS AND STATISTICAL ANALYSES

2.4.1 MLINK

All two-point linkage analyses were originally carried out using the FASTLINK (Cottingham *et al.*, 1993; Schaffer *et al.*, 1994) version of MLINK which are part of the LINKAGE package (Lathrop and Lalouel, 1984a; Lathrop *et al.*, 1984b; Lathrop *et al.*, 1986). These were performed on the computer system of the Human Genome Mapping Project (HGMP) Computing Resource Centre in Cambridge.

2.4.2 Model parameters

For parametric linkage analysis, information on the disease model parameters must be inputted into the linkage program. These include disease allele frequency, penetrance and phenocopy risk values. Based on some observable data, such as the disease population prevalence and the segregation model (e.g. dominant or recessive), and some rough estimations such as full or reduced penetrance and the rate of phenocopies (non genetic cases), allele frequencies and phenocopy risk values are obtained using the following formula:

$$Kp = f_1(p^2)+f_2(2pq)+f_3(q^2) \qquad (1)$$

where **Kp** is the population disease frequency; **p** is the disease allele (A) frequency, **q** is the wild type allele (a) frequency. $\mathbf{p^2}$ is the genotype frequency of individuals homozygous (AA) for the disease allele, $\mathbf{q^2}$ is the genotype frequency of individuals homozygous (aa) for the wild type allele and **2pq** is the genotype frequency for heterozygotes (Aa). f_1,f_2 and f_3 are the penetrance values for AA, Aa and aa respectively and are defined as the probability of being affected (i.e. expressing the phenotype), while having that particular genotype.

So for a dominant model, f_1 and f_2 will be the same and will be equal to the penetrance of the disease predisposing genotypes (AA and Aa) and f_3 is the

phenocopy risk (associated with the wild-type aa genotype). For a recessive model, f_2 and f_3 will be the equal to the phenocopy risk. For a dominant model, phenocopy rate is defined as the probability of having the aa genotype while being affected.

The allele and genotype frequencies are the standard Hardy-Weinberg equilibrium (HWE) terms based on the following formulae:

$$p^2+2pq+q^2 = 1$$

$$p+q = 1$$

Model parameters used for the parametric analyses

Although the most appropriate model for PCOS has not been determined, segregation ratio calculations in most family studies have suggested a dominant model with high penetrance (Legro, 1995). Recently, it has been suggested that LOD score calculations based on two simple single locus (SL) models, one dominant and one recessive, are more powerful than non-parametric approaches in detecting linkage in complex diseases (Durner *et al.*, 1999) (refer to **section 1.8.1**).

Therefore, the approach taken in this thesis has been to carry out non-parametric linkage analysis as well as two parametric analyses (one based on a dominant model and the other on a recessive model) for each locus being tested.

Model parameters were estimated using a disease frequency of 20% (based on the population prevalence of polycystic ovaries (Polson *et al.*, 1988)) and a phenocopy rate of 10% (this is a rough estimate which should allow for the possibility of genetic heterogeneity). It has been shown that variation in the penetrance value has a relatively modest effect on the magnitude of the maximum LOD score (Clerget-Darpoux *et al.*, 1986; Greenberg, 1989a; Greenberg *et al.*, 1998). Since most family studies of PCOS have suggested a dominant mode of inheritance with high

penetrance (see **section 1.5.1**), relatively high (but still arbitrary) penetrance values have been used for the two simple SL analyses.

For the dominant model, a penetrance of 95% has been used. Based on these model parameters, equation (1) has been used to estimate the disease allele frequency and the phenocopy risk to be 10% and 2.5% respectively.

For the recessive model, a penetrance of 72% has been used to derive an estimated allele frequency of 0.5% and a phenocopy risk of 2.8%.

2.4.3 GENEHUNTER

With the release of the GENEHUNTER program in 1996 (Kruglyak *et al.*, 1996), all further linkage analyses (two parametric (refer to **section 2.4.2** above) and a non-parametric analysis) were carried out using this program as it was considered to be more appropriate for our studies (see **section 1.8.1**, page 57). Therefore only the output from this program has been presented in this thesis.

Input parameters used for all GENEHUNTER analyses

Score all: all affected individuals are considered simultaneously.

Hap on: maximum-likelihood reconstruction of marker haplotypes will be carried out.

Skip large off: this ensures that large pedigrees are used but to make them manageable they are trimmed (unaffected individuals).

Disc off: this ensures that unaffected children are used in the analysis.

Count on: based on the maximum-likelihood reconstruction of marker haplotypes this will count the observed recombinations estimated for each map interval. The output from this function is used to carry out a z-test (**section 2.4.6**) to check the integrity of the results with respect to map inflations.

Off end 3: scans will be done 3cM beyond the ends of the map.

Inc steps 3: scans will be carried out in 3 equal steps per map interval.

Key to GENEHUNTER output results

Position: the centimorgan (cM) distance from the first marker.

LOD score^{D}: is the total parametric LOD score obtained, under the dominant model, at that position.

LOD score^{R}: is the total parametric LOD score obtained, under the recessive model, at that position.

HLOD: is the maximum LOD score obtained for the given alpha (proportion of families linked to this locus) when an allowance for heterogeneity is made.

NPL score: is the non-parametric linkage score which is a Z score.

p-value: is specific for the NPL score and is a measure of the pointwise significance of the NPL.

Info.: is a measure of the information content extracted at each point, where "1" is the maximum.

Interpretation of GENEHUNTER results

Lander and Kruglyak have proposed guidelines for the interpretation of linkage results in complex traits (Lander and Kruglyak, 1995). These standards are designed to be sufficiently rigorous to take into account the multiple testing aspect of linkage studies (particularly genome scans), while allowing for enough sensitivity to detect true hints of linkage. They have defined *suggestive linkage* as the statistical evidence that would be expected to occur one time at random in a genome scan. *Significant linkage* is defined as the statistical evidence expected to occur 0.05 times by chance in a genomescan (i.e. a 5% probability). The criteria that have been proposed are as follows: for a dense, complete genome scan a non-parametric p-value of between 0.001 and 0.0005 and a p-value of between 0.00005 and 0.00001 may be taken as evidence of suggestive and significant linkage respectively. For a parametric multipoint test, a LOD score of +1.9 may be considered as suggestive and a score of +3.3 as significant for evidence of linkage and a LOD score of -2 or less as evidence for exclusion. These criteria take into account the multiple testing aspect of a genome

scan. Since we studied only a small number of relevant candidate genes (which by definition have a reasonable prior probability of being involved in disease susceptibility), these criteria are likely to be overly stringent for our study.

2.4.4 Power limitations of parametric and non-parametric linkage analysis

There are two aspects to the power limitations of linkage analysis: one is the linkage method applied and the other is the type and size of the data set used in the analysis. The limitations of both of these factors are dependent on the model of the disease under investigation. Both parametric (e.g. two simple SL method of Greenberg et al. (Greenberg *et al.*, 1998)) and non-parametric (e.g. NPL statistic of GENEHUNTER (Kruglyak *et al.*, 1996)) linkage methods have been proposed as valid and appropriate approaches to the study of complex traits (see **section 1.8.1**). Simulation studies have been used to test the power of these approaches in detecting susceptibility loci that contribute to the aetiology of various models of complex disease (Risch and Merikangas, 1996; Kruglyak *et al.*, 1996; Greenberg *et al.*, 1998; McCarthy *et al.*, 1998; Durner *et al.*, 1999). Depending on the true genetic architecture of the disease (e.g. the number of susceptibility loci involved, their relative contribution to disease aetiology and the way they interact with one another (e.g. additively or multiplicatively)), a great range in the power of the analysis and the number of pedigrees required to detect linkage has been observed. For example, Durner *et al.* showed that for certain disease models (e.g. a two locus heterogeneity model where the locus being tested accounted for disease in a small proportion of the population), the power to detect linkage is low even under a parametric analysis using the correct model (Durner *et al.*, 1999). Also, Risch and Merikangas have shown (using the non-parametric ASP approach) that linkage analysis is only likely to be successful in detecting major genes (e.g. those with a genotype relative risk (GRR), defined as the increased chance that an individual with a particular genotype has the disease, of about 4) (Risch and Merikangas, 1996). They also showed that linkage analysis has

limited power to detect genes of modest effect (e.g. greater than 2,500 sibpairs may be required to detect a locus with a GRR of less than 2) (Risch and Merikangas, 1996).

Therefore, with the size and structure of our PCOS pedigrees, the power to detect linkage is likely to be limited to loci of relatively major effect. Hence, any exclusion data obtained will only indicate that the locus under investigation is unlikely to be a major susceptibility locus but it does not necessarily rule out the presence of a gene of modest effect.

2.4.5 Case control association studies

Candidate genes were assessed by comparing genotype (or allele) distribution differences between the cases and controls in this data set. Statistical evaluation of any differences observed was carried out using a contingency table and the χ^2 test. The odds ratio (OR), which is a measure of the relative risk of developing the disease given a certain genotype, was also estimated where applicable (i.e. where evidence for association had been obtained). This is calculated as the cross product ratio using a 2x2 contingency table and once again a χ^2 test is used to assess the significance. This data set was also used to assess any significant relationship between serum testosterone levels (a quantitative variable) and marker genotypes of candidate genes (eg. *CYP17* and *CYP11a*). Serum testosterone levels were initially $\log_e$ transformed to ensure a normal distribution, and then significance was assessed by using a non-parametric test such as the Mann Whitney or the Kruskal Wallis test.

Statistical analyses

We used a Statview spreadsheet as a database for storing patient information details as well as all biochemical data. All the statistical analyses carried out for the case control association studies were performed using the Statview package for Apple Macintosh (Version 4.0).

Limitations of case control association studies

Disease-marker association at the population level may arise for several different reasons (refer to **section 1.8.2**). Firstly, the disease associated allele may directly cause susceptibility to disease (i.e. the allele is the susceptibility locus). Secondly, it may be due to the presence of linkage disequilibrium (LD). LD may be an indication of allelic association between the marker allele and an allele at a tightly linked locus that influences susceptibility to disease. This may reflect the presence of a founder effect (i.e. most of the disease-bearing chromosomes are descended from one or a few ancestral chromosomes). Alternatively, LD can arise either as a result of random genetic drift, selection or genetic admixture. Random genetic drift is the random changes in allele or haplotype frequencies from one generation to the next in a finite population. Selection may arise, for example, when the effect of an allele at one locus, that has an influence on reproductive fitness, is potentiated by the presence of another allele at a different locus, resulting in a disproportionate increase in the frequency of individuals with both alleles. Population admixture is the recent mating between two ethnically distinct populations giving rise to individuals that are a genetic mixture of the two populations. Spurious associations may be obtained as a result of population stratification, for example, such as in the case of genetic admixture (particularly in the earliest generations where LD extends over the whole genome) or as a result of inadequately matched controls. Alternatively, erroneous results may be obtained by chance due to small sample size (i.e. a type I error).

There are two main considerations when carrying out an association study. The first is to minimise the chances of obtaining false positive associations, which, for a case control study, most often, relates to the use of an inappropriate control group. Ethnic differences between cases and non-related controls can lead to spurious apparent association between a candidate gene and disease phenotype. This can occur if there are ethnic differences in both candidate gene allele frequencies and disease incidence.

For these reasons we have only included Europid individuals in both our patient and control groups in an attempt to minimise this effect.

The second consideration is a question of the size of the data set required (i.e. the power) to detect a statistically significant allelic association. This is dependent on a number of factors such as the quantitative contribution of the locus to disease susceptibility, the mode of transmission and the degree of LD between the susceptibility and the marker loci (Cox and Bell, 1989). Since none of these parameters is usually known with certainty, it is virtually impossible to rationally specify sample sizes. Attempts by Cox and Bell (1989) at estimating the sample size required to detect NIDDM susceptibility loci based on different assumptions revealed that this may vary by several orders of magnitude depending on the model tested. Essentially, apart from the influence of mode of inheritance, penetrance and phenocopy rate (genetic heterogeneity), the strength of LD between the marker and susceptibility locus as well as the frequency of the disease associated marker allele appeared to have a highly significant effect.

Therefore, in general, case control association results should be interpreted with some care. Alternative explanations for positive data need to be considered and negative results (i.e. failure to detect association) may simply reflect a lack of power of the data set to detect the contribution of that locus to disease susceptibility.

2.4.6 z-test

Using the "count recs" output from GENEHUNTER (refer to **section 2.4.3**), a z-test may be used as a way of checking the integrity of the linkage results with respect to map inflations (Riley, 1996). The "count recs" function counts the observed recombinations in each interval, and based on the user defined map distance, defines the number of recombinations expected for that interval. The z-test is a test for the comparison of two proportions (using the continuity correction for normal

approximation to the binomial distribution) and assesses the significance of the difference observed. The z-score is derived using the following formula:

$$z = \frac{|p - \pi| - 1/(2n)}{\sqrt{\{\pi(1-\pi)/n\}}}$$

where **n** is the total number of meioses, π is the expected recombination fraction (θ), and **p** is the observed recombination fraction (= observed no. of recombination /n).

A one-tailed standard normal distribution table is then used to obtain a p-value for the z-score calculated. A significant p-value indicates either genotyping or data entry errors have been made or that an incorrect map distance has been used.

2.4.7 M-test

Several methods have been proposed for the detection of locus heterogeneity. The method proposed by Smith (Smith, 1959) assumes the presence of two disease loci that are unlinked to each other. This so-called *admixture model* assumes that the marker being tested is possibly linked to the disease locus in a proportion of the families (defined by α), but is definitely unlinked in the rest. The HLOD score (heterogeneity LOD score) estimated is the maximised LOD score for the subset α. Therefore to take this extra parameter α into account, and to test whether there is significant evidence for heterogeneity, an *M-test* is carried out. This compares the likelihood of linkage under homogeneity ($\alpha = 1$) with the likelihood of linkage under the assumption of heterogeneity ($0 \leq \alpha \leq 1$). The M-test, or likelihood ratio test statistic, is derived using:

$$2 (\ln L_2 - \ln L_1)$$

where **ln L_1** is the maximum log-likelihood of the entire data set over $0 \leq \theta \leq 0.5$; **ln L_2** is the maximum log-likelihood under the assumption of heterogeneity (i.e. α-specific). This statistic has a chi-squared distribution (one-tailed) and has one degree

of freedom. A heterogeneity test is considered an appropriate test when there are reasons for dividing the data into subsets (eg. some prior evidence for linkage).

2.4.8 Bonferroni correction

When a large number of independent significance tests are carried out (each with the significance level set at 5%), then some of the tests will be significant due to chance (even in the absence of any real effect). The Bonferroni correction is a simple method of controlling for the Type I error rate associated with such multiple comparisons (Bland and Altman, 1995). This can be done in one of two ways: either by using:

$$\alpha = \frac{0.05}{k}$$

where, **k** is the number of independent significance tests and α is the level set at which all the null hypotheses are true. If any of the p-values are less than α then a significant difference at the 5% level will have been achieved. Alternatively the p-value obtained for each test may be multiplied by the number of independent significance tests, k, if the corrected p-value ($\mathbf{P_c}$) is less than 0.05 then a significant result has been obtained at the 5% level (any P_c exceeding one is ignored).

The Bonferroni correction has only been used for the correction of association results in this thesis. For linkage results, the accepted significance levels have already been set at a sufficiently stringent level to take into account any multiple testing (refer to **section 2.4.3**).

2.4.9 Measure of linkage disequilibrium

Linkage disequilibrium (LD), or allelic association, is defined as the co-inheritance in a population, of a pair of alleles at two linked polymorphic loci at a frequency higher (or lower) than would be expected by chance. If we consider two tightly linked loci

A and B, each with two alleles (A, a and B, b respectively), then there are four possible combinations of alleles that could exist on the same chromosome (A_B, A_b, a_B and a_b). If allele A has a frequency of P_A and allele B has frequency P_B, then in the absence of LD (i.e. the alleles occur independently on haplotypes) the frequency of A_B haplotypes would be P_AP_B (known as the equilibrium haplotype frequency). If alleles A and B are in LD, the observed A_B haplotype frequency would differ from its equilibrium value (P_AP_B). The difference between the observed haplotype frequency and its equilibrium value is a measure of the magnitude of LD between the alleles.

LD between two loci can be generated in finite populations by founder effects, random drift, mutation, selection and non-random mating (see **section 2.4.5**, page 99).

A commonly used measure of LD is the correlation coefficient Δ(Hill and Weir, 1994). The Δ for two linked loci can be derived using the following formula:

$$\Delta = \frac{\text{(freq. of coupling haplotypes) - (freq. of repulsion haplotypes)}}{\sqrt{\{\text{product of allele frequencies of both loci}\}}}$$

Δ has a maximum value of +1 and a minimum of -1, where a positive Δ indicates coupling (i.e. association between common alleles of the two loci), a negative value indicates association between the common allele at one locus and the rare allele at the other (repulsion) and $\Delta = 0$ indicates that there is no LD between the two loci. The significance of the Δ value obtained can be tested by taking $N\Delta^2$ (where N is the total number of haplotypes observed), which is distributed as χ^2 with 1 degree of freedom.

3.0 RESULTS

3.1 CHROMOSOME 15q

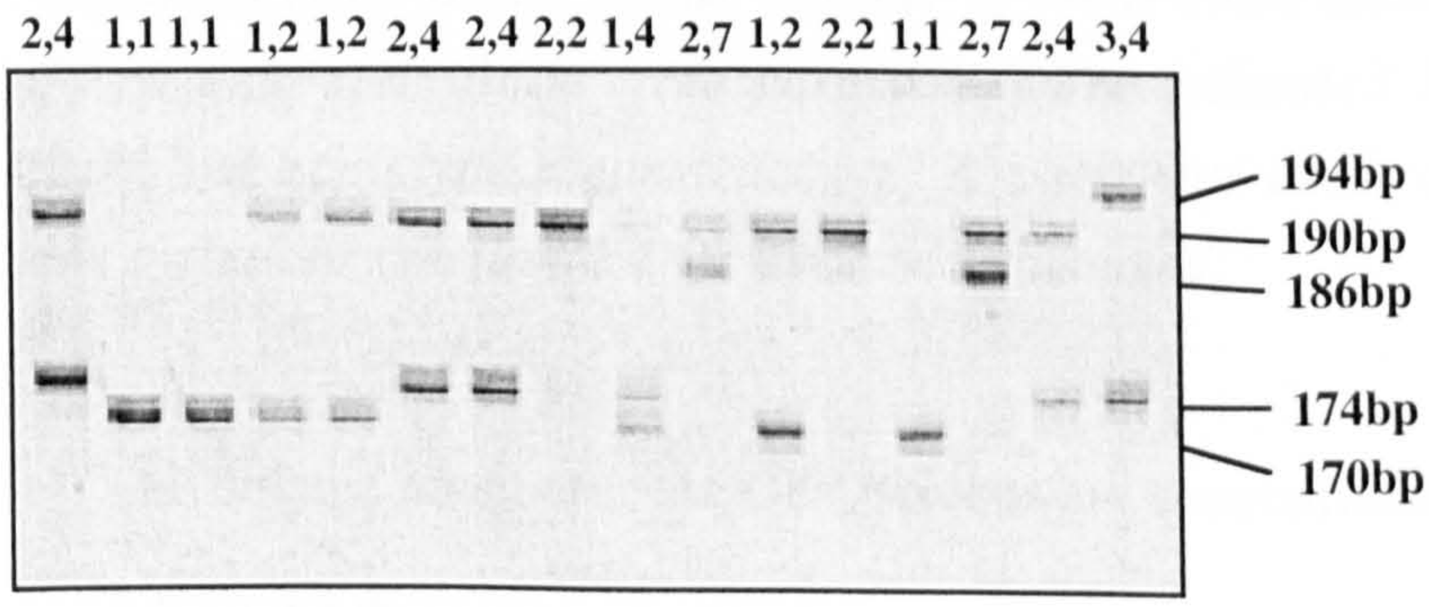
The genes coding for the steroidogenic enzymes P450 side chain cleavage (*CYP11a*) and P450 aromatase (*CYP19*) are present on the long arm of chromosome 15, separated by a genetic distance of less than 35cM, at positions 15q24 and 15q21.1 respectively. In the following section, each of these is investigated in turn as a candidate for involvement in the aetiology of PCOS. The preliminary results of this study were published in 1997 (Gharani *et al.*, 1997). These results have since been updated and are presented below.

Parametric and non-parametric linkage analyses were carried out for both the *CYP19* and the *CYP11a* loci using the GENEHUNTER program (**section 2.4.3**) (Kruglyak *et al.*, 1996). For each of these genes an intragenic microsatellite marker (*CYP19*(ttta)n and *CYP11a*(ac)n) was used in the linkage analysis. The marker information, position and chromosomal order of the markers used in the analyses are given in **section 2.2.1**.

3.1.1 *CYP19* locus linkage results

Examples of *CYP19*(ttta)n genotypes, the intragenic marker used in the linkage analysis, are illustrated below in **figure 3.1.1**.

Figure 3.1.1 Autoradiograph of*CYP19*(ttta)n genotypes.



The results of the *CYP19* locus linkage analysis are as follows:

- Information on the expected and observed number of recombinations for each marker locus interval is provided by the GENEHUNTER output and is presented below in **table 3.1.1**. A simple z-test (**section 2.4.6**) may then be used to test the integrity of the data with respect to map inflations. If map inflations are detected, this would indicate, either that there are genotyping or data entry errors, or that the map distance used in the analysis is wrong.

- No significant map inflations were detected for any of the marker intervals (**table 3.1.1**).

Table 3.1.1 Recombination data for the *CYP19* locus linkage analysis.

Locus Interval	Recombination Fraction (θ)	Expected cross-overs	Observed cross-overs	z-test	p-value
D15S118- D15S123	0.099	17.31	18.73	-0.078	0.468
D15S123 - CYP19(ttta)$_n$	0.017	3.19	5.26	0.847	0.198
CYP19(ttta)$_n$- D15S103	0.000	0.00	0.00		
D15S103- D15S126	0.036	6.63	5.42	0.371	0.356

Information on the number of expected and observed recombinations, for the CYP19 locus linkage analysis, was generated by the GENEHUNTER program. There were a total of 191 meioses observed. Recombination fractions (θ) are based on the interval distance values used in the analysis. The "Expected cross-overs" is the number calculated based on the θand the number of observed meioses. The "Observed cross-overs" is the number of cross-overs estimated during the analysis of this data set. Both these parameters are estimated based on the maximum-likelihood haplotype reconstruction. A z-test was used to test for the presence of map inflations (section 2.4.6): none were detected.

- The linkage results for the *CYP19* locus are summarised in **table 3.1.2** and **figures 3.1.2 - 3.1.4.**

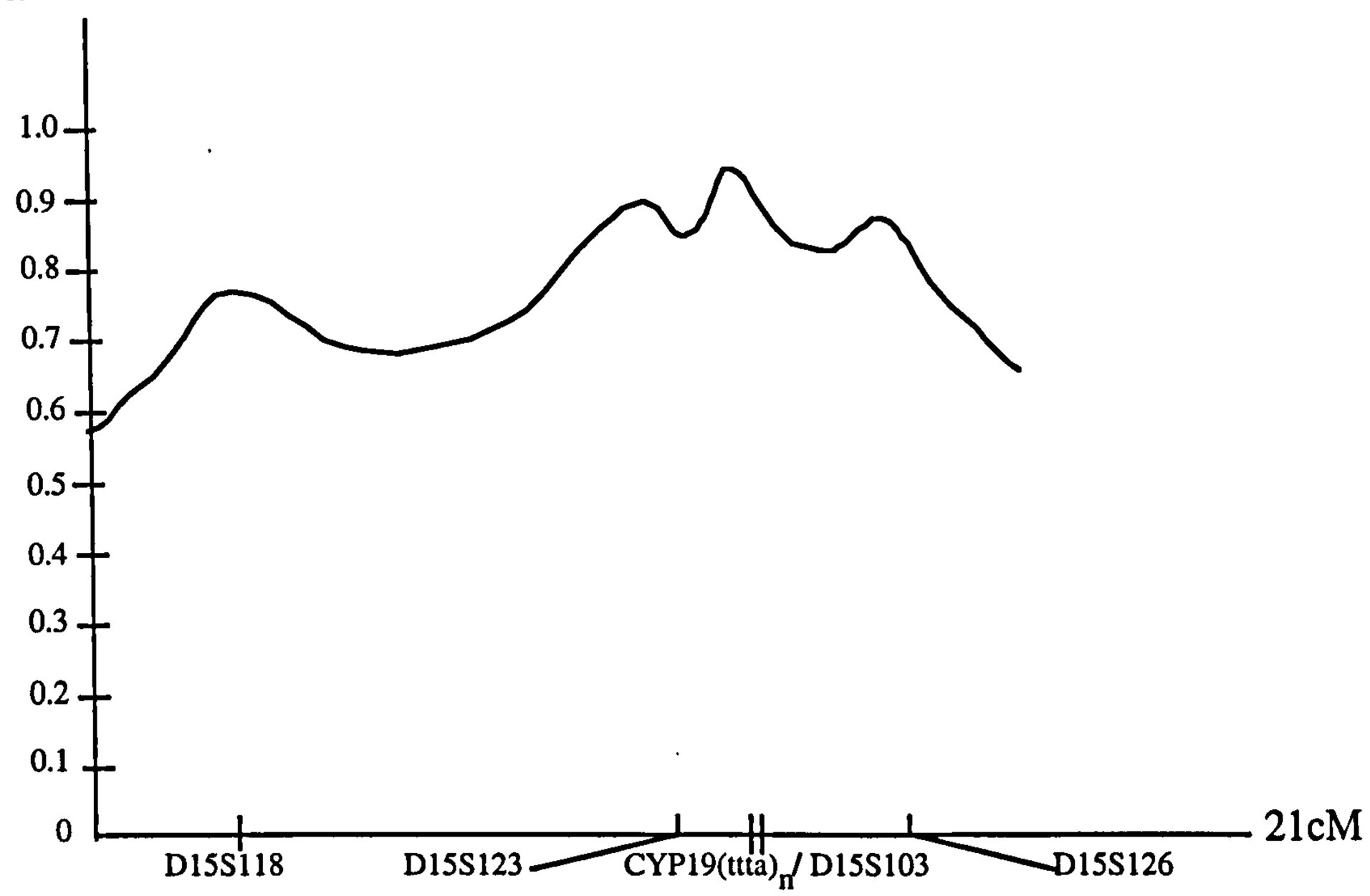
Table 3.1.2 GENEHUNTER results for *CYP19* locus

Marker	Position (cM)	LOD score^R	LOD score^D	NPL score	p-value	info.
	-3.00	-1.72	-3.27	0.28	0.37	0.58
	-2.00	-1.84	-3.59	0.28	0.37	0.63
	-1.00	-1.96	-3.95	0.29	0.37	0.69
D15S118	0.00	-2.11	-4.39	0.29	0.36	0.77
	3.33	-2.18	-4.41	0.20	0.40	0.70
	6.67	-2.37	-5.08	0.11	0.44	0.75
D15S123	10.00	-2.71	-6.69	0.03	0.47	0.92
	10.57	-2.57	-5.78	0.03	0.47	0.90
	11.13	-2.47	-5.86	0.03	0.47	0.91
CYP19$(ttta)_n$ / D15S103	11.70	-2.41	-7.00	0.03	0.47	0.95
	12.90	-2.44	-5.77	0.27	0.37	0.89
	14.10	-2.50	-5.18	0.51	0.29	0.87
D15S126	15.30	-2.61	-4.88	0.75	0.21	0.89
	16.30	-2.44	-4.23	0.73	0.21	0.80
	17.30	-2.29	-3.70	0.71	0.22	0.73
	18.30	-2.14	-3.26	0.68	0.23	0.67

Refer to section 2.4.3 for GENEHUNTER input parameters used in the analysis and for the key to the output results. The "p-value" above, is specific for the non-parametric linkage score ("NPL score") and is a measure of the pointwise significance of the NPL. "Info." is a measure of the information content extracted at each point, where "1" is the maximum. For both the dominant "LOD score^D" and the recessive "LOD score^R" analyses, parametric LOD scores of less than -2 have been obtained across the whole region (except for the 3cM region beyond marker D15S118 at the end of the map, where the LOD score^R does not reach criteria for proof or rejection of linkage). Non-parametric analysis has provided no evidence for excess allele sharing. There is zero genetic distance between the D15S103 and CYP19$(ttta)_n$ markers. The position of the markers along the chromosome are given in centimorgans (cM) moving from the most centromeric marker to the most telomeric.

- Information content (extent of information extracted by the marker data at each point along the linear map) is generally high at all markers (**table 3.1.2** and **figure 3.1.2**). The highest value was obtained at the *CYP19(ttta)$_n$/D15S103* locus (0.95) and the minimum value of 0.77 was obtained at marker D15S118.

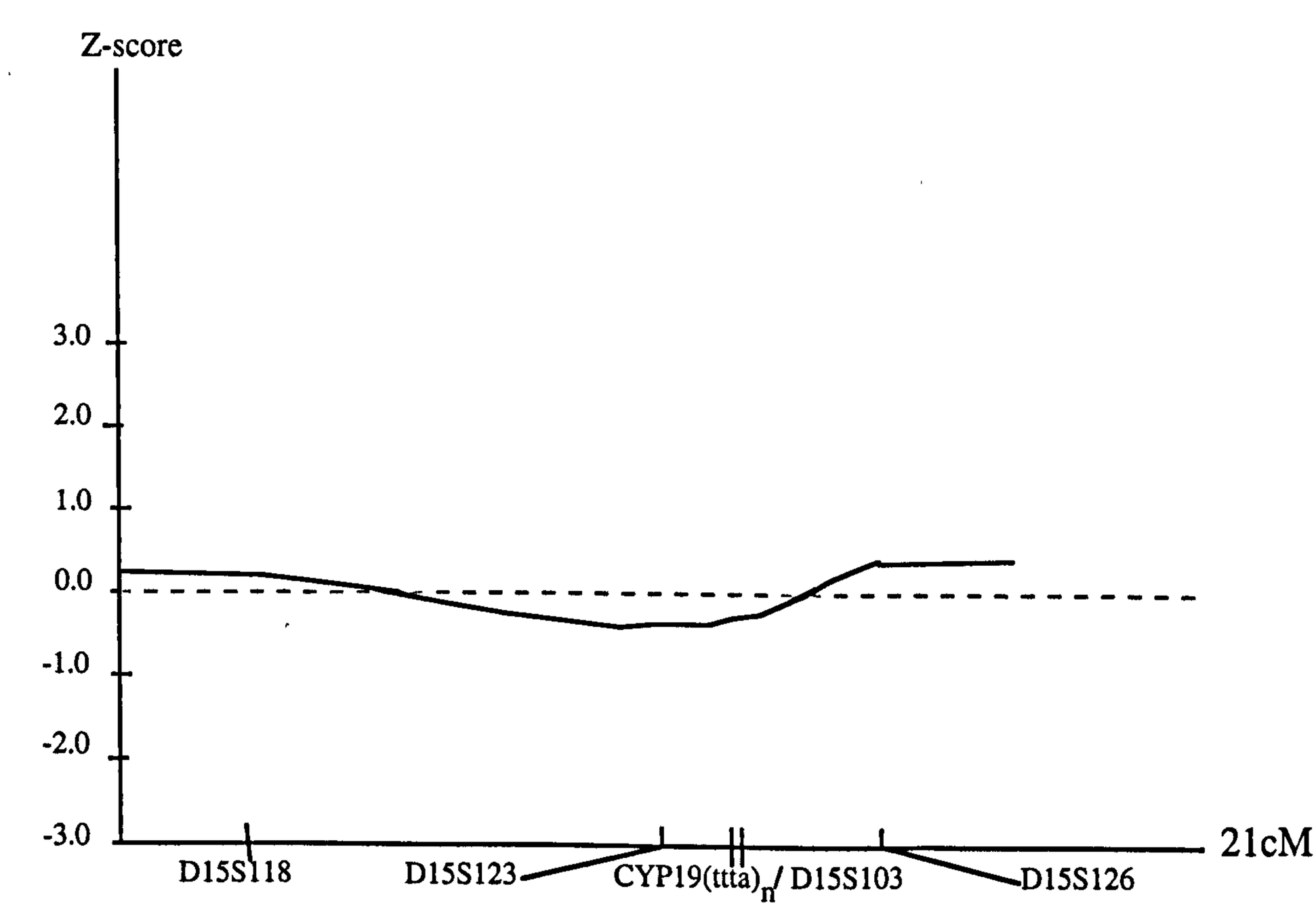
Figure 3.1.2 Information plot for the *CYP19* locus linkage analysis



Information plot is a graphic representation of the information extracted at each point along the linear map (see table 3.1.2). Information content is generally high at all markers.

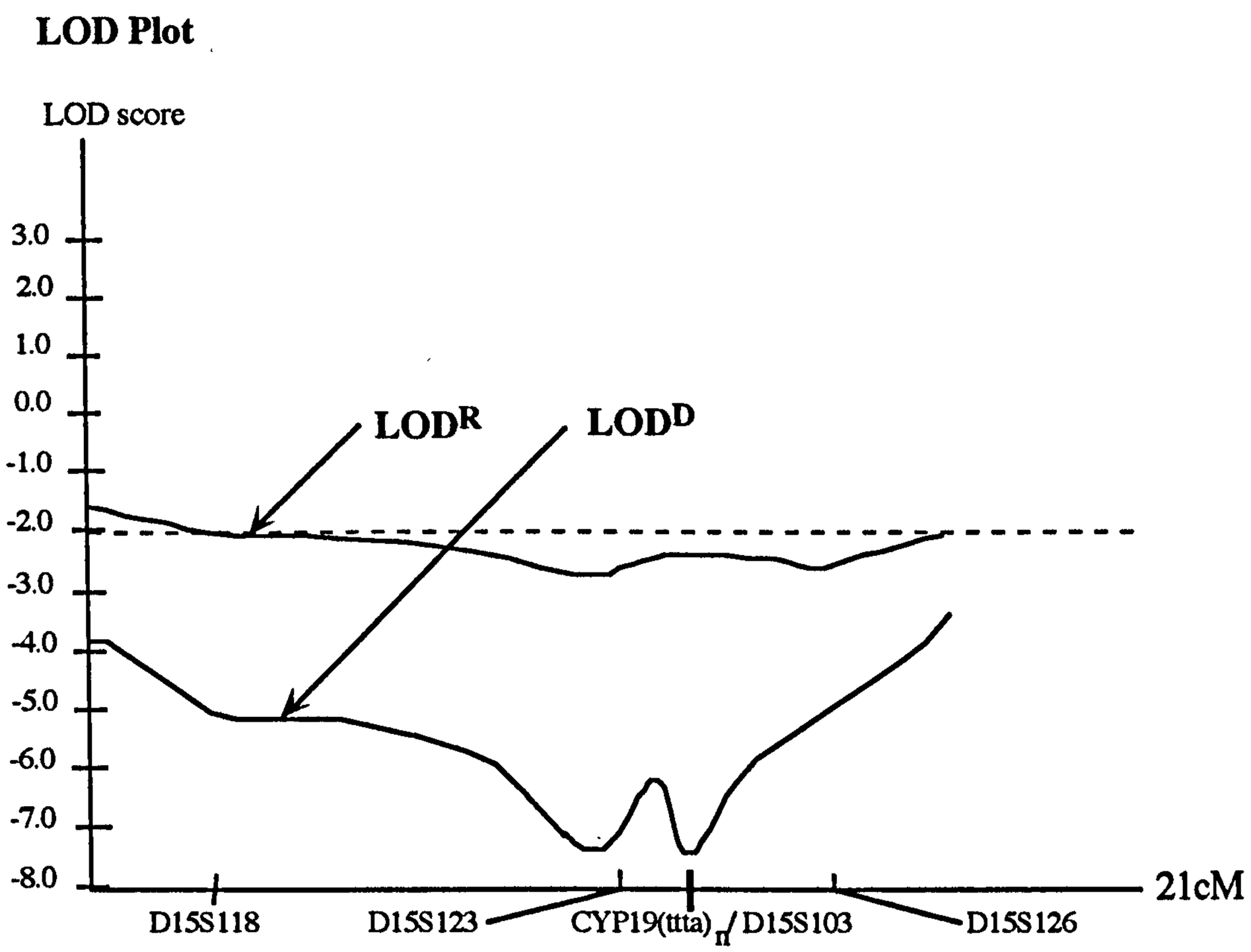
- Clear exclusion of the *CYP19* locus as a major susceptibility locus (refer to **section 2.4.4** for limitations of linkage analysis) was obtained (**table 3.1.2** and **figures 3.1.3** and **3.1.4**). There was no evidence for excess allele sharing and a negative parametric LOD score of -2 or less was obtained across the whole region for the dominant parametric analysis. For the recessive parametric analysis, most of this region was excluded (including the *CYP19(ttta)n/D15S103* locus) with a LOD score of -2 or less. Only the 3cM region beyond marker *D15S118* at the end of the map could not be excluded and remained inconclusive. This is probably due to the diminishing information content over that region.

Figure 3.1.3 NPL plot for the *CYP19* locus linkage analysis.



NPL plot is a graphic representation of the non-parametric linkage results (refer to table 3.1.2). The dashed line represents the mean of expected allele sharing distribution. There is no evidence for excess allele sharing across the entire map.

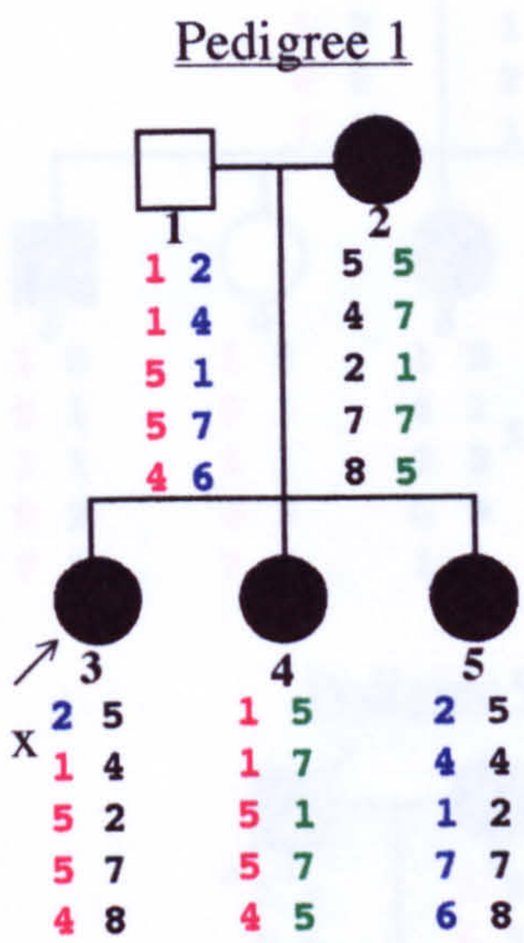
Figure 3.1.4 LOD plots for the *CYP19* locus linkage analysis.

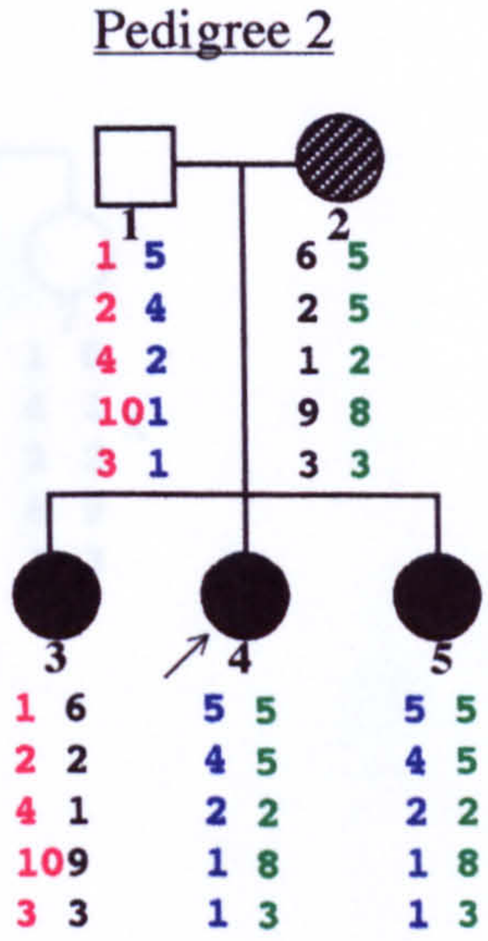


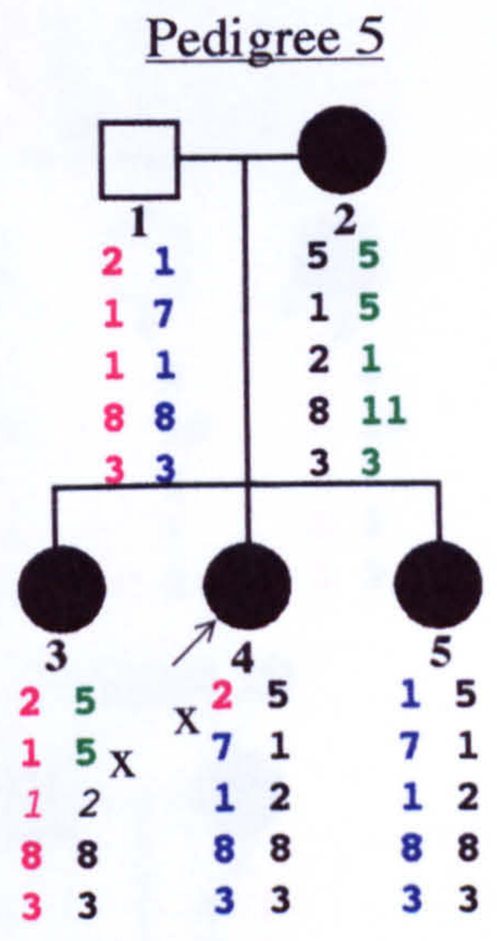
The above LOD plots are graphic presentations of the parametric linkage results (see table 3.1.2). The black and red curves represent the parametric linkage analysis under the dominant and recessive models respectively. The dashed line at -2 represents the minimum criterion required for rejection of linkage (section 2.4.3). Both the dominant and recessive analyses have excluded the presence of a major susceptibility locus from the entire region.

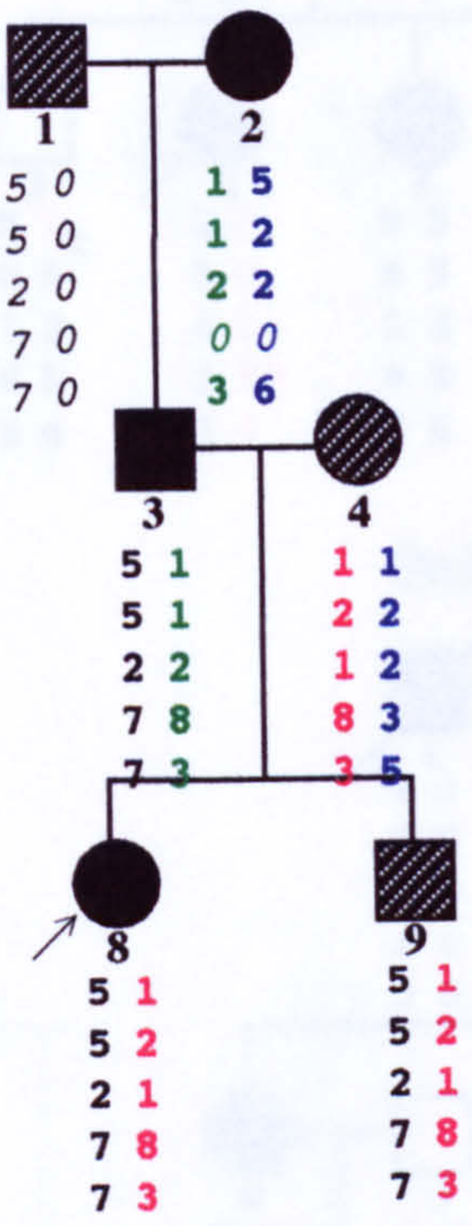
- The most likely reconstruction of the pedigree haplotypes (generated by the GENEHUNTER program) are given in **figure 3.1.5.**

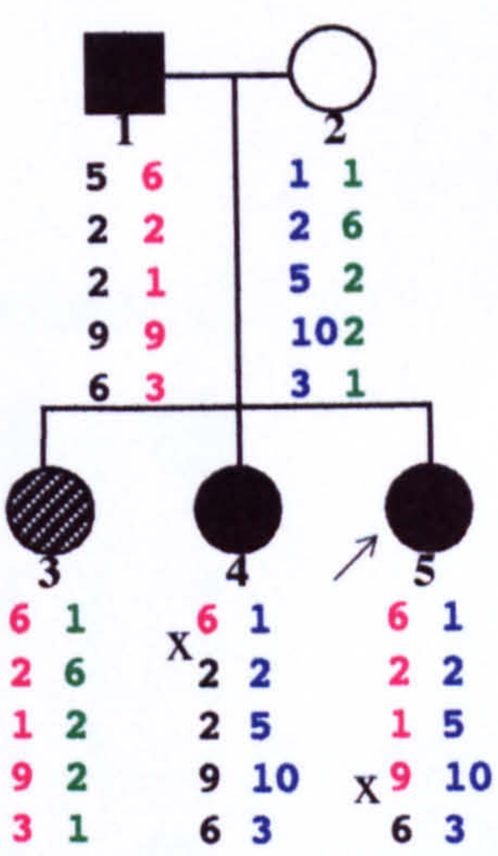
Figure 3.1.5 *CYP19* locus haplotypes in the PCOS/MPB pedigrees.

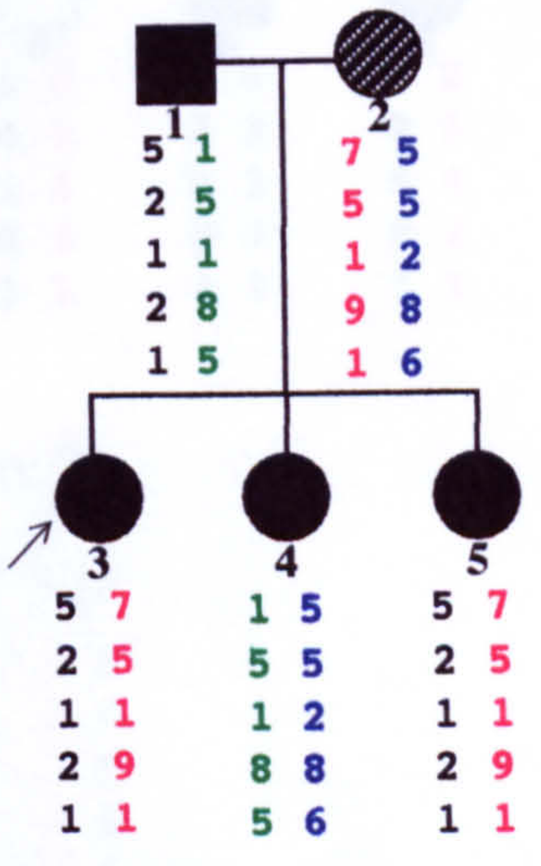






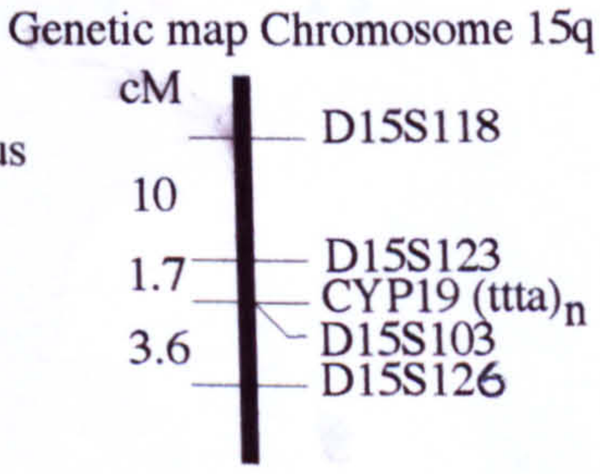




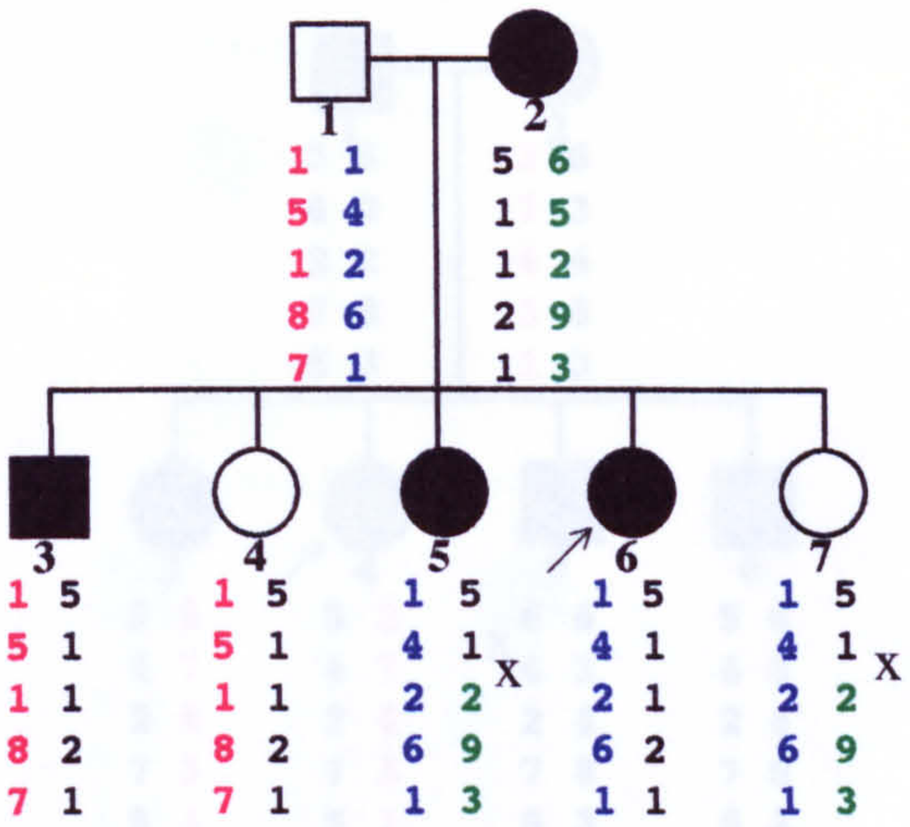


Key

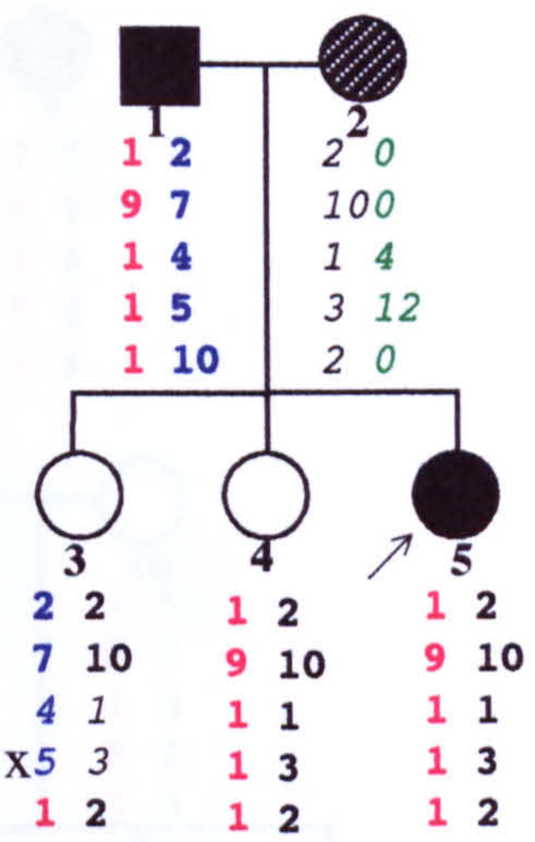
- unaffected male
- affected male
- unknown affection status
- unaffected female
- affected female
- unknown affection status
- X crossover
- *'Italic'* assumed genotype



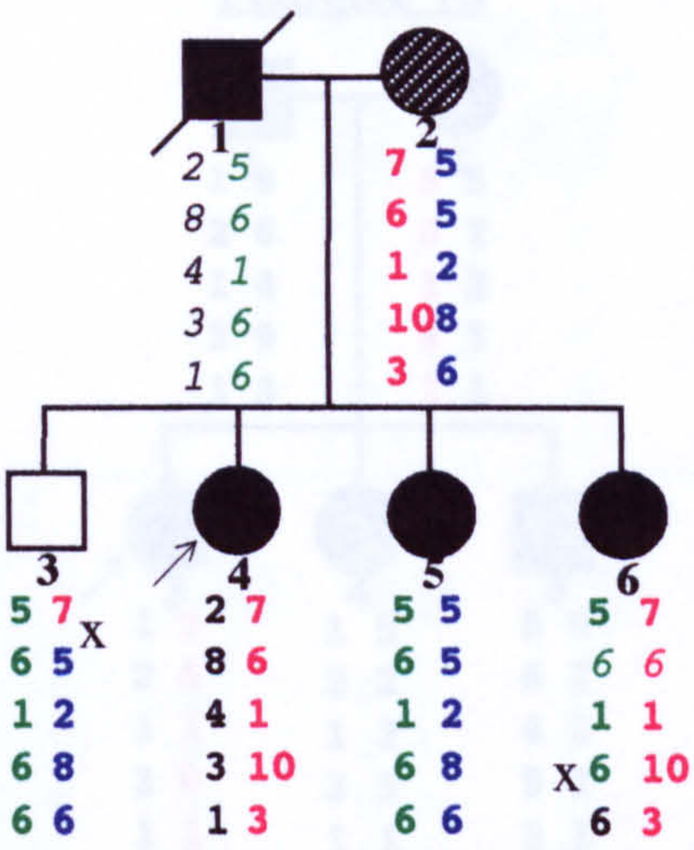
Pedigree 7



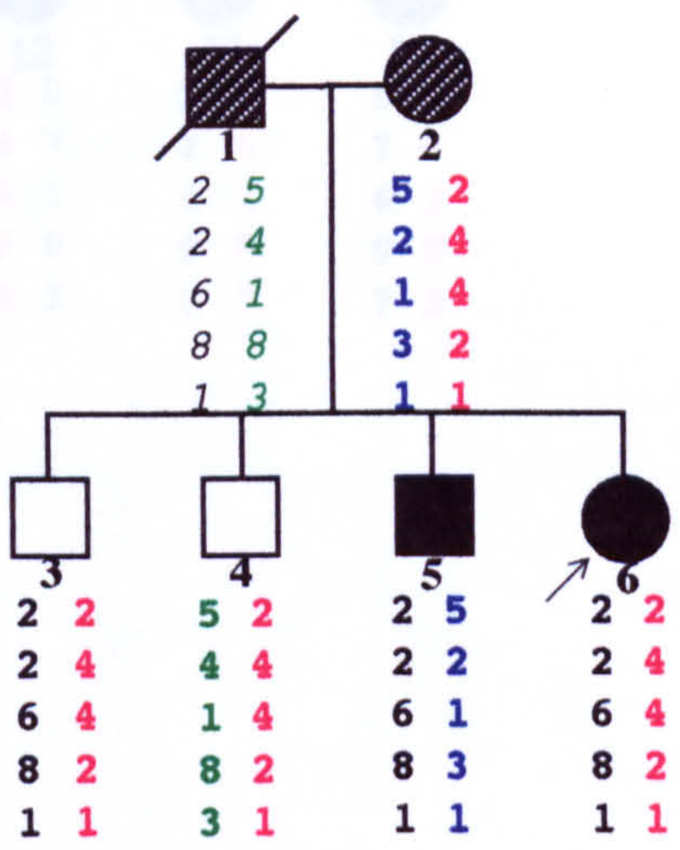
Pedigree 8



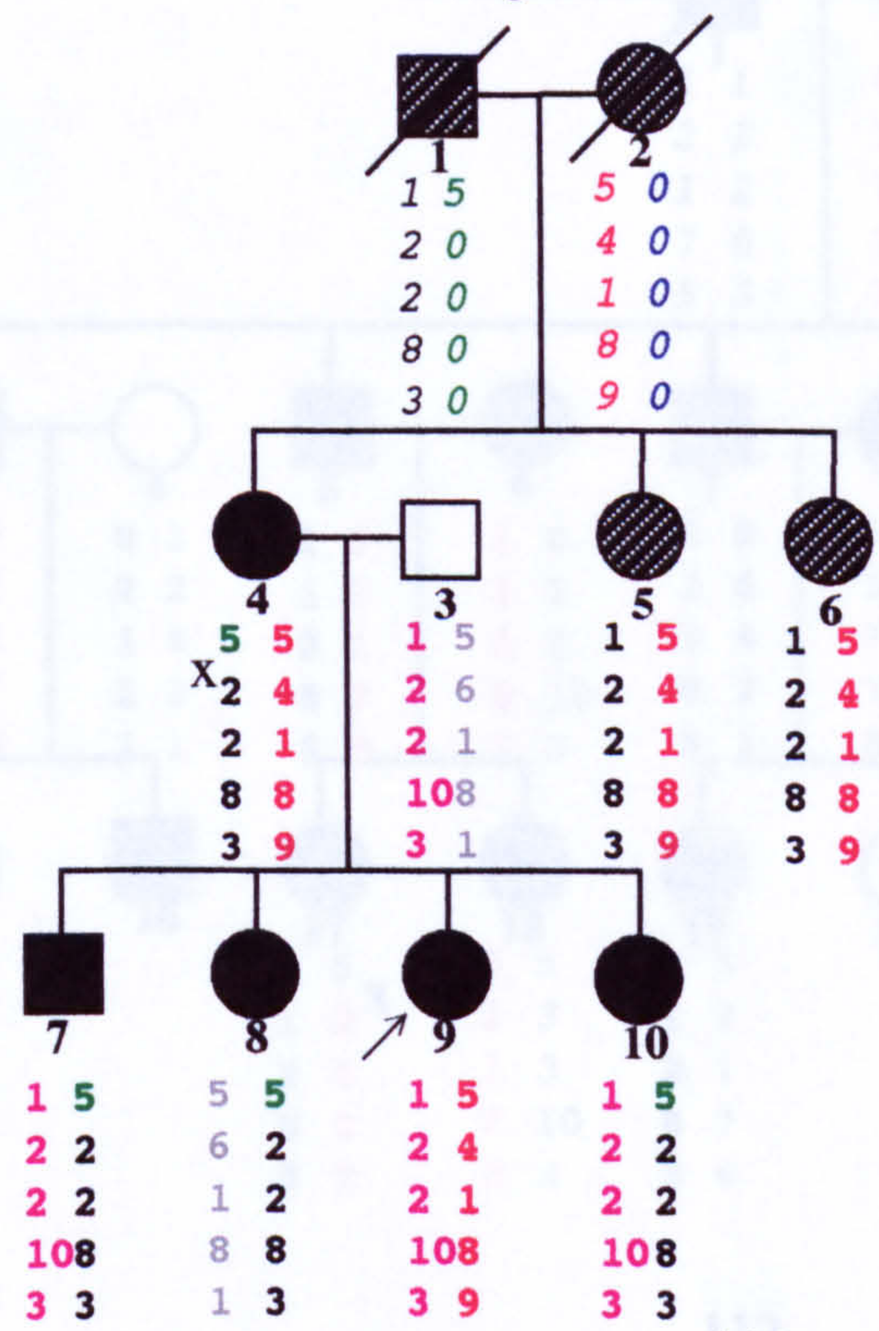
Pedigree 9



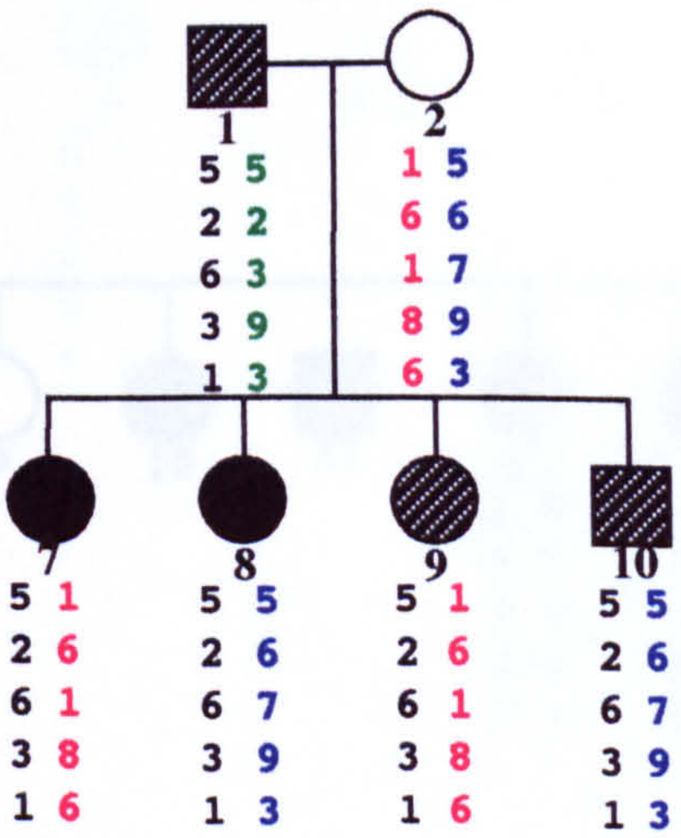
Pedigree 10



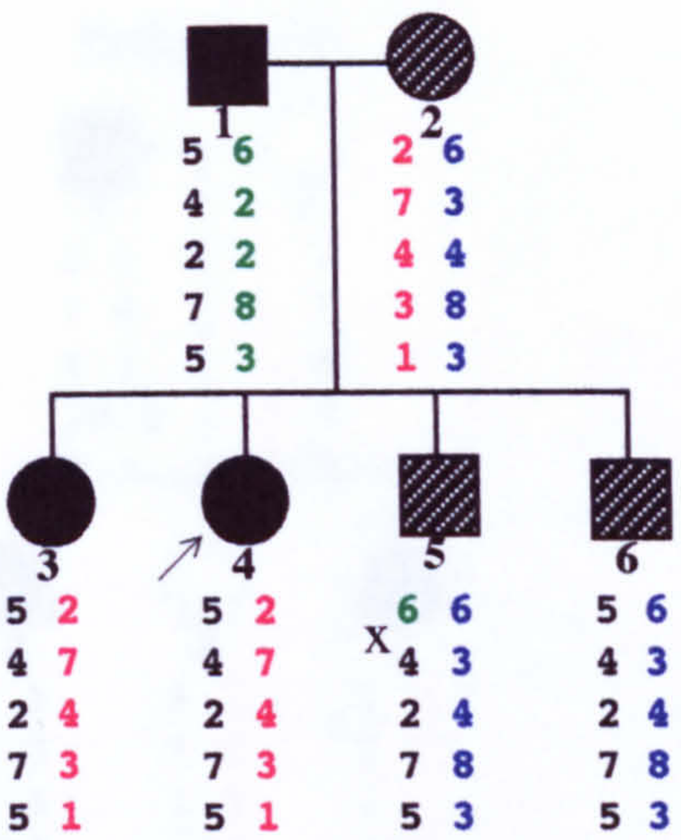
Pedigree 11



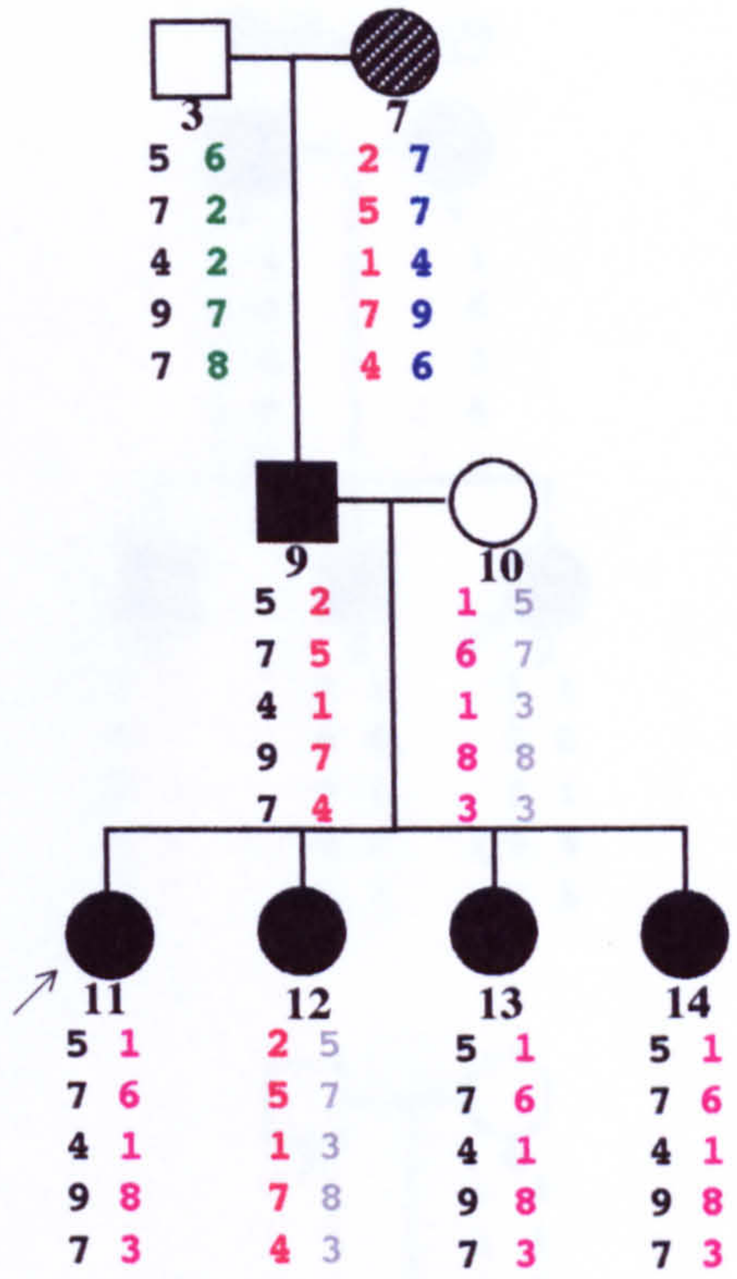
Pedigree 12



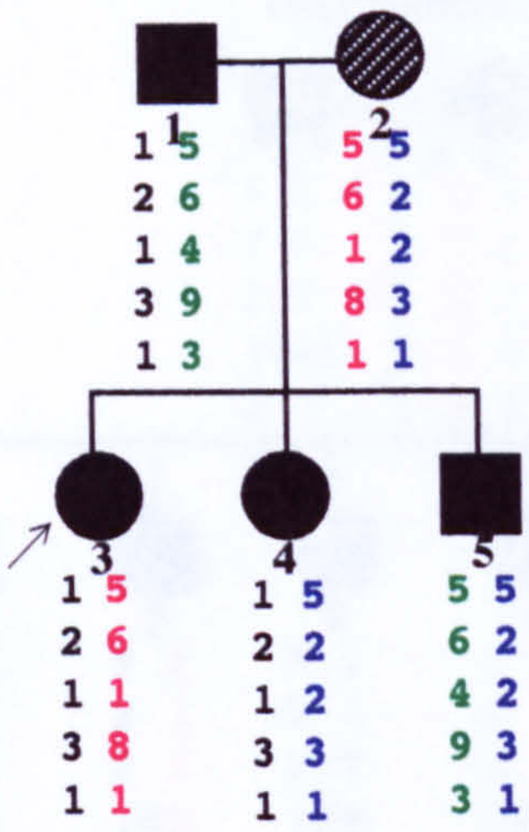
Pedigree 13



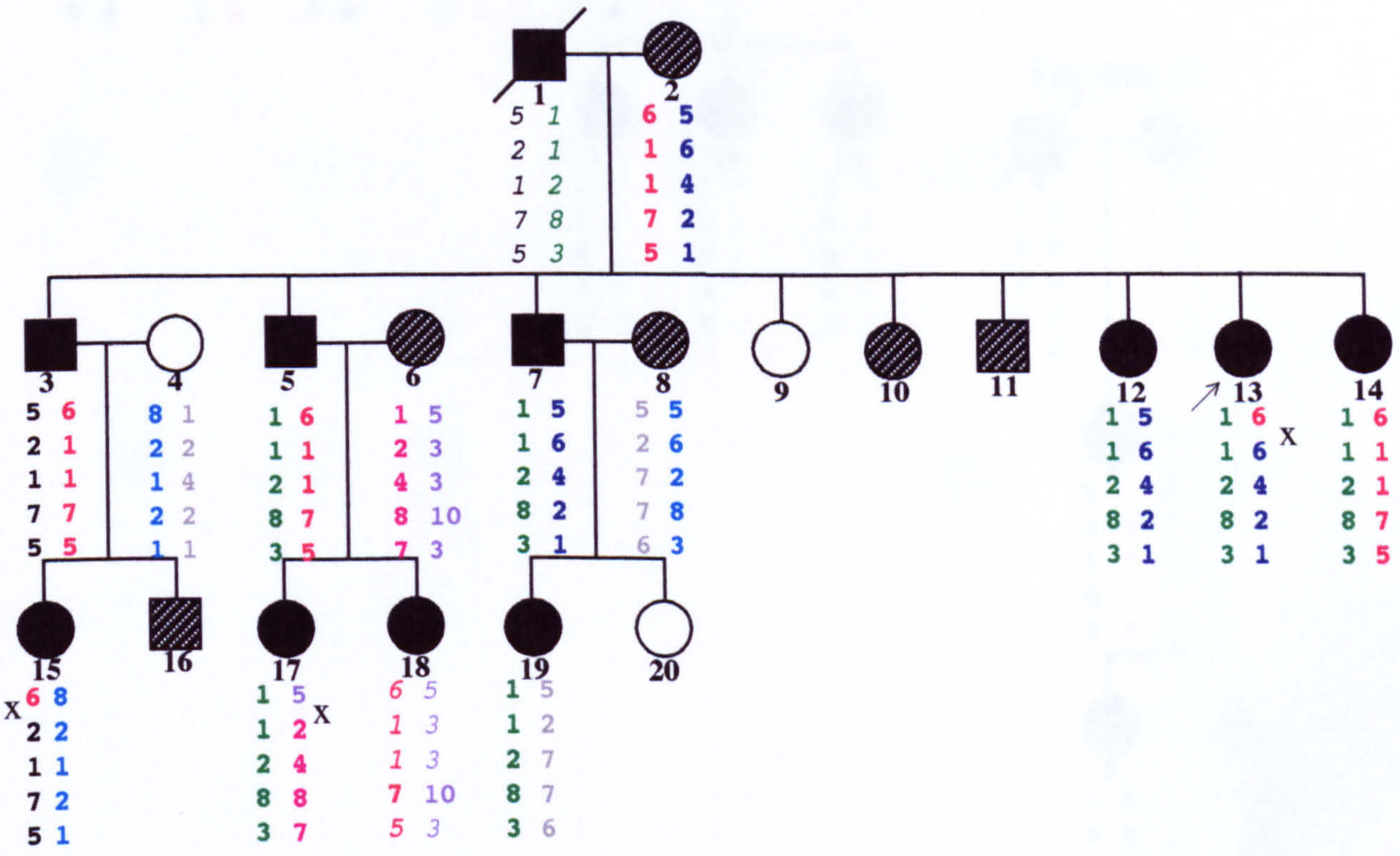
Pedigree 14



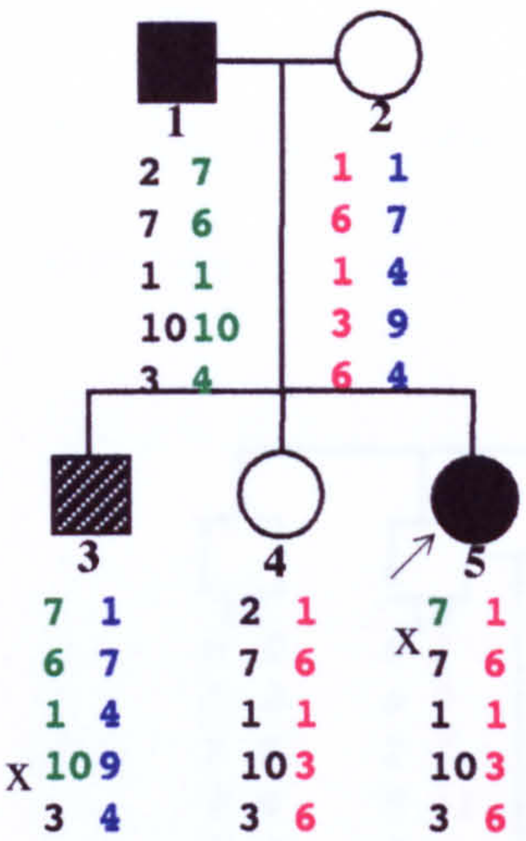
Pedigree 16



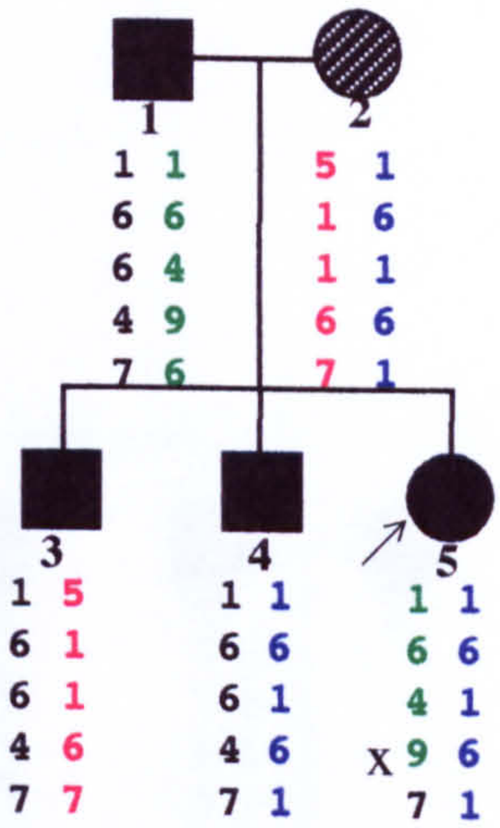
Pedigree 17



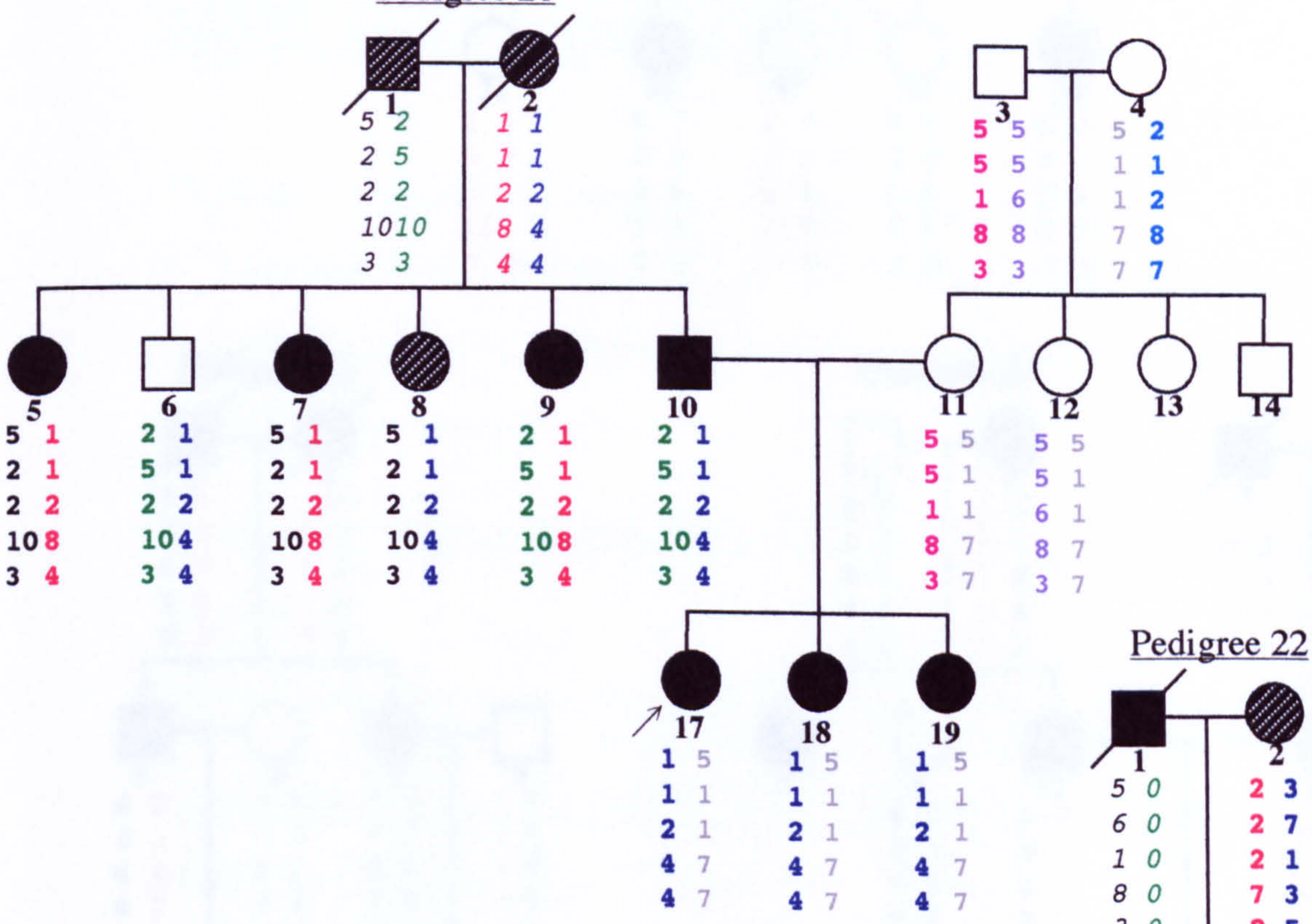
Pedigree 18



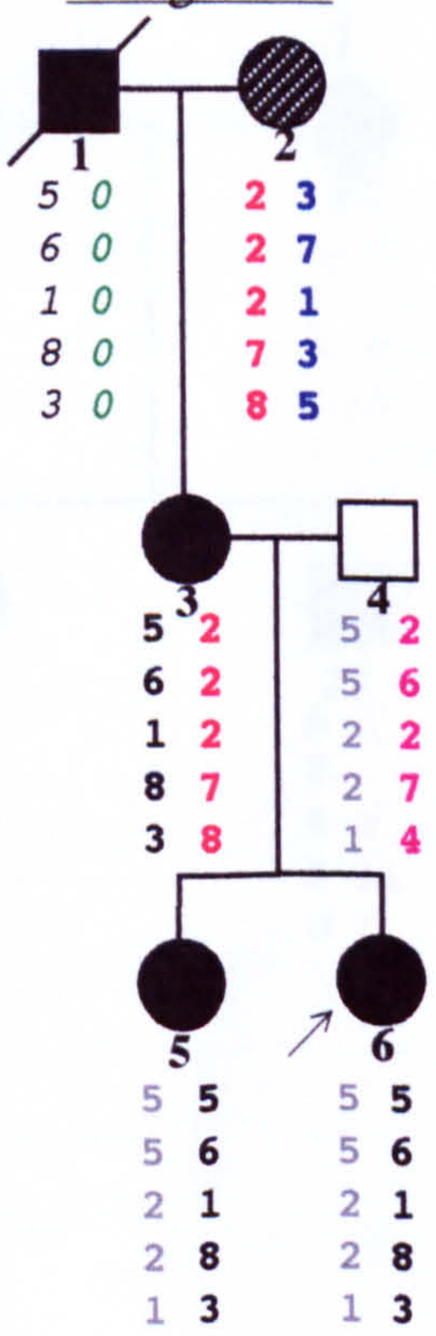
Pedigree 19



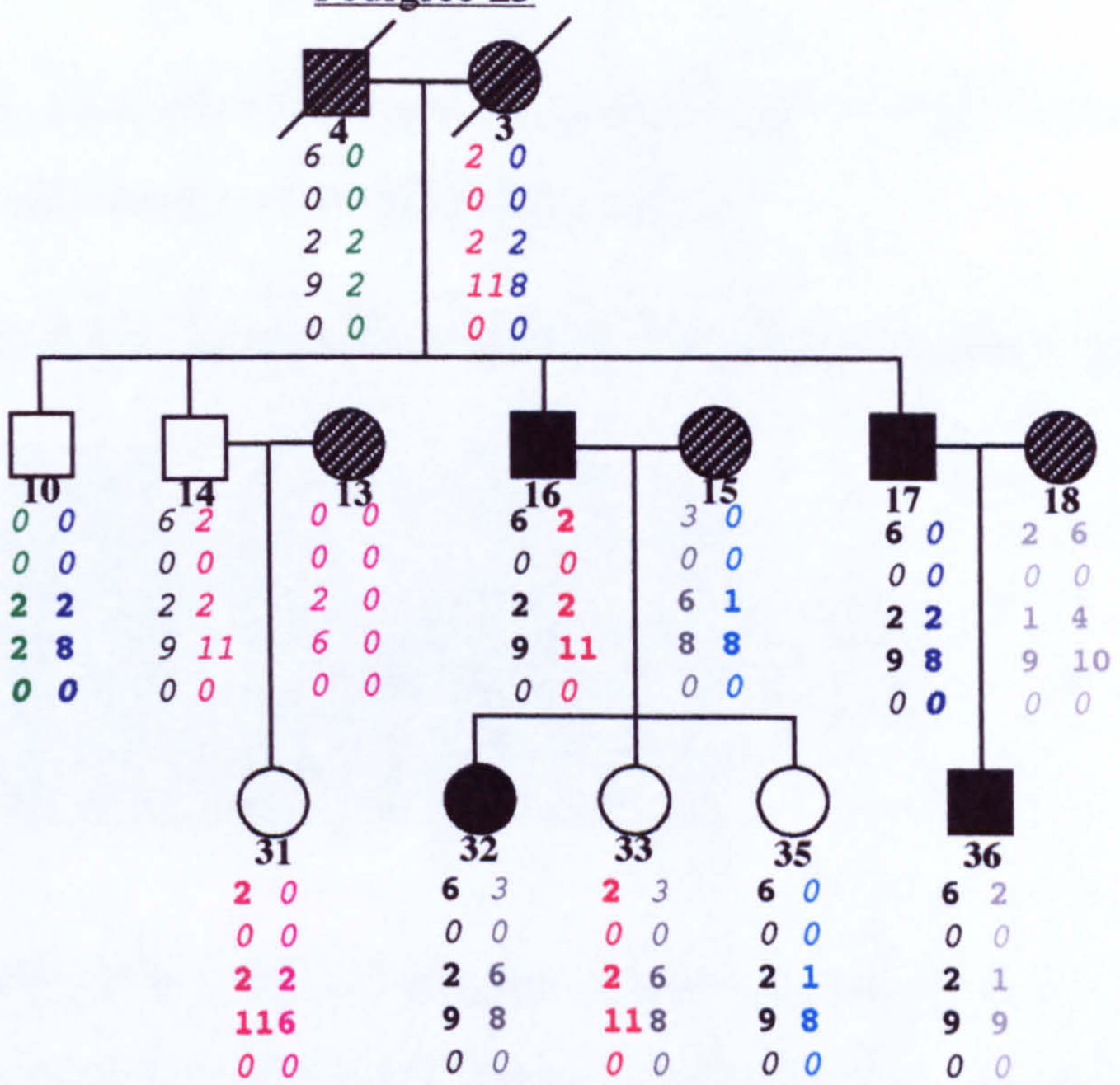
Pedigree 20



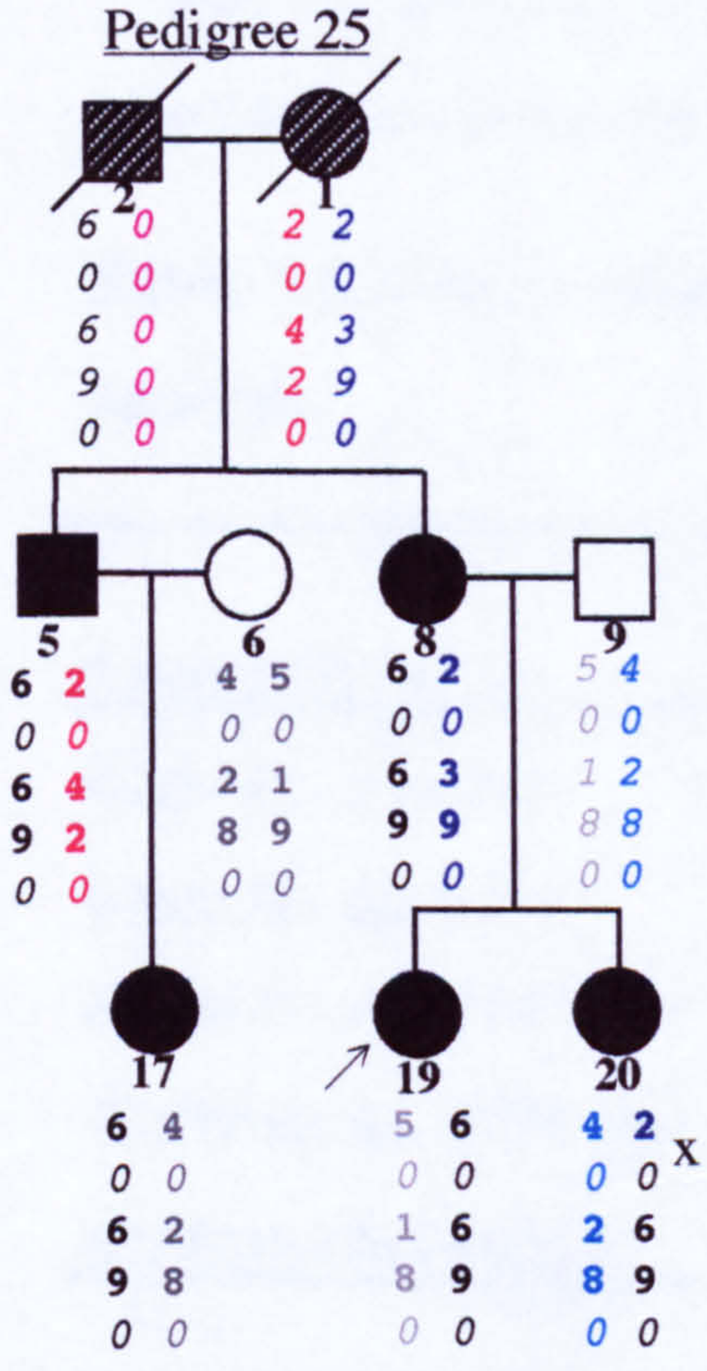
Pedigree 22



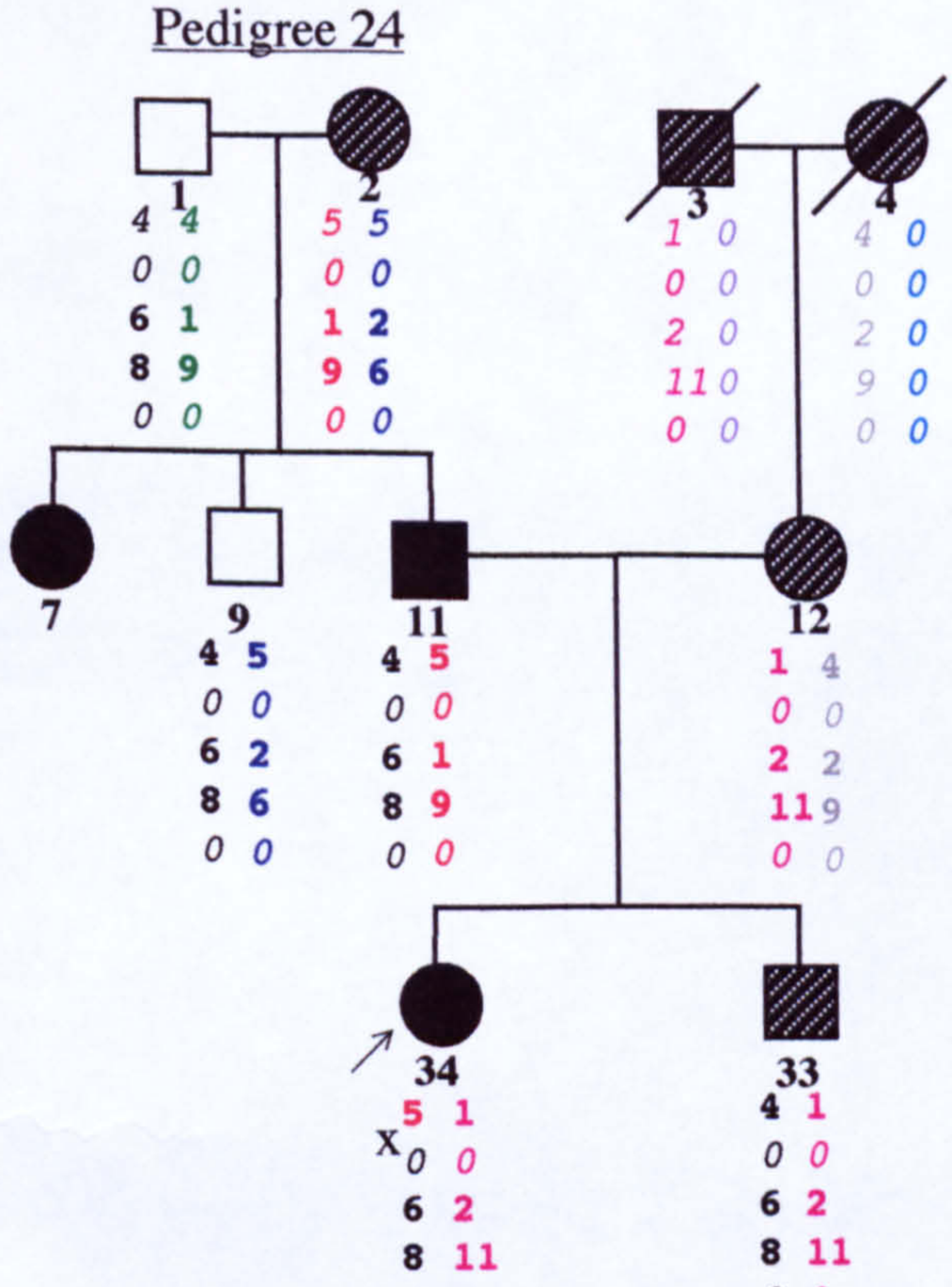
Pedigree 23



Pedigree 25



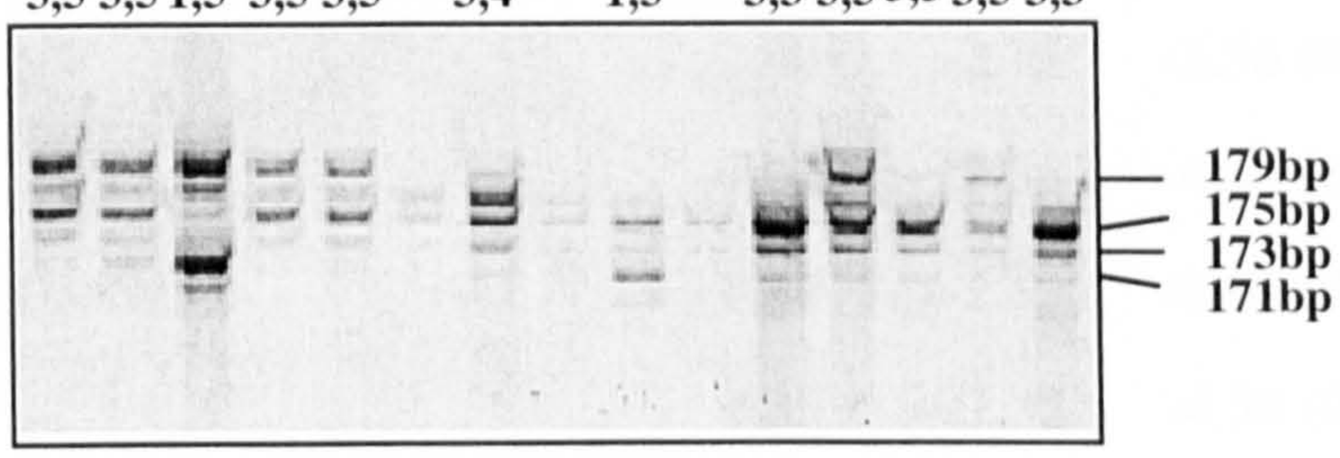
Pedigree 24



3.1.2 *CYP11a* locus linkage results

Figure 3.1.6 shows part of an autoradiograph of *CYP11a*(ac)n genotypes, the intragenic marker used in the linkage analysis.

Figure 3.1.6 Autoradiograph of *CYP11a*(ac)n genotypes.



The results of the *CYP11a* locus linkage analysis are as follows:

- The expected and observed number of recombinations for each of the inter-marker intervals are given in **table 3.1.3** and compared using the z-test. There is no evidence for map inflation.

Table 3.1.3 Recombination data for the *CYP11a* locus linkage analysis.

Locus Interval	Recombination Fraction (θ)	Expected cross-overs	Observed cross-overs	z-test	p-value
D15S155 - D15S153	0.099	19.84	18.10	0.307	0.378
D15S153 - D15S125	0.025	5.02	6.03	0.228	0.409
D15S125 - CYP11a(ac)$_n$	0.089	17.90	22.71	1.070	0.151
CYP11a(ac)$_n$ - D15S169	0.000	0.00	0.00		
D15S169 - D15S211	0.048	9.62	10.74	0.195	0.421

Information on the number of expected and observed recombinations, for the CYP11a locus linkage analysis, was generated by the GENEHUNTER program. There were a total of 201 meioses observed. Refer to the legend of Table 3.1.1 for abbreviations. There is no evidence for significant map inflations for any of the intervals tested (z-test).

- The linkage results for the *CYP11a* locus are summarised in **table 3.1.4** and **figures 3.1.7 - 3.1.9.** The *CYP11a* locus haplotypes are given in **figure 3.1.10.**

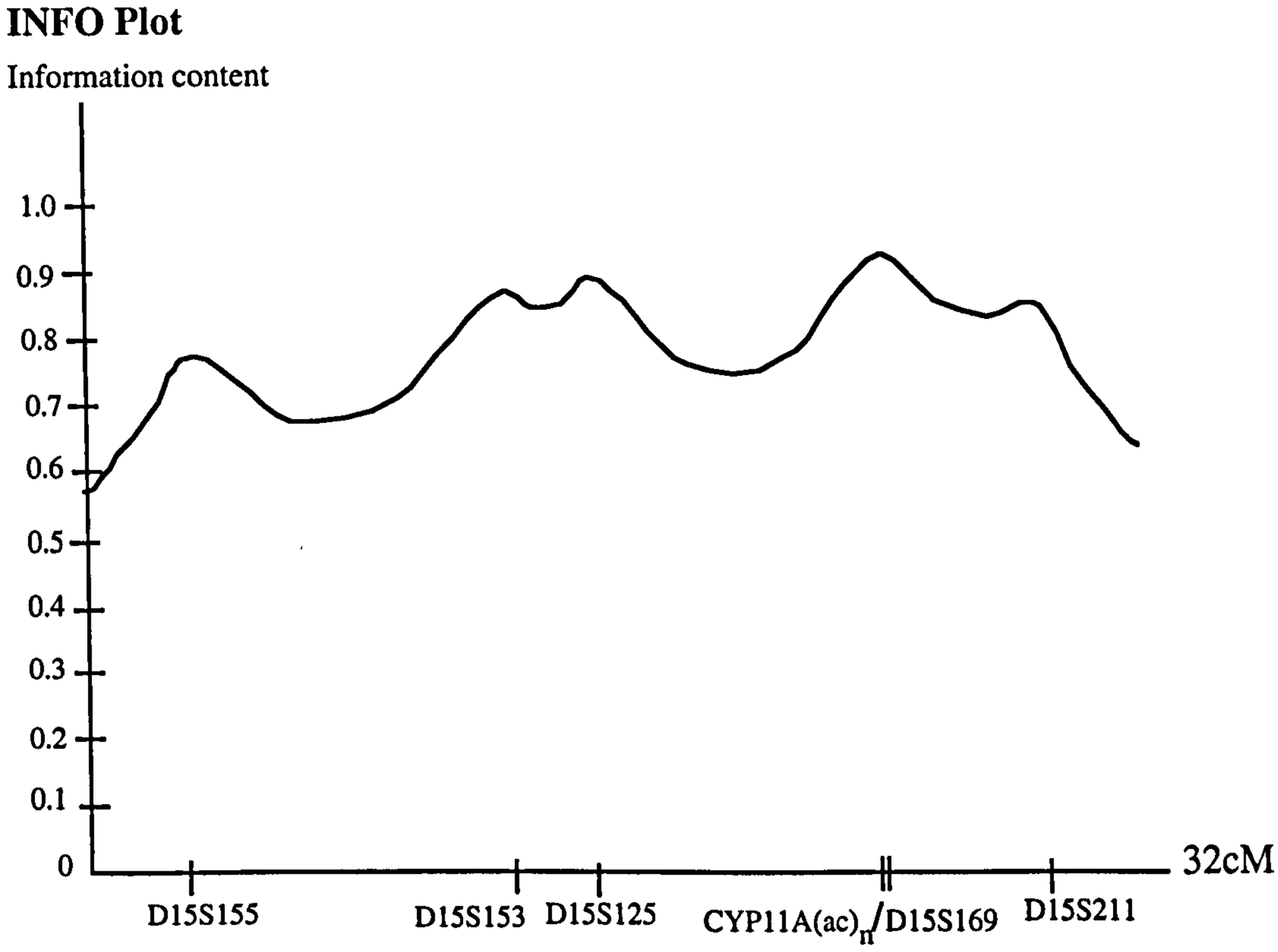
Table 3.1.4 GENEHUNTER results for *CYP11a* locus

Marker	Position (cM)	LOD score^R	LOD score^D (alpha,HLOD)	NPL score	p-value	info.
	-3.00	-2.04	-3.52 (0.16, 0.18)	0.78	0.029	0.70
	-2.00	-2.19	-3.96 (0.15, 0.19)	0.81	0.027	0.74
	-1.00	-2.36	-4.47 (0.15, 0.19)	0.84	0.026	0.78
D15S155	0.00	-2.53	-5.06 (0.15, 0.20)	0.86	0.025	0.84
	3.33	-2.32	-4.30 (0.17, 0.25)	0.94	0.021	0.78
	6.67	-2.18	-4.08 (0.19, 0.29)	1.04	0.018	0.81
D15S153	10.00	-2.23	-4.33 (0.20, 0.34)	1.15	0.014	0.92
	10.83	-2.18	-3.29 (0.21, 0.37)	1.22	0.012	0.89
	11.67	-2.15	-2.83 (0.22, 0.39)	1.26	0.011	0.90
D15S125	12.50	-2.16	-2.78 (0.23, 0.39)	1.25	0.011	0.93
	15.50	-1.68	-0.14 (0.48, 1.13)	1.72	0.004	0.83
	18.50	-1.48	1.33 (0.61, 2.12)	2.30	0.001	0.83
CYP11a(ac)$_n$ / D15S169	21.50	-1.51	*1.92 (0.62, 2.84)*	*2.95*	*0.0005*	*0.95*
	23.10	-1.50	1.60 (0.62, 2.48)	2.59	0.0008	0.90
	24.70	-1.54	0.92 (0.58, 1.92)	2.22	0.001	0.88
D15S211	26.30	-1.62	-0.78 (0.41, 0.98)	1.87	0.003	0.90
	27.30	-1.50	-0.47 (0.44, 1.00)	1.80	0.003	0.84
	28.30	-1.39	-0.21 (0.46, 1.02)	1.72	0.004	0.79
	29.30	-1.29	0.01 (0.48, 1.03)	1.65	0.005	0.75

Refer section 2.4.3 for GENEHUNTER input parameters and to the legend of table 3.1.2 and section 2.4.3 for the key to the output results. For the recessive parametric analysis, a LOD score^R of -2 or less was obtained across the region from marker D15S155 to D15S125, but the LOD score^R for the region from CYP11a(ac)n/D15S169 to D15S211 did not achieve criteria for proof or rejection of linkage. Maximum multipoint parametric (under the dominant model LOD score^D) and non-parametric scores are shown in italics. There is zero genetic distance between the D15S169 and CYP11a (ac)$_n$ markers.

- High information content has been obtained for all the markers used, with a maximum of 0.95 at the *CYP11a(ac)n/D15S169* locus and a minimum value of 0.84 at marker *D15S155* (**table 3.1.4** and **figure 3.1.7**).

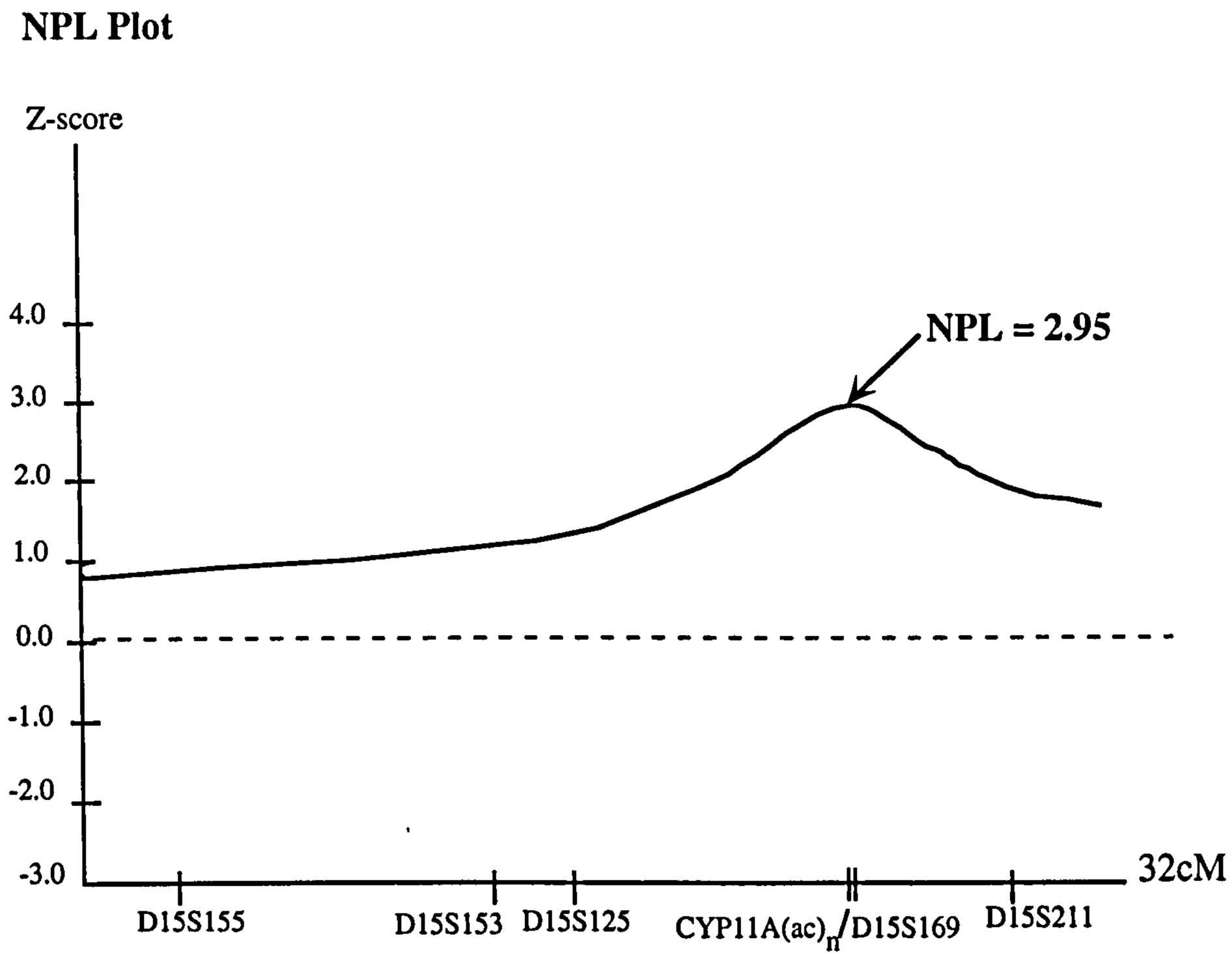
Figure 3.1.7 Information plot for the *CYP11a* locus linkage analysis.



Information plot is a graphic representation of the information extracted at each point along the linear map (see table 3.1.4). Information content is generally high at all markers.

- The non-parametric results for the *CYP11a* locus have provided evidence for excess allele sharing, generating a maximum NPL score of 2.95 (p = 0.0005) at the *CYP11a* locus (**table 3.1.4, figure 3.1.8**).

Figure 3.1.8 NPL plot for the *CYP11a* locus linkage analysis.

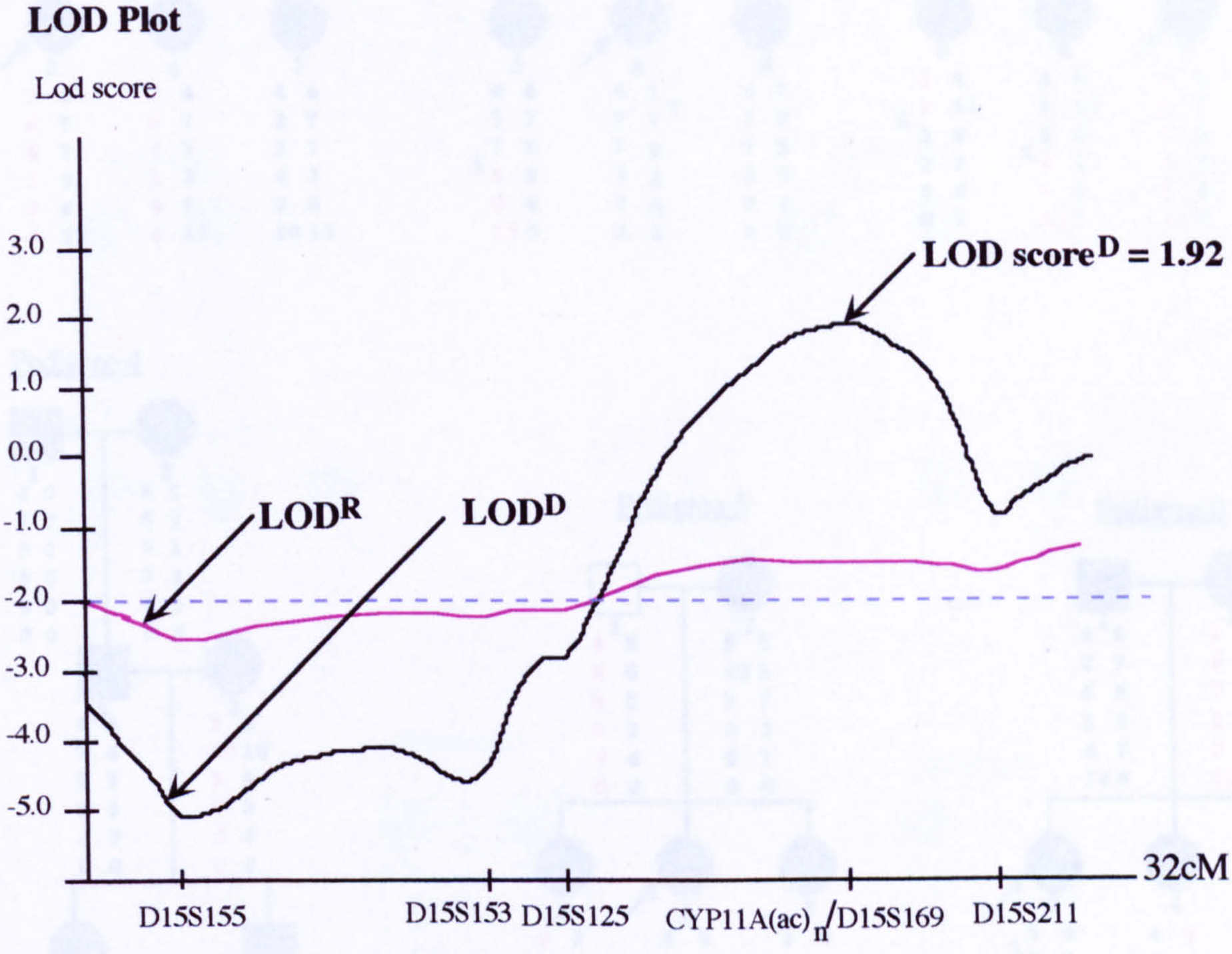


NPL plot is a graphic representation of the non-parametric linkage results (refer to table 3.1.4). The dashed line represents the mean of expected allele sharing distribution. The maximum NPL score of 2.95 (p = 0.0005), which is suggestive of linkage, was obtained at the CYP11a(ac)n/D15S169 locus.

- For the parametric analysis, the results under the dominant model have provided suggestive evidence for linkage. The overall multipoint LOD score^{D} obtained at the *CYP11a* $(ac)_n$/*D15S169* locus is 1.92 (which has reached the criteria of minimum LOD score of +1.9 for suggestive linkage (refer to **section 2.4.3**)). Since evidence for linkage has been obtained, a heterogeneity test on these data was also carried out. A maximum HLOD score of 2.8, $\alpha = 0.62$, was computed at the *CYP11a* $(ac)_n$/D15S169 locus. An M-test was then carried out on these data (refer to **section 2.4.7**). This test allowed for the extra parameter α (proportion of families linked to this locus) and provided statistically significant evidence ($\chi^2 = 4.24$, $p = 0.04$) for a better fit of the data under heterogeneity (compared to the

analysis under homogeneity (i.e. $\alpha = 1$)). The results of the heterogeneity test indicate that about 60% of the pedigrees may be linked to this locus (**table 3.1.4**, **figure 3.1.9**).

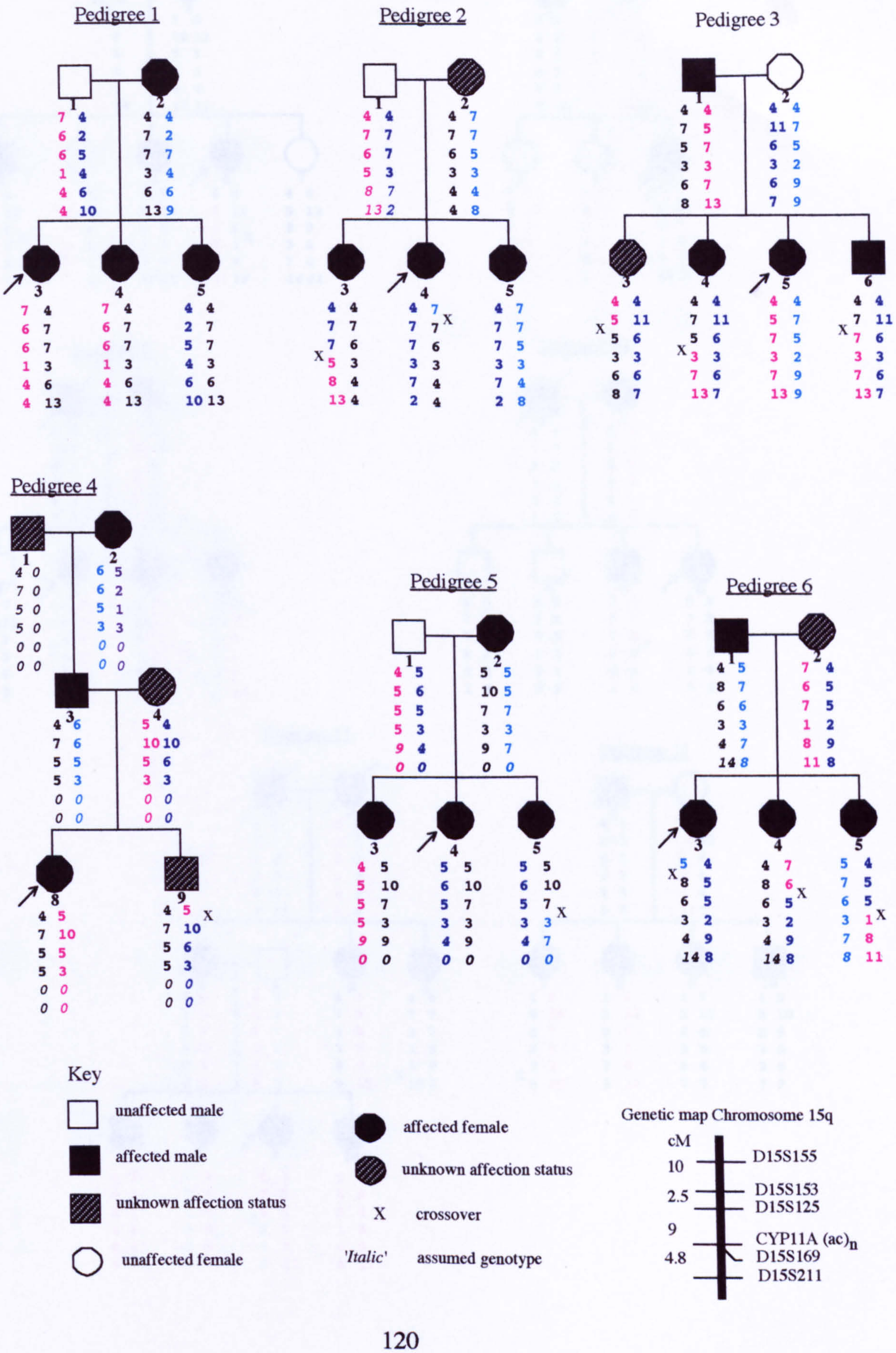
Figure 3.1.9 LOD plots for the *CYP11a* locus linkage analysis.



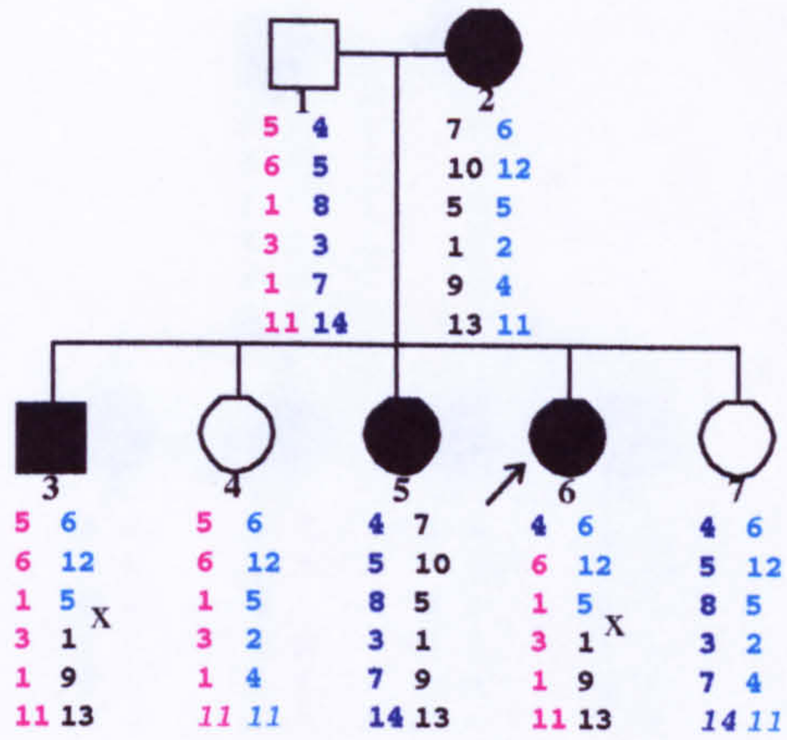
LOD plots are graphic presentations of the parametric linkage results (see table 3.1.4). The black and red curves represent the parametric linkage analysis under the dominant and recessive models respectively. The dashed line at -2 represents the minimum criterion required for rejection of linkage (section 2.4.3). Suggestive evidence for linkage to the CYP11a(ac)n/D15S169 locus has been obtained under the dominant analysis (LOD score = 1.92).

- Thus both the non-parametric and the parametric (under the dominant model) linkage results are consistent with evidence of linkage to the *CYP11a* locus (refer to **section 2.4.3**).

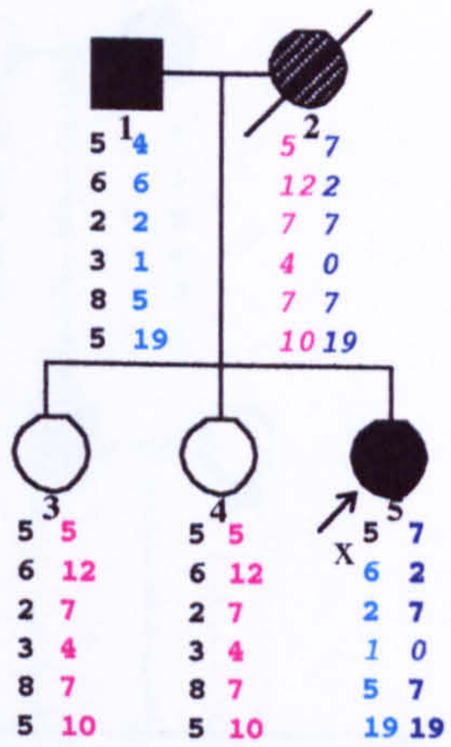
Figure 3.1.10 *CYP11a* locus haplotypes in the PCOS/MPB pedigrees



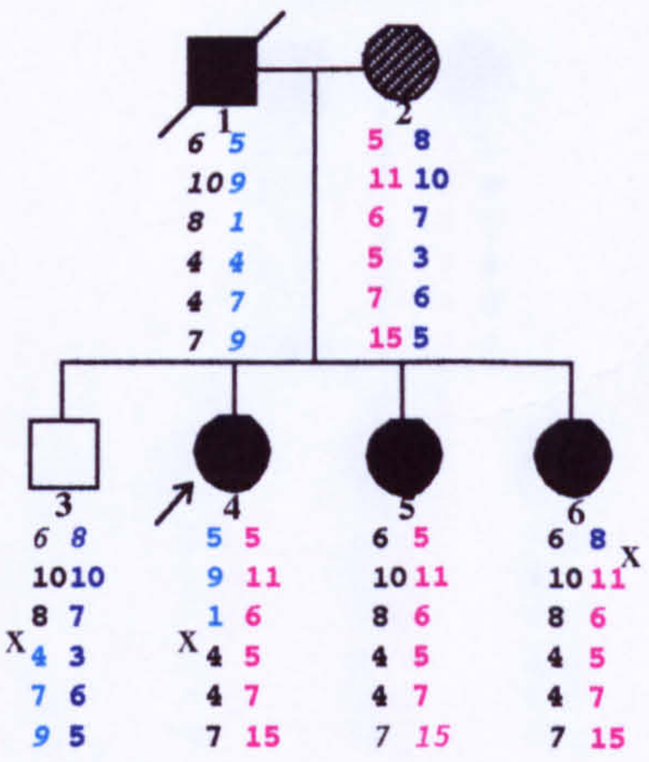
Pedigree 7



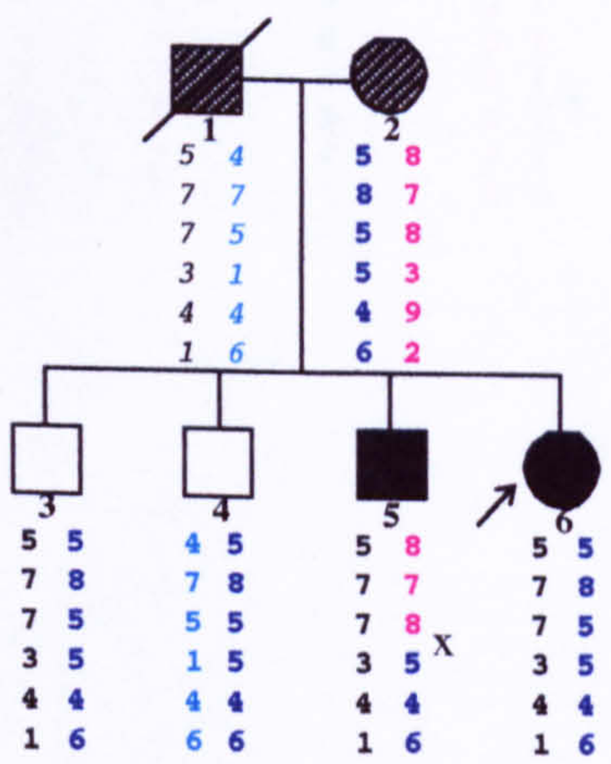
Pedigree 8



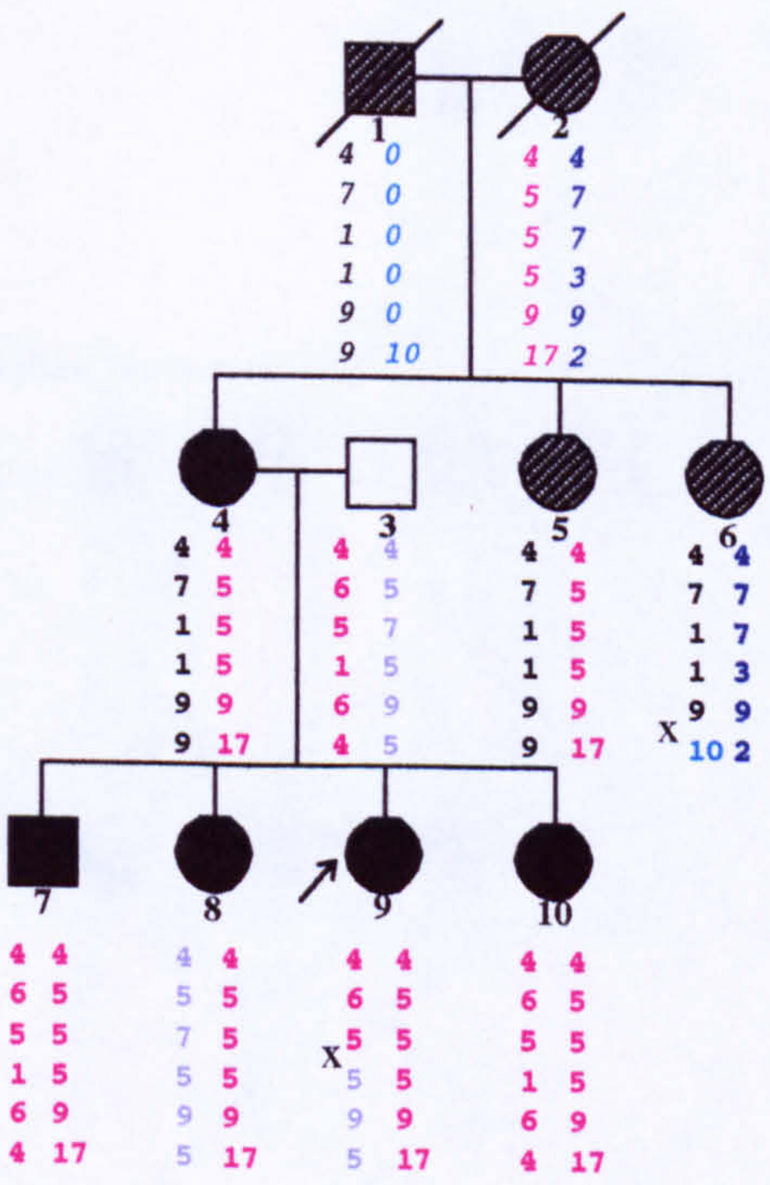
Pedigree 9



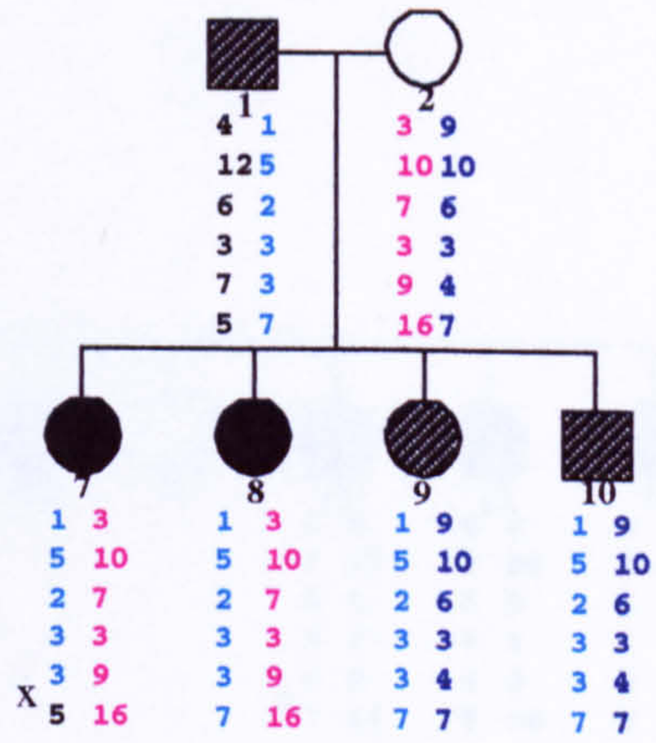
Pedigree 10



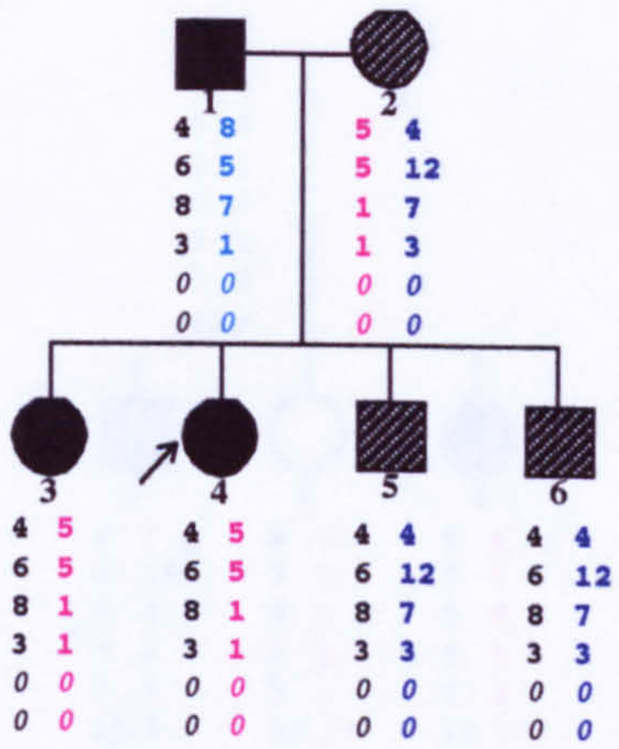
Pedigree 11



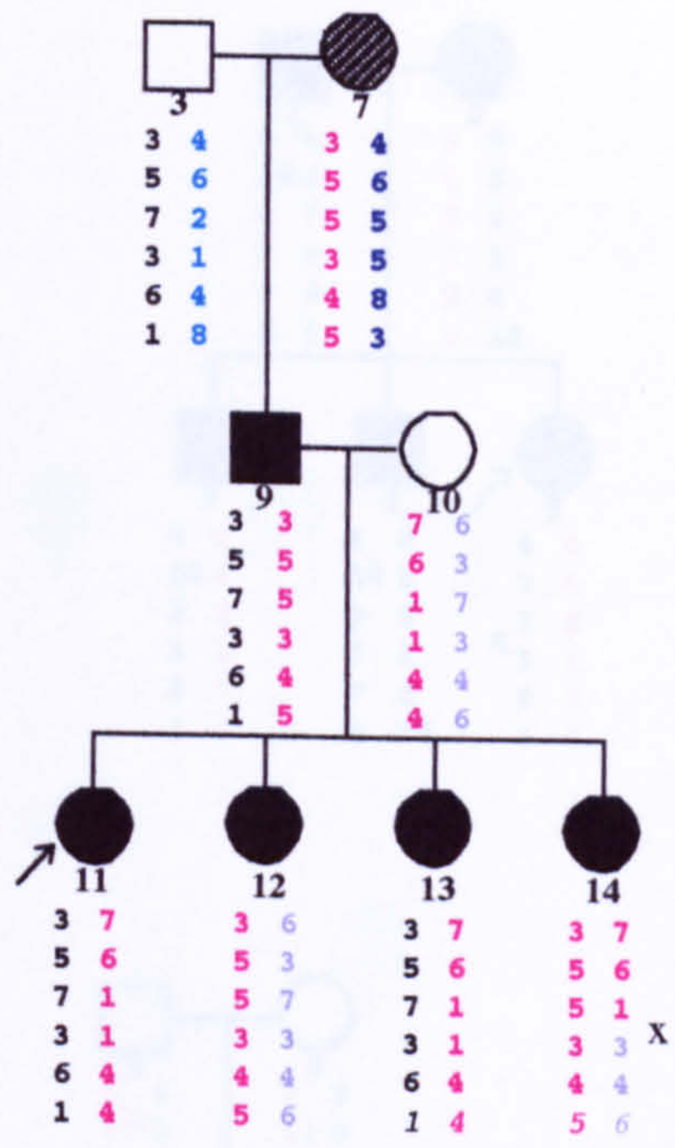
Pedigree 12



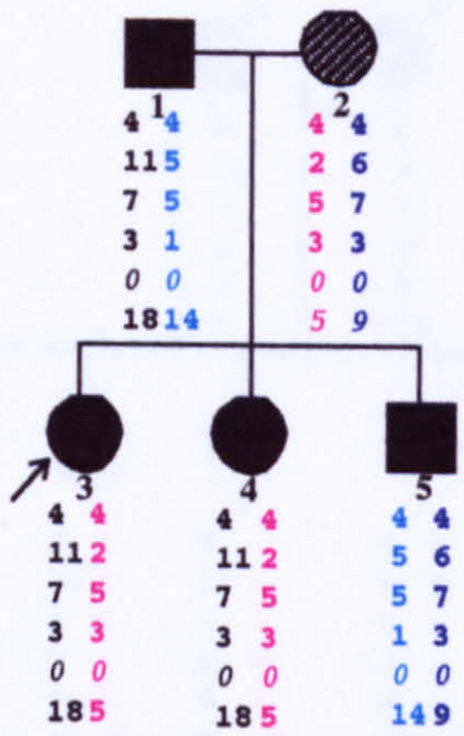
Pedigree 13



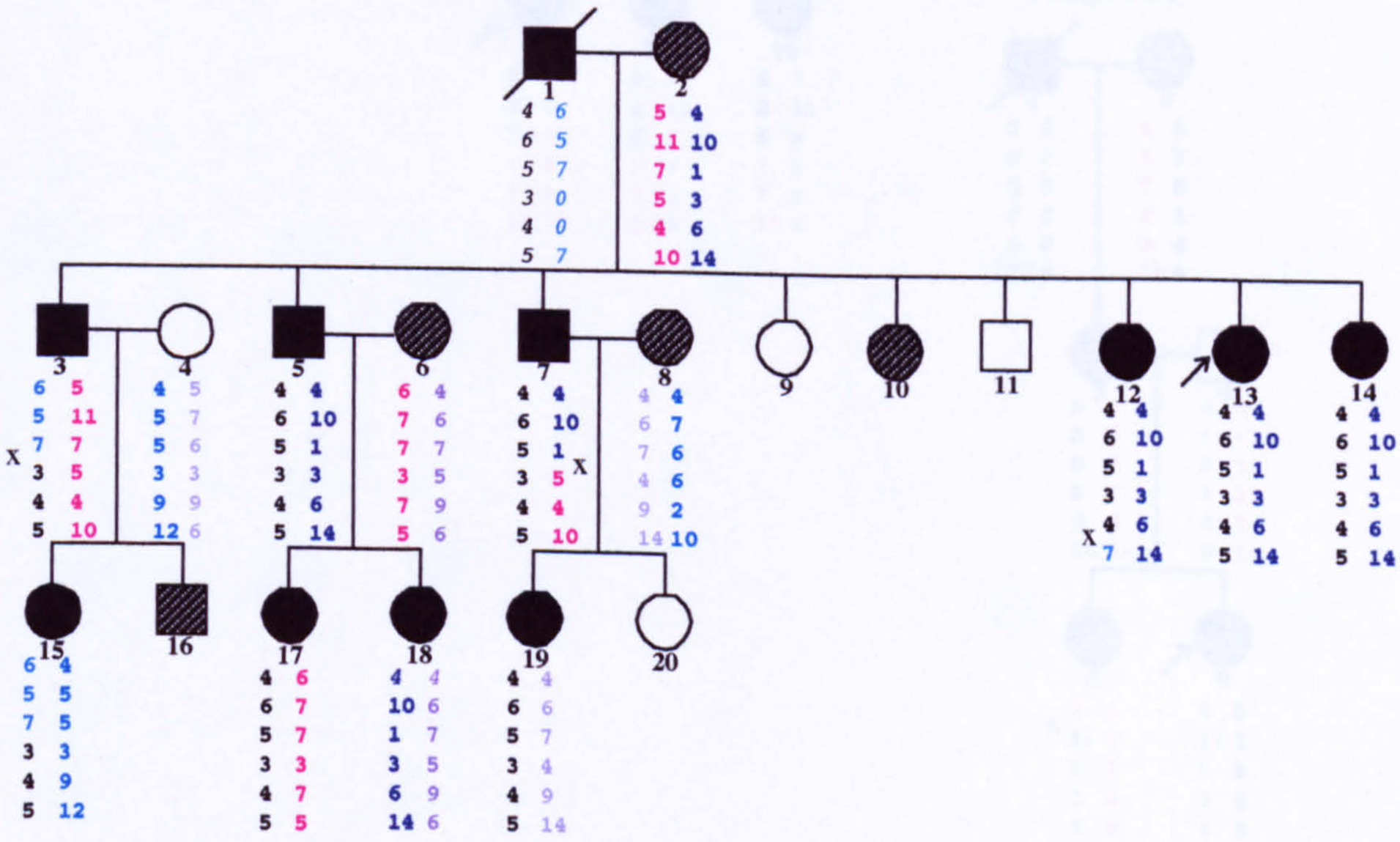
Pedigree 14



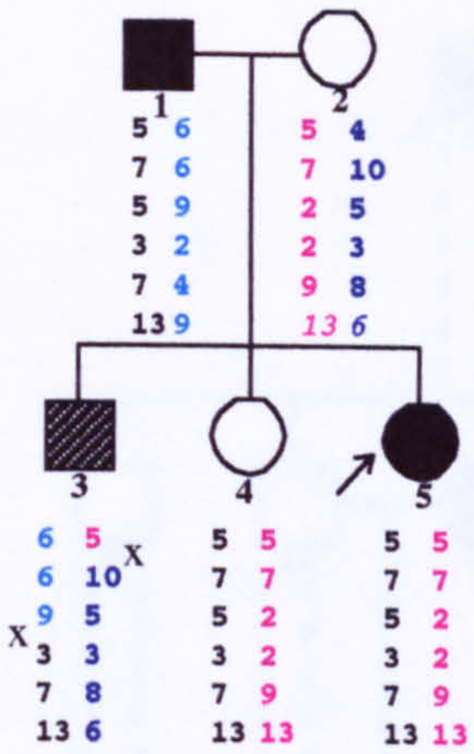
Pedigree 16



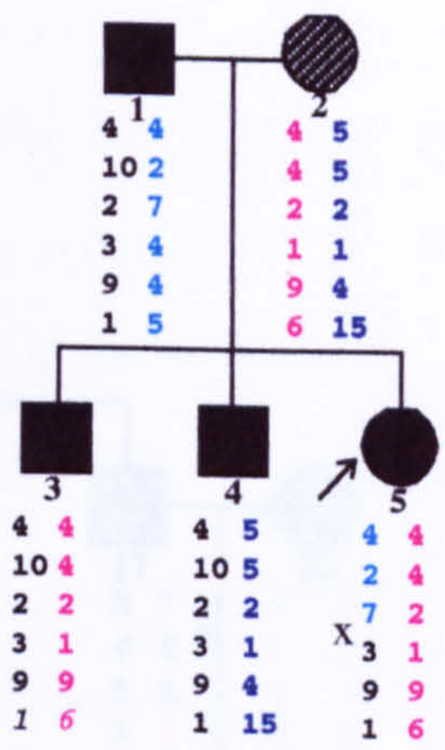
Pedigree 17



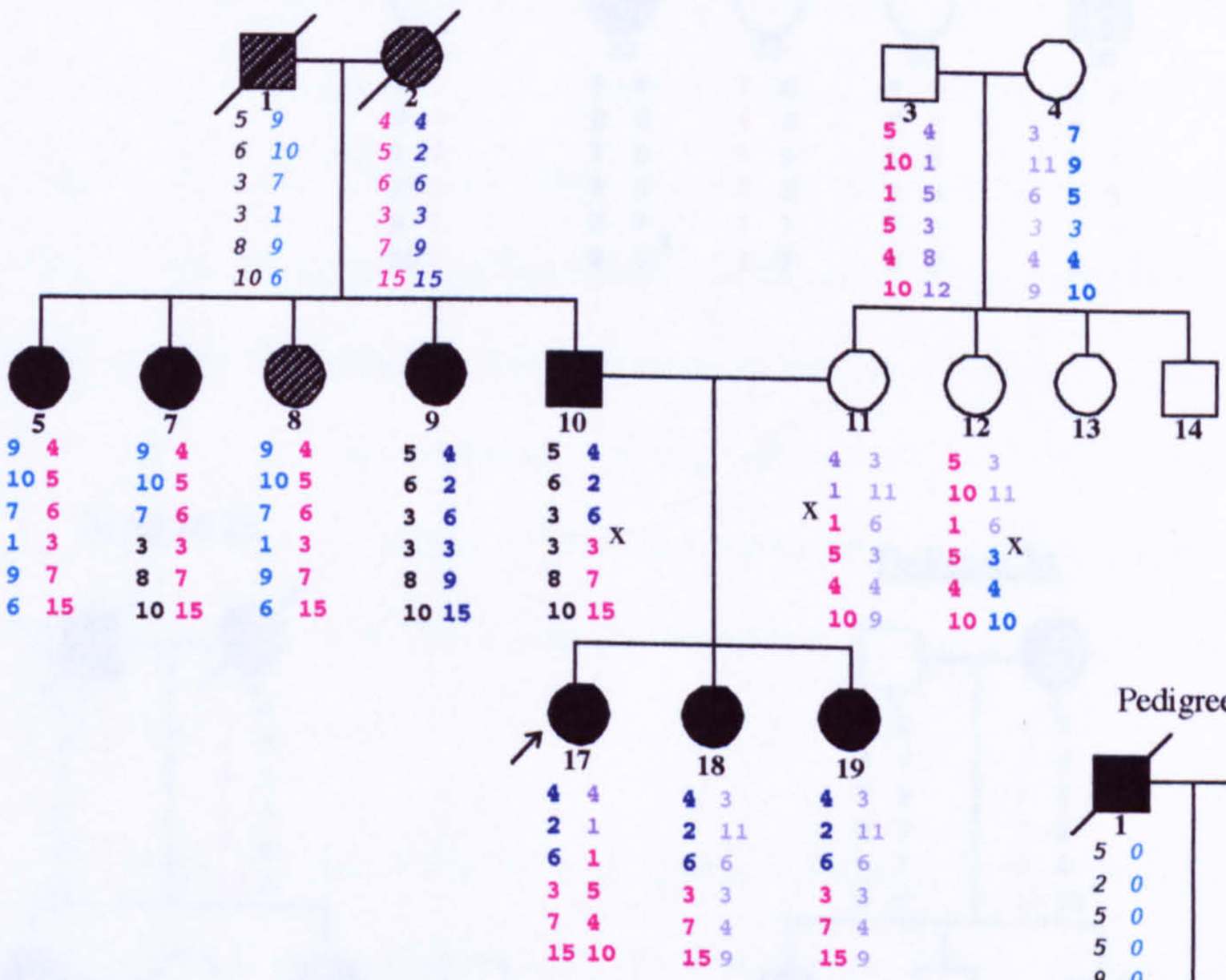
Pedigree 18



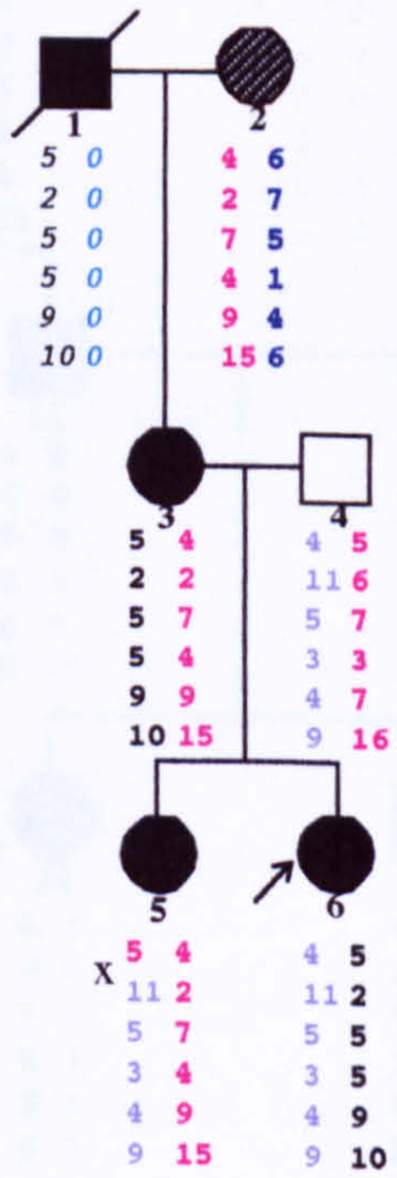
Pedigree 19



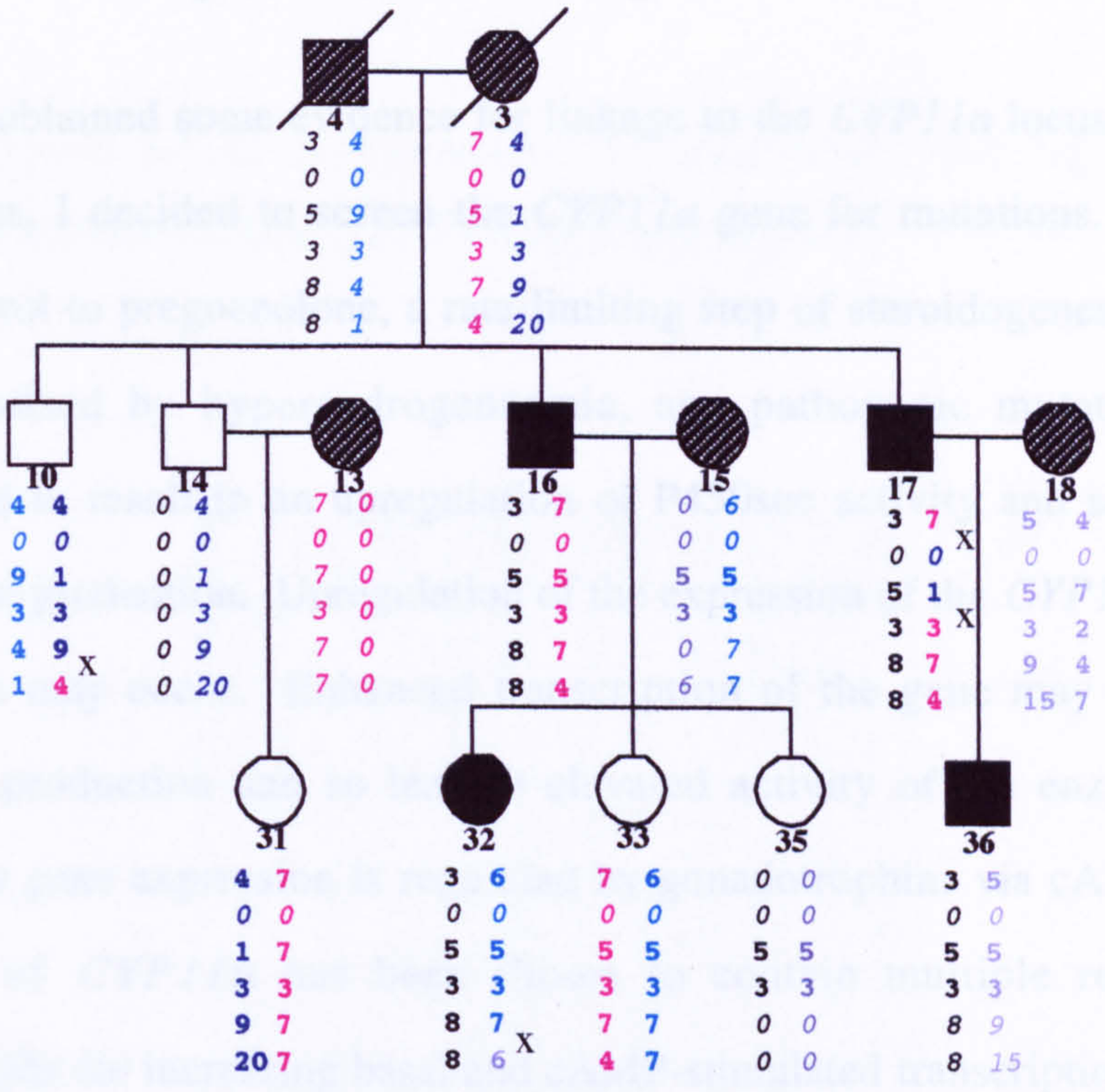
Pedigree 20



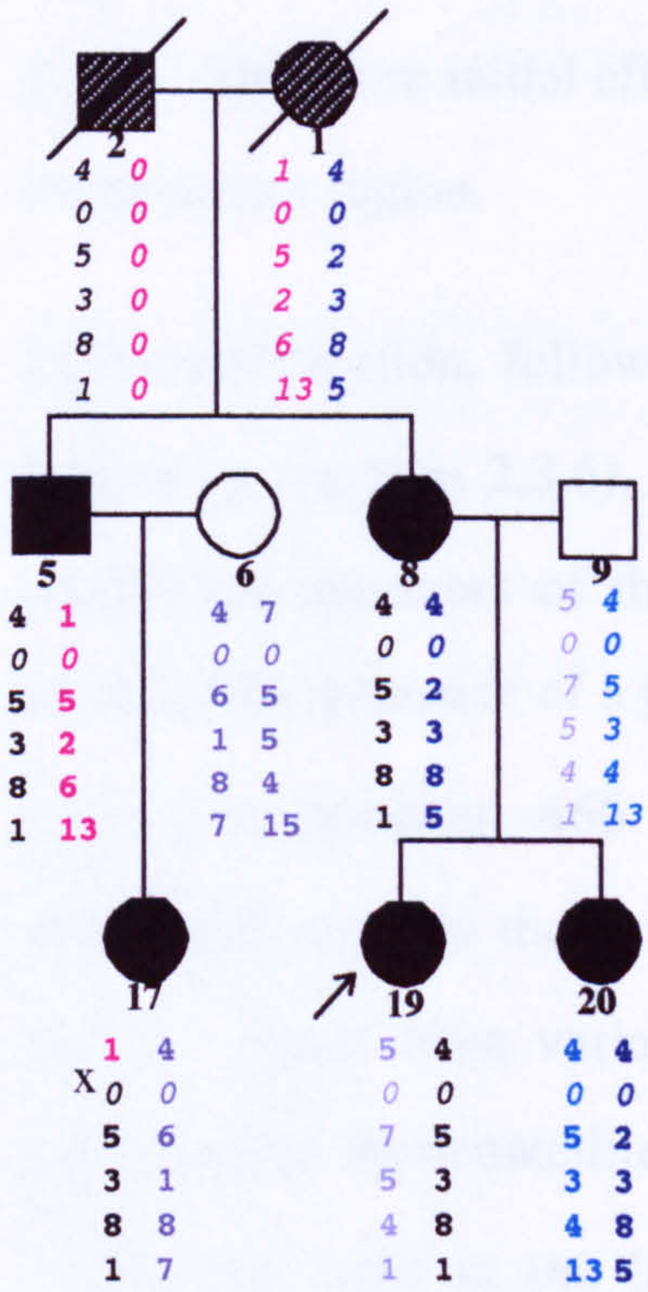
Pedigree 22



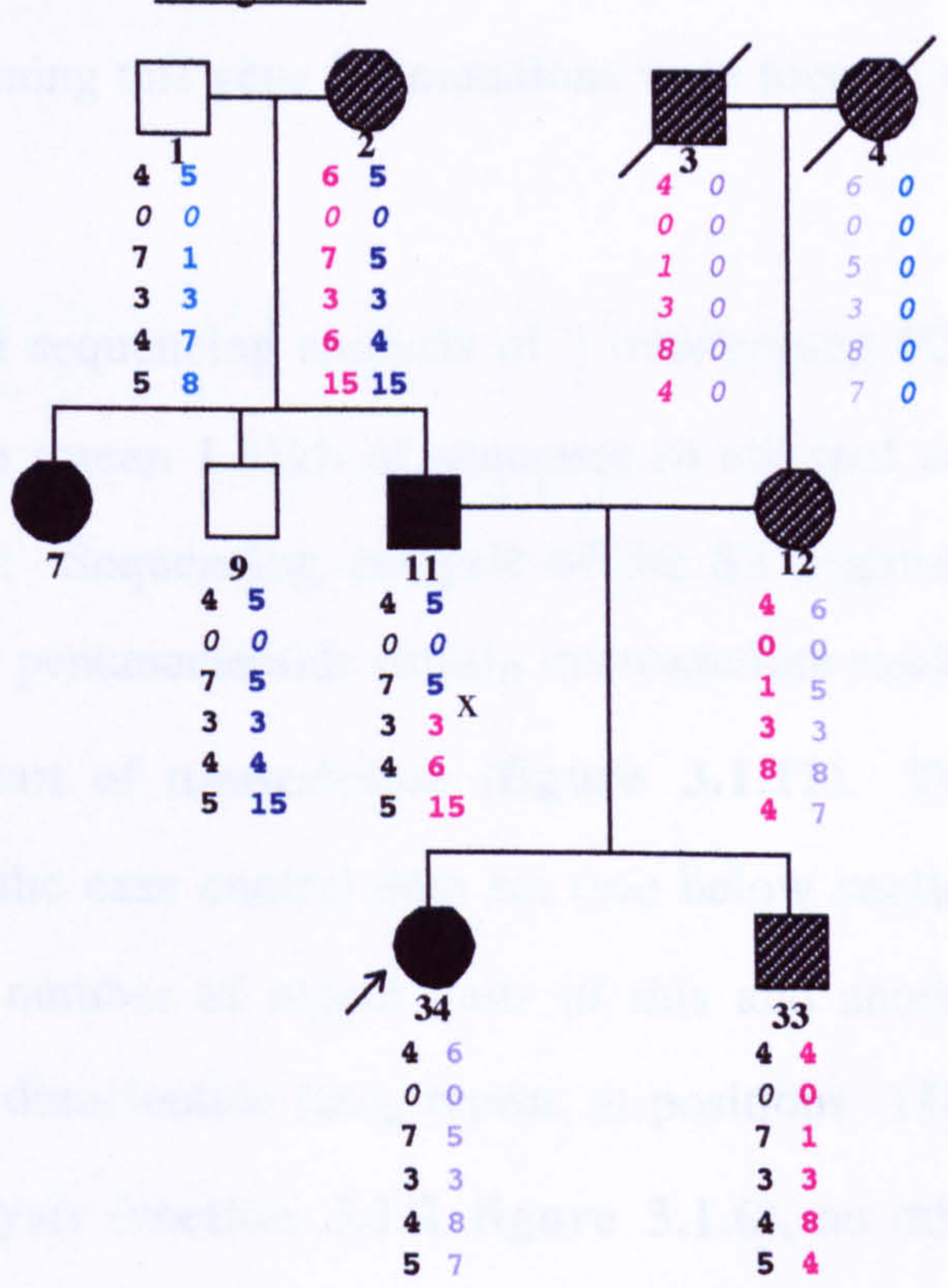
Pedigree 23



Pedigree 25



Pedigree 24

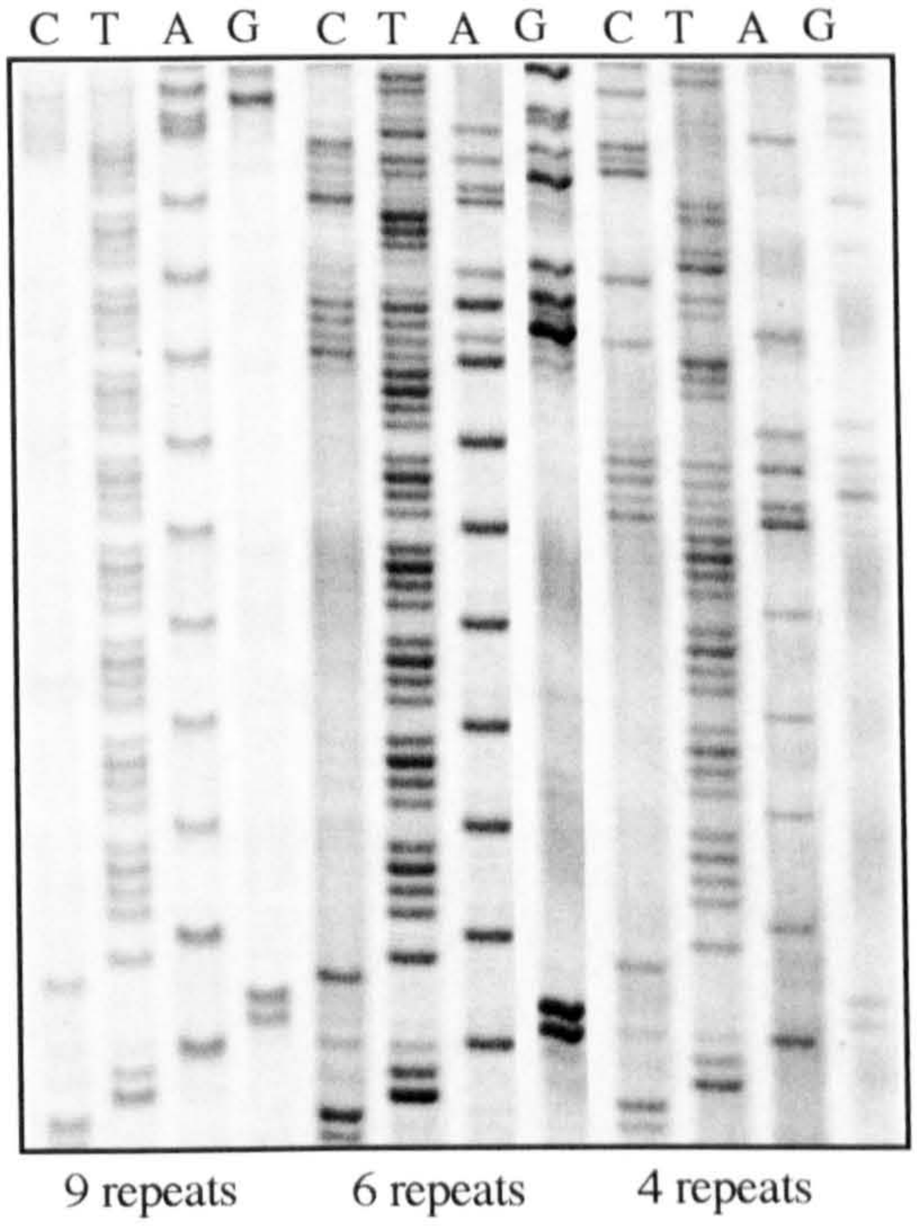


3.1.3. *CYP11a* promoter screening

Having obtained some evidence for linkage to the *CYP11a* locus in the PCOS/MPB pedigrees, I decided to screen the *CYP11a* gene for mutations. P450scc converts cholesterol to pregnenolone, a rate-limiting step of steroidogenesis. Since PCOS is characterised by hyperandrogenaemia, any pathogenic mutation in *CYP11a* is expected to result in an upregulation of P450scc activity and so lead to increased androgen production. Upregulation of the expression of the *CYP11a* gene is one way that this may occur. Enhanced transcription of the gene may result in increased protein production and so lead to elevated activity of the enzyme. In the ovary, *CYP11a* gene expression is regulated by gonadotrophins via cAMP. The promoter region of *CYP11a* has been shown to contain multiple regulatory elements responsible for increasing basal and cAMP-stimulated transcriptional activity, as well as elements involved in repression of expression (Moore *et al.*, 1990; Takayama *et al.*, 1994; Liu and Simpson, 1997b; Monté *et al.*, 1998). These include cAMP-regulated elements (CREs), SP-1 and steroidogenic factor 1 (SF-1) elements (**figure 2.3.6**). Therefore initial efforts at screening this gene for mutations were focused on the promoter region.

PCR amplification, followed by direct sequencing analysis of 3 overlapping PCR fragments (**section 2.3.6**), was used to screen 1.85kb of sequence in affected and unaffected members of the pedigrees. Sequencing analysis of the S3 fragment revealed the presence of a polymorphic pentanucleotide $(tttta)_n$ microsatellite marker present at position -466 from the start of transcription (**figure 3.1.11**). This polymorphism was then evaluated in the case control data set (see below **section 3.1.4**). Apart from variation in the number of repeat units of this and another polymorphic microsatellite marker (a dinucleotide $(ac)_n$ repeat, at positions -1314, which was used in the linkage analysis (**section 3.1.2, figure 3.1.6**), no other alterations to the published sequence were found in any individual.

Figure 3.1.11 Sequence of S3 fragment showing presence of the *CYP11a* (tttta)n repeat.

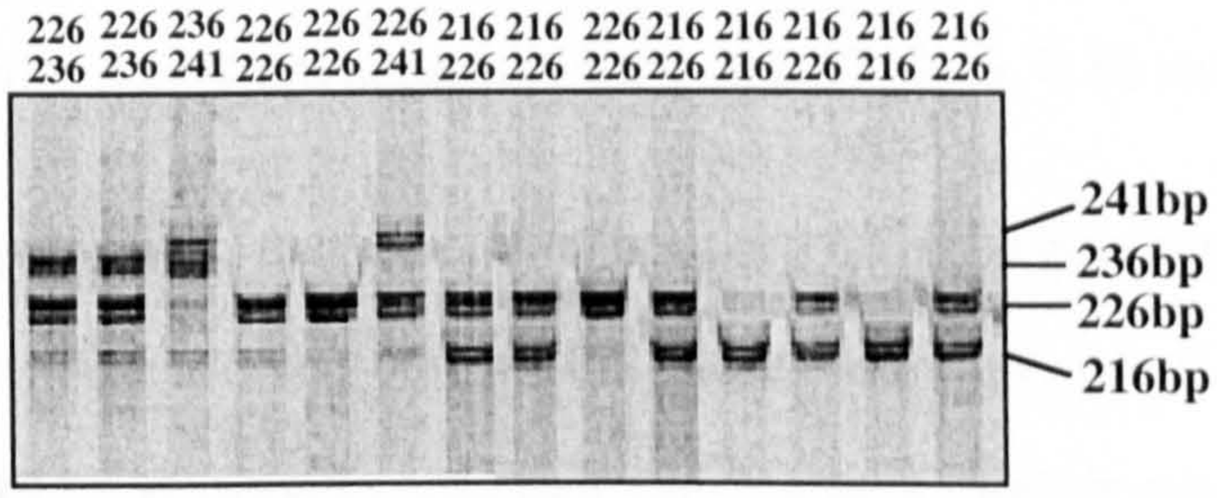


DNA sequence from three individuals, homozygous for the 216 allele (4 repeats), 226 allele (6 repeats) and 241 allele (9 repeats) respectively, is shown above. Sequencing analysis of the S3 fragment was carried out using the S3 reverse primer (refer to section2.3.6).

3.1.4. *CYP11a* $(tttta)_n$ association results

Individuals in the case-control study were genotyped for the *CYP11a* $(tttta)_n$ marker identified in the promoter region of *CYP11a* (see above). To facilitate the genotyping process, a new reverse PCR primer was designed and this was used together with the S3 forward primer to give a smaller PCR product (≈ 200bp) which could easily be analysed on a denaturing polyacrylamide gel (**section 2.3.4**). Four alleles were observed in our Europid data set and these were designated alleles 216, 226, 236 and 241, corresponding to the base pair size of the band observed on the gel (**figure 3.1.12**).

Figure 3.1.12 Autoradiograph of *CYP11a* (tttta)n genotypes.



The number of repeat units were 4, 6, 8 and 9 (**figure 3.1.11**), and the observed allele frequencies were 0.59, 0.28, 0.04 and 0.09 respectively. Since the initial cohort screened with this polymorphism was relatively small (<180 women), there were only a few individuals with the rarer genotypes. Therefore, individuals were assigned based on the presence or absence of the most common (4 repeat units) allele, the 216 allele. Individuals were allocated to two groups according to their genotype, those having at least one copy of the 216 allele (designated 216+) and those with no 216 allele (designated 216-). To compensate for the possibility of a grouping bias (which may be considered as equivalent to carrying out multiple significance tests) a Bonferroni correction (**section 2.4.8**) was applied to all subsequent significant associations obtained. This was done by multiplying the p-value obtained by the number of potential statistical comparisons that could be made for this polymorphism

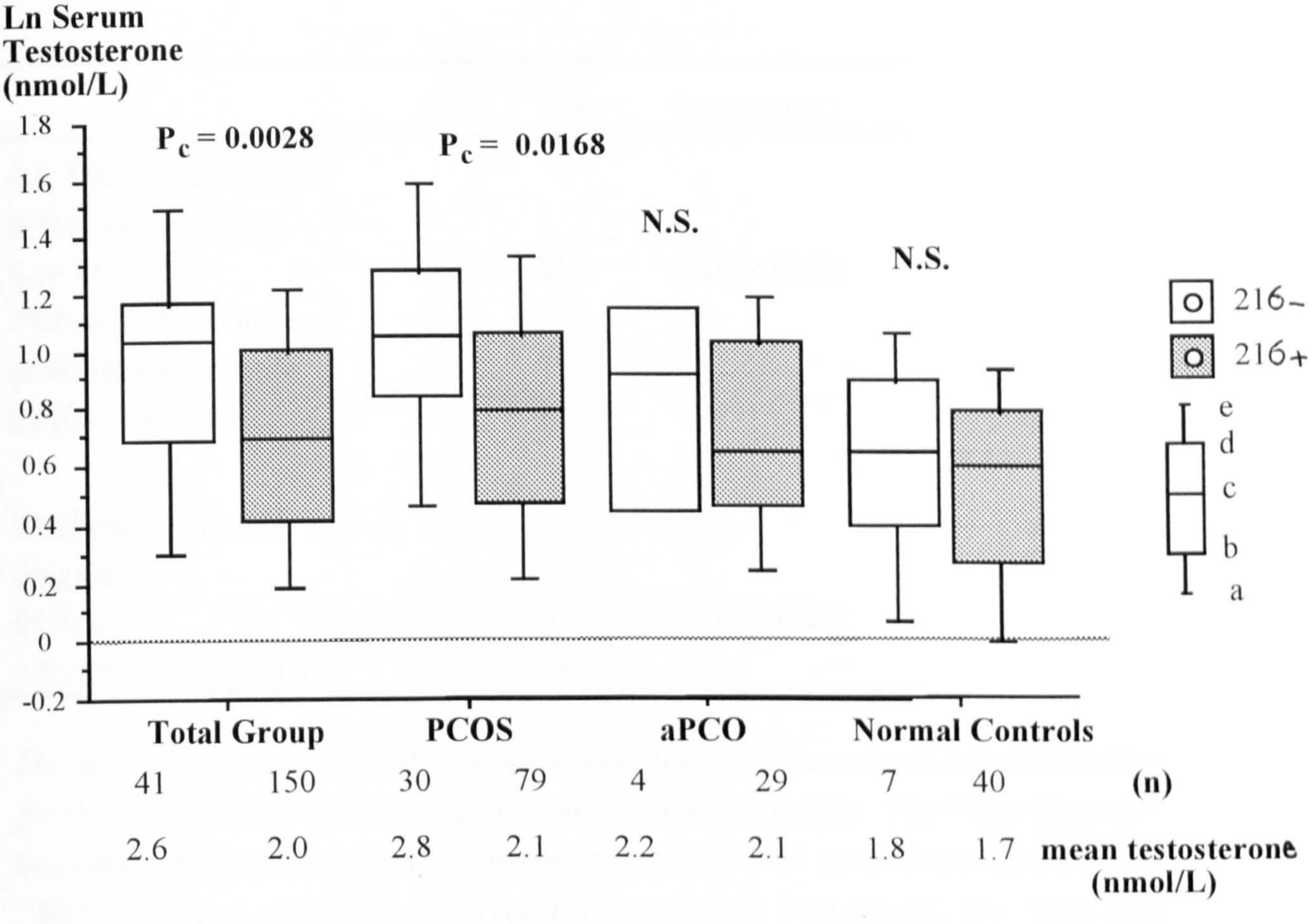
number of potential statistical comparisons that could be made for this polymorphism (i.e. by a factor of 4 since a similar significance test may be carried out for each of the 4 alleles). A comparison was then made between the geometric mean of serum testosterone levels between these two genotype groups. The results obtained are as follows:

- An association between this polymorphism and serum testosterone levels was found in the total data set giving a Mann Whitney p-value of 0.0007 (**figure 3.1.13**). This remains highly significant even after a Bonferroni correction (P_c = 0.0028).

- Another way of presenting this association, is to compare the *CYP11a* (tttta)n genotype distributions between each quartile of serum testosterone level within the total data set (**table 3.1.5**). The proportion of individuals with 216- genotype in each quartile was (from highest to lowest) 0.37, 0.15, 0.16 and 0.14 (P_c = 0.03).

- On subgroup analysis, the association between *CYP11a* genotype and testosterone was evident in the women with PCOS (Mann Whitney, p = 0.0042 (P_c = 0.017) **figure 3.1.13**). However, although there was a similar trend in testosterone distribution in the aPCO (asymptomatic PCO) and normal control groups, this was far from significance. This is probably due to a type II error given the small 216- genotype group sizes for both of these subgroups (with only 4 and 7 individuals for the aPCO and normal control groups respectively (**figure 3.1.13**)).

Figure 3.1.13 Association of *CYP11a* with the geometric mean of total serum testosterone levels in the case control data-set.



The Mann-Whitney test was used to compare the geometric means (Ln) of total serum testosterone levels between the two CYP11a genotype groups. Bonferroni corrected P-values are given above (except for the "aPCO" and "normal control" groups which were not significant (N.S.), the uncorrected p-values are: aPCO, p = 0.53; normal controls, p = 0.52). The "Total group" included all the women in the data set for whom serum testosterone measurements were available. The "PCOS" group are the women with polycystic ovaries and symptoms of anovulation and/or hirsutism. The "aPCO" group are the asymptomatic women with polycystic ovaries and the "normal control" group are the control subjects with normal ovarian morphology. "216+" sub-group are individuals with at least one 216 allele and "216-" are those with no 216 allele. "a", "b", "c", "d" and "e" are the 10th, 25th, 50th, 75th and 90th percentiles of Ln serum testosterone levels respectively. The mean (antilog of the geometric mean) and standard deviation range (derived from antilog of (geometric mean + SD) to antilog (geometric mean - SD)) of serum testosterone levels for each group are also given above.

Tables 3.1.5 Comparison of *CYP11a* genotype distributions between different sub-groups in the case control data set.

	Genotypes 216+	216-	P-value(P_c)
<u>Analysis by quartile of serum testosterone:</u>			
Top Quartile	37	21	0.008 (0.03)
50th-75th Percentile	34	6	
25th-50th Percentile	37	7	
Bottom Quartile	42	7	
<u>Analysis by clinical diagnosis:</u>			
PCOS	115	35	0.029 (0.12)
aPCO/Normal controls	111	16	

The quartiles of serum testosterone were obtained from the natural Log distribution for the total data set and levels quoted are the antilog values. The "Top Quartile" includes women with serum testosterone levels of 2.80 nmol/L and greater, the "50th-75th Percentile" includes levels between 2.10 to 2.80 nmol/L, the "25th-50th Percentile" includes levels between 1.50 to 2.10 nmol/L and the "Bottom Quartile" includes those with levels of 1.50 nmol/L and lower, for the entire data-set (affecteds and controls). The "genotypes" column is the observed number of individuals with either the 216+ or the 216- genotype for each sub-group. "216+" genotype includes individuals with at least one 216 allele and "216-" includes those with no 216 allele. The "PCOS" group are the women with polycystic ovaries and symptoms of anovulation and/or hirsutism. The combined "aPCO/normal control" group includes the asymptomatic women with polycystic ovaries and the normal control subjects with normal ovarian morphology. Chi Square contingency tables were used to compare genotype distributions between the different groups (both the uncorrected and corrected (P_c) p-values are given above).

- Since the association between serum testosterone levels and *CYP11a* genotypes was only seen in the PCOS group, taking the entire data set (including those without testosterone levels available), the *CYP11a* genotypes of the women with PCOS was compared with that of a combined normal control/aPCO group. A

significant difference in the distribution of genotypes between these groups was obtained ($p = 0.03$). However, after correcting for multiple testing ($P_c = 0.12$) this difference was no longer statistically significant (**table 3.1.5**).

Given the association of this gene with increased serum androgen levels, the possibility of an association with the phenotypic expression of hyperandrogenism was also examined. Therefore, a comparison was made between the genotype distributions of women classified according to hirsutism (hirsute vs. non-hirsute). Initially, in a group of 39 hirsute women and 62 non-hirsute subjects, a significant difference in genotype distributions $p = 0.01$ ($P_c = 0.04$) was obtained (Gharani *et al.*, 1997). However, with the recent updating of these results (by typing an extra 39 hirsute and 64 non-hirsute subjects) this difference was no longer observed ($p = 0.28$). This raises the possibility that the initial results may have been due to a type I error, and suggests that serum testosterone concentration is not the only determinant in the aetiology of hirsutism, which is also dependent on the possible contribution of end-organ factors. This is supported by the observation that some non-hirsute women with PCOS have similar levels of hyperandrogenaemia as hirsute PCOS women and there are hirsute subjects with polycystic ovaries with normal serum androgen concentrations (Franks, 1989; Franks, 1995).

In summary, the results of the chromosome 15 data show that the gene coding for P450 aromatase (*CYP19*) can be excluded as a major determinant of risk for PCOS/MPB. In contrast, the linkage results for *CYP11a* (encoding P450 side chain cleavage) provide support for the involvement of this gene in the aetiology of PCOS/MPB. Furthermore, the association data suggest that allelic variants of *CYP11a* may mediate the development of hyperandrogenaemia in women with PCOS.

3.2 CHROMOSOME 2p

The gene for the luteinizing hormone receptor (*LHR*) plays a crucial role in gonadal steroidogenesis and was therefore investigated as a candidate gene for involvement in the development of PCOS. The *LHR* gene had previously been cytogenetically mapped to chromosome 2p21 (Rousseau-Merck *et al.*, 1990). However, its precise genetic location relative to known polymorphic microsatellite markers had not been established. Therefore, this gene was initially mapped using the Stanford G3 radiation hybrid panel. Subsequently, microsatellite markers spanning the *LHR* locus were selected and used to carry out linkage analysis in the PCOS/MPB data set.

3.2.1 Radiation Hybrid panel screening with *LHR*

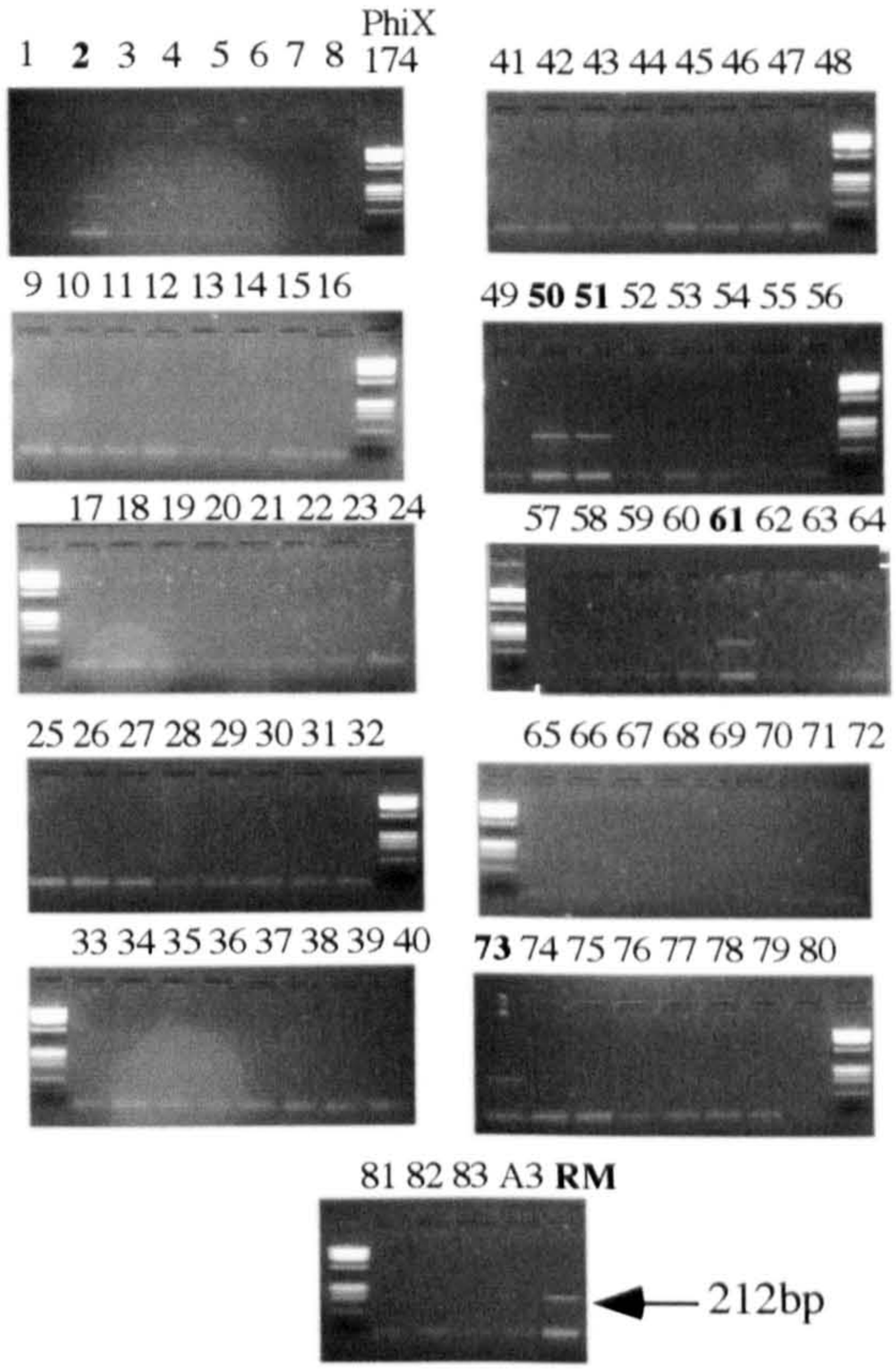
The *LHR* gene was mapped relative to known polymorphic microsatellite markers using the Stanford G3 radiation hybrid panel and the *LHR* 1a/1b primer set (**section 2.3.10**). Screening of the G3 RH panel was carried out in duplicate (**figure 3.2.1**).

Legend of figure 3.2.1.

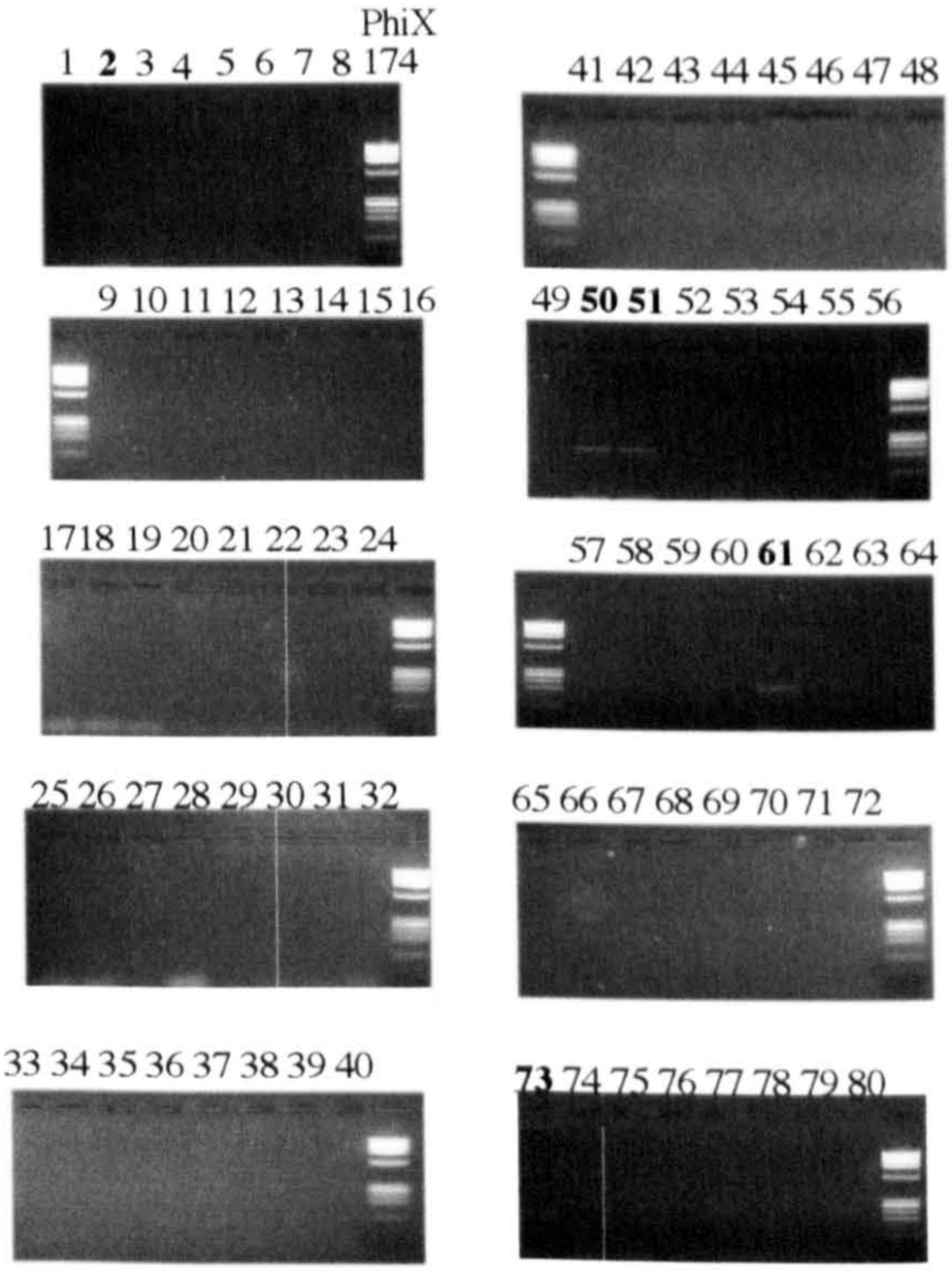
Panel A represents the first screening of the RH panel with the LHR 1a/1b primer set and panel B represents the second screening. Hybrids 1 to 83 are labelled accordingly for both panels A and B and positive clones are shown in bold font. Sample A3 and RM in panel A represent the negative (hamster DNA only) and positive (human DNA only) controls respectively. A 212bp product was only obtained in the positive control sample RM and five of the hybrid clones: 2,50,51,61 and 73. The result of the RH panel screening was scored as a vector (refer to methods section 2.3.10) and is given below. A smaller excess primer band is also visible in some samples.

Figure 3.2.1. Screening of G3 RH panel with the *LHR* 1a/1b primer set.

A



B



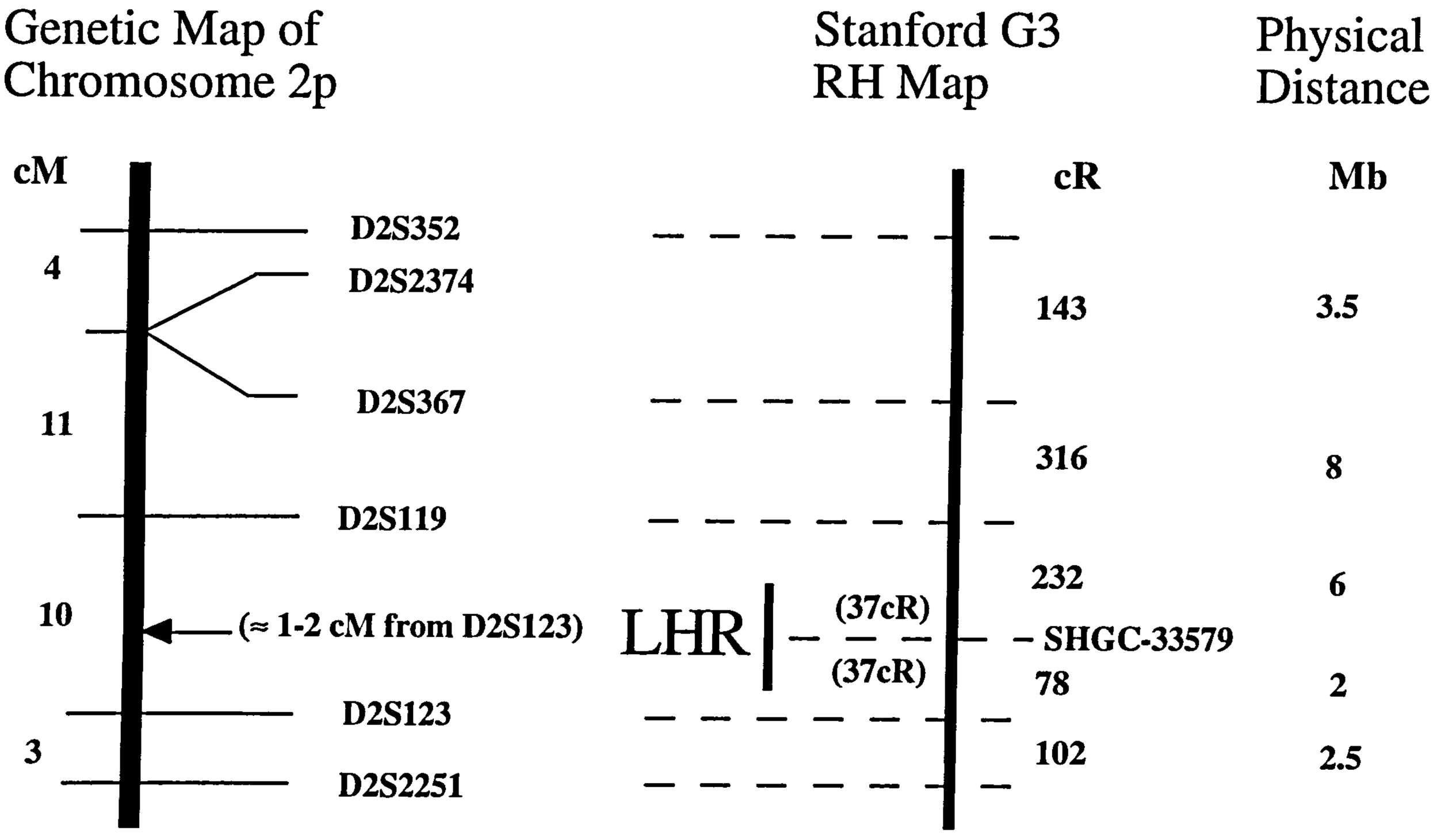
The results of the RH panel screening were scored as a vector, starting from hybrid 1 to hybrid 83 (refer to **section 2.3.10**). The raw scoring data (given below) was then submitted to the radiation hybrid mapping web server at the Stanford Human Genome Center (**section 2.3.10**), where a two point linkage analysis was carried out using the RHMAP Version 2.0 program.

Raw scoring data submitted to radiation hybrid mapping web server at the Stanford Human Genome Center

0100000000000000000000000000000000000000001100000000100000000
00010000000000

A maximum 2 point LOD score of 4.95 was obtained between the *LHR* marker and the marker SHGC-33579 at a distance of 36.72cR. For the G3 RH panel $1cR_{10,000}$ is approximately equal to 25Kb. Breakage probabilities and distance estimates from RH vector analysis were consistent with the *LHR* gene being at a distance of 2 (±1) megabases from marker *D2S123*, 6 (±1) Mb from marker *D2S119*, 14 (±1) Mb from *D2S2374*, 17 (±1) Mb from *D2S352* and 4.5 (±1) Mb from *D2S2251*. The results of the radiation hybrid linkage analysis are given in **Figure 3.2.2.**

Figure 3.2.2 Radiation hybrid mapping of the LHR.



A Maximum 2 point LOD scores of 4.59 was obtained between the LHR marker and the marker SHGC-33579 at a distance of 36.72 cR. Distance is measured in centirays (cR), and for the Stanford G3 Radiation Hybrid Panel 1cR10,000 ≈ 25Kb. Estimated megabase (Mb) distance is calculated from the cR distance obtained. Tle LHR has been mapped to a distance of ±1Mb from marker SHGC-33579. The genetic map of chromosome 2 showing intermarker distances (cM) has been obtained from the 1996 Genethon map (Dib et al., 1996). A rough estimate of the cM distance of the LHR from marker D2S123 was obtained by averaging the physical and genetic distances for this region.

3.2.2. Chromosome 2 Linkage results

Parametric and non-parametric linkage analyses were carried out using the GENEHUNTER program (Kruglyak *et al.*, 1996). The marker information, position and chromosomal order of the markers used in the analysis are given in **section 2.2.1.**

The results of the chromosome 2 linkage analysis were as follows:

- Information on the expected and observed numbers of recombinations is given in **table 3.2.1.** The z-test was used to compare observed and expected marker-marker interval distances. No evidence for map inflation was observed.

Table 3.2.1 Recombination data for the *LHR* locus linkage analysis.

Locus Interval	Recombination Fraction (θ)	Expected cross-overs	Observed cross-overs	z- test	p-value
D2S352 - D2S367	0.038	7.69	6.08	0.377	0.352
D2S367 - D2S2374	0.000	0.00	0.00		
D2S2374 - D2S119	0.099	19.55	22.82	0.252	0.401
D2S119 - D2S123	0.091	17.95	21.37	0.540	0.295
D2S123 - D2S2251	0.029	5.82	7.03	0.300	0.382

Information on the number of expected and observed recombinations for the LHR locus linkage analysis were generated by the GENEHUNTER program and are given above. There were a total of 200 meioses observed. The z-test indicates that there are no significant map inflations in any of the marker-marker intervals. Refer to the legend of table 3.1.1 for abbreviations.

- The results of the linkage analysis are summarised in **table 3.2.2** and **figures 3.2.3-3.3.5.** The pedigree haplotypes for the *LHR* locus are given in **figure 3.3.6.**

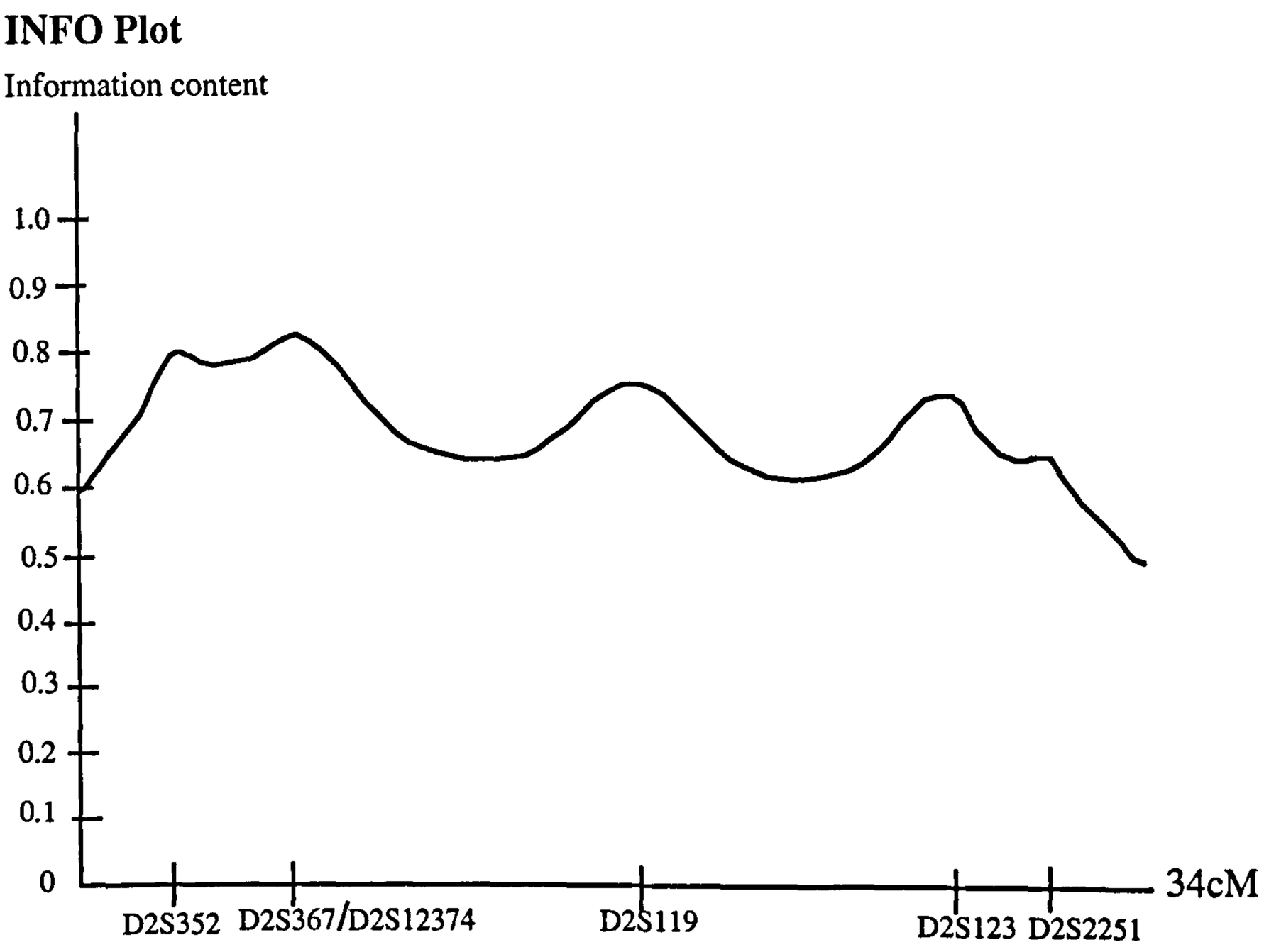
Table 3.2.2 GENEHUNTER results for *LHR* locus.

Marker	Position (cM)	LOD score^R	LOD score^D	NPL score	p-value	info.
	-3.00	-0.08	-2.26	0.63	0.25	0.59
	-2.00	-0.12	-2.57	0.65	0.25	0.65
	-1.00	-0.16	-2.93	0.66	0.24	0.71
D2S352	0.00	-0.21	-3.36	0.68	0.23	0.79
	1.33	-0.24	-3.51	0.70	0.23	0.78
	2.67	-0.30	-3.75	0.73	0.22	0.79
D2S367/ D2S2374	4.00	-0.37	-4.16	0.75	0.22	0.82
	7.67	-0.56	-3.72	0.41	0.32	0.67
	11.33	-1.08	-4.02	0.10	0.44	0.64
D2S119	15.00	-1.84	-5.02	-0.20	0.57	0.75
	18.33	-1.81	-4.60	-0.68	0.75	0.63
	21.67	-2.00	-4.82	-1.15	0.89	0.65
D2S123	25.00	-2.47	-5.79	-1.64	0.96	0.73
	26.00	-2.40	-5.74	-1.54	0.95	0.67
	27.00	-2.34	-5.79	-1.44	0.94	0.64
D2S2251	28.00	-2.30	-5.98	-1.34	0.92	0.64
	29.00	-2.17	-5.43	-1.27	0.91	0.58
	30.00	-2.04	-4.98	-1.23	0.90	0.53
	31.00	-1.93	-4.59	-1.18	0.89	0.49

Refer to section 2.4.3 for GENEHUNTER input parameters used in the analysis and for the key to the output results. The LHR is located in the interval D2S119-LHR-D2S123 at a distance of about 1-2cM from D2S123 (refer to figure 3.2.2). Under the dominant parametric analysis, a negative LOD score^D of -2 or less has been obtained across the whole region. Under the recessive parametric analysis only the region distal to position 21.67cM, which encompasses the LHR locus, has been excluded by a LOD score^R of -2 or less. The LOD score^R for the rest of the map remains inconclusive. Non-parametric analysis provides no evidence for excess allele sharing. There is zero genetic distance between the D2S367 and D2S2374 markers.

- Information content is high at markers *D2S352* (0.79), *D2S367/D2S2374* (0.82), *D2S119* (0.75) and *D2S123* (0.73). The lower information content at marker *D2S2251* (0.64), is partly due to the lower observed heterozygosity for this marker in our data set. Information at this position may be increased by typing another marker close to *D2S2251*. However, since the results have provided sufficient information at the *LHR* locus (which is in the interval *D2S119-LHR-D2S123*, at a distance of about 1-2cM from *D2S123*), this was not considered essential for the purposes of this study.

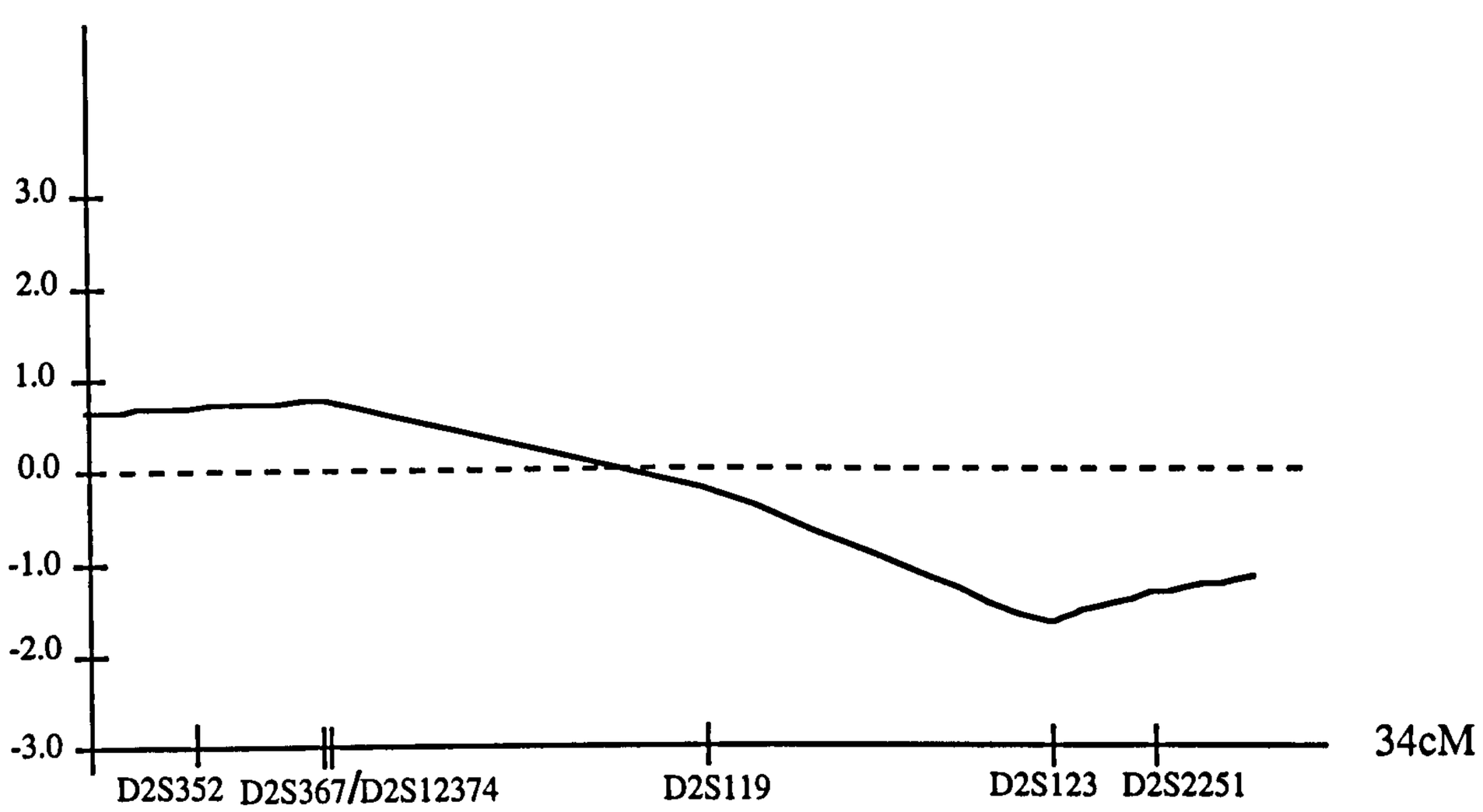
Figure 3.2.3 Information plot for the chromosome 2 linkage analysis.



Information plot is a graphic representation of the information extracted at each point along the linear map (see table 3.2.2). Information content is generally high at all markers.

- There was no evidence for excess allele sharing (non-parametric analysis) across the entire region.

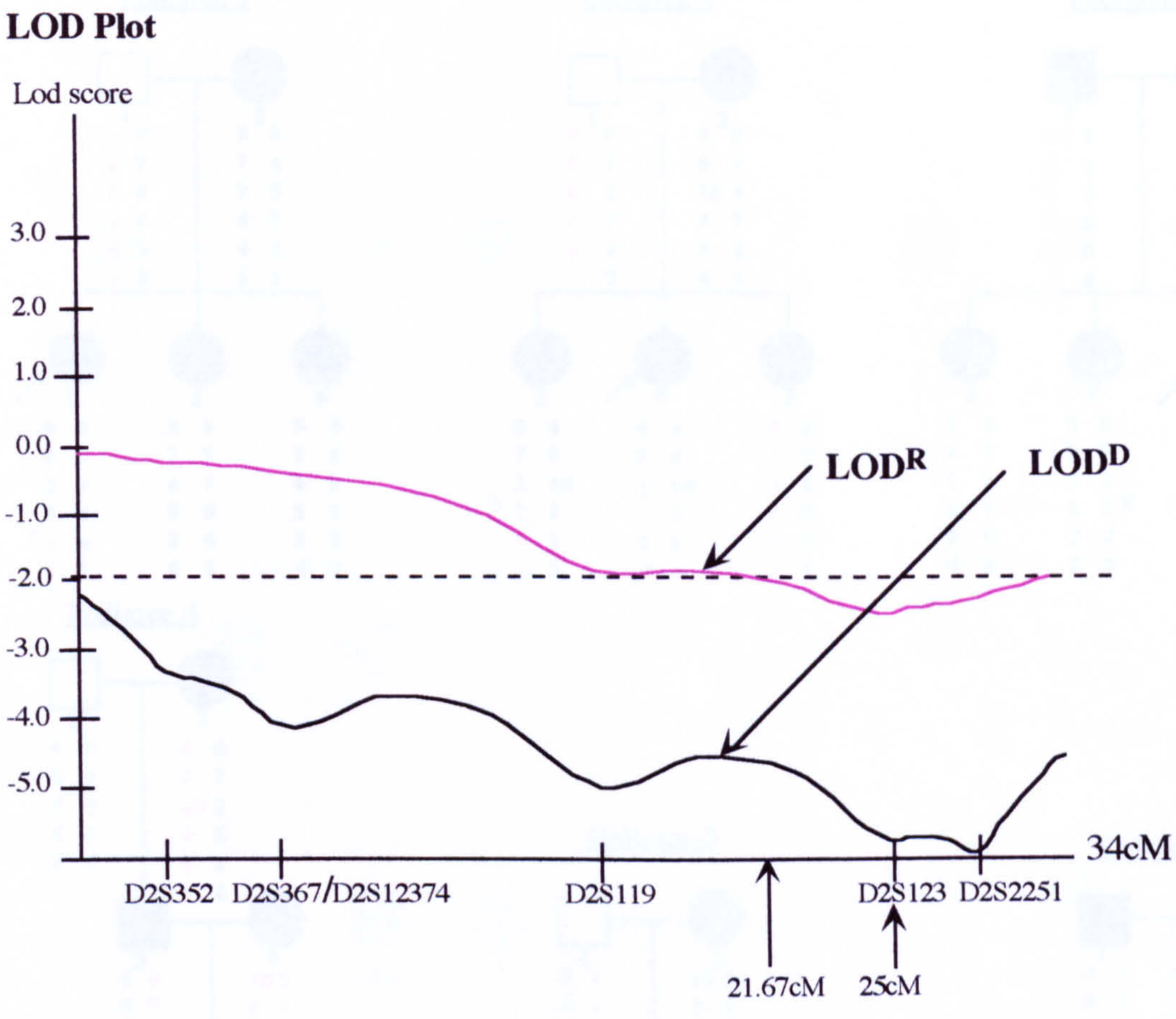
Figure 3.2.4 NPL plot for the chromosome 2 linkage analysis.



NPL plot is a graphic representation of the non-parametric linkage results (refer to table 3.2.2). The dashed line represents the mean of expected allele sharing distribution. There is no evidence for excess allele sharing across the entire map.

- Under the dominant parametric analysis a LOD score^{D} of -2 or less was obtained across the whole region.
- The recessive parametric analysis has also provided a LOD score^{R} of -2 or less for the region distal to position 21.67cM (which encompasses the *LHR* locus). However, the LOD score^{R} obtained for the region proximal to this position is inconclusive (i.e. does not achieve criteria for proof or rejection of linkage).

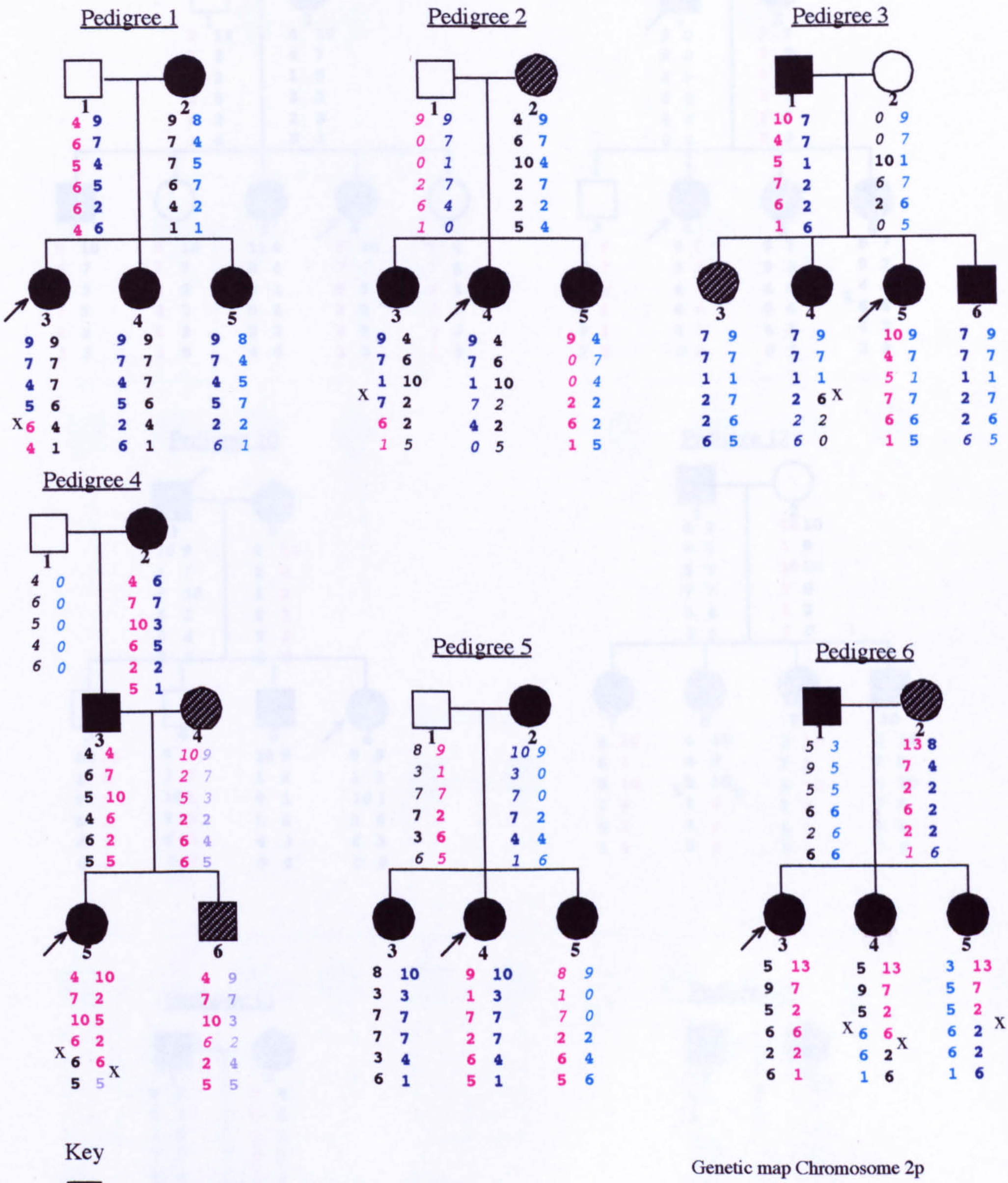
Figure 3.2.5 LOD plots for the chromosome 2 linkage analysis.



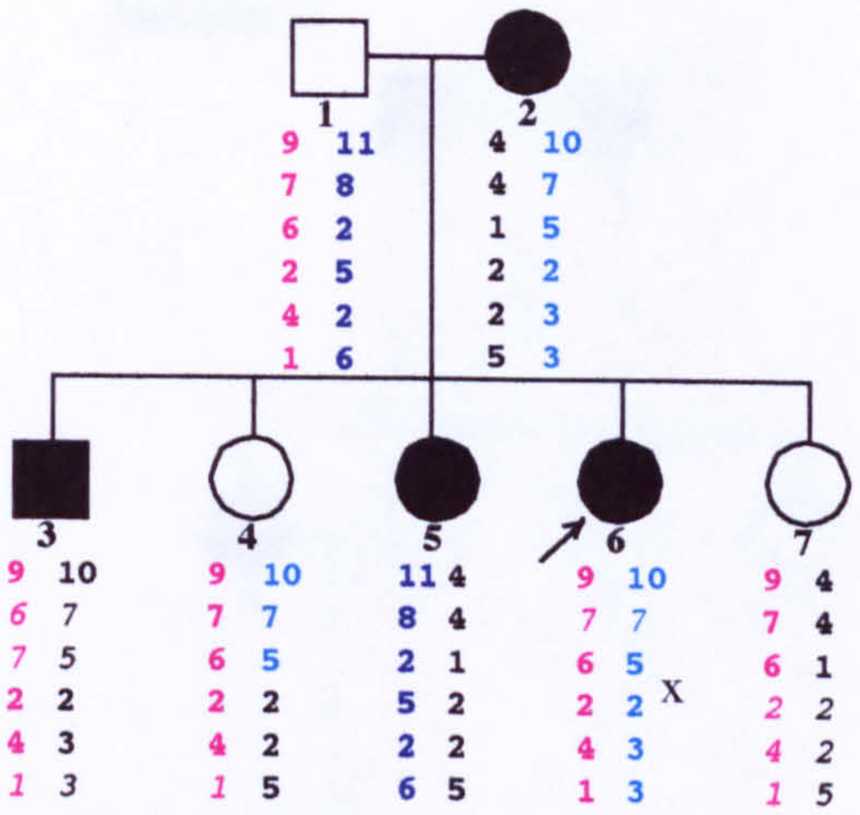
The above LOD plots are graphic presentations of the parametric linkage results (see table 3.2.2). The black and red curves represent the parametric linkage analysis under the dominant and recessive models respectively. The dashed line at -2 represents the minimum criterion required for rejection of linkage (section 2.4.3). Under the dominant analysis the entire region has been excluded. The recessive analysis has excluded the region distal to position 21.67cM (which encompasses the LHR locus, which is at a distance of 1-2cM proximal to D2S123). However, the LOD score^{R}s obtained for the region proximal to this position, do not achieve criteria for proof or rejection of linkage.

Overall, these results indicate that the *LHR* gene is not a major susceptibility locus in the aetiology of PCOS/MPB.

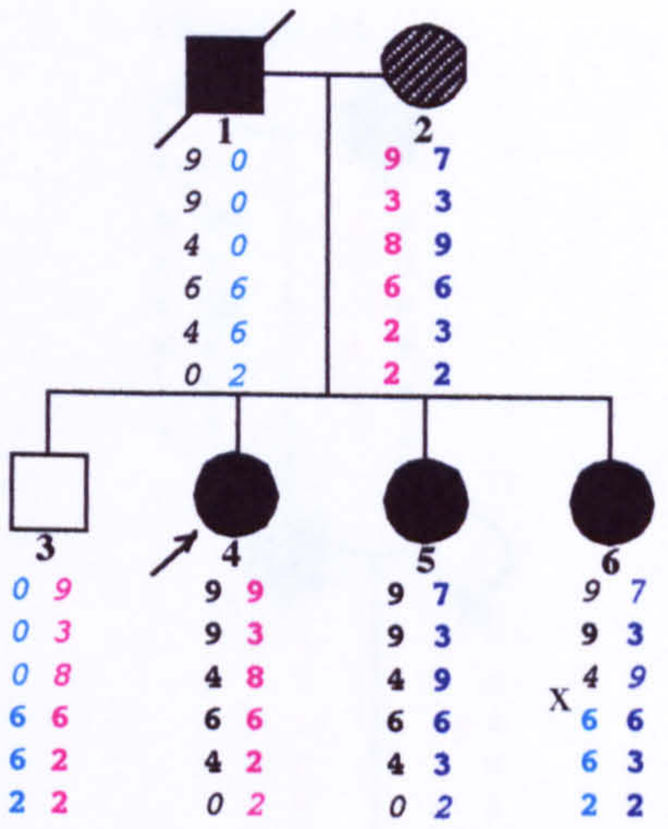
Figure 3.2.6 *LHR* locus haplotypes in the PCOS/MPB pedigrees



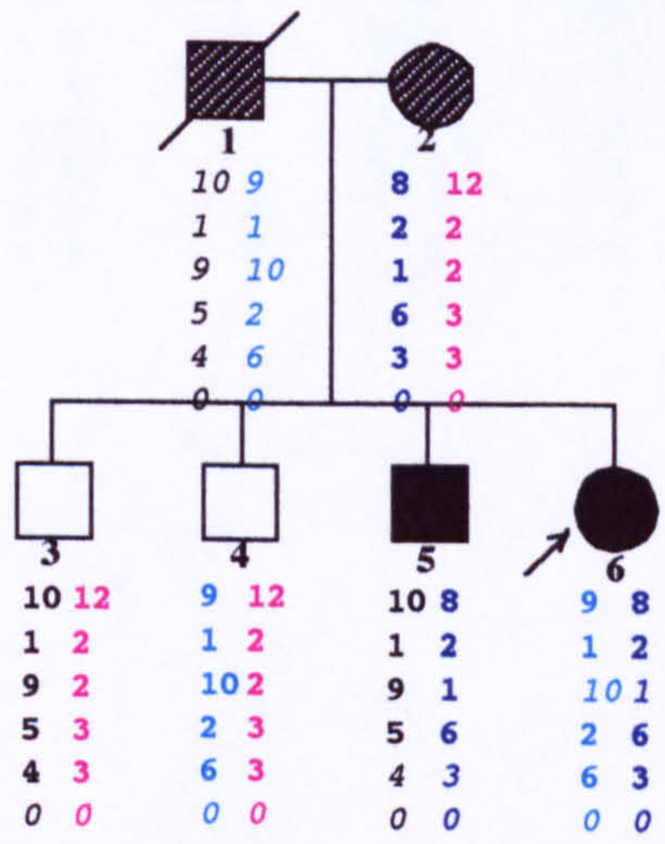
Pedigree 7



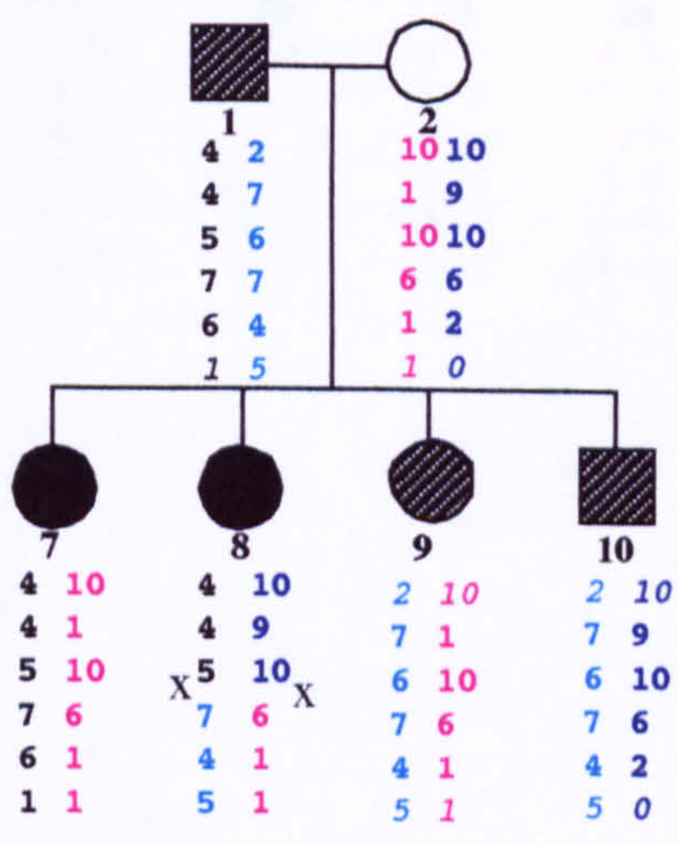
Pedigree 9



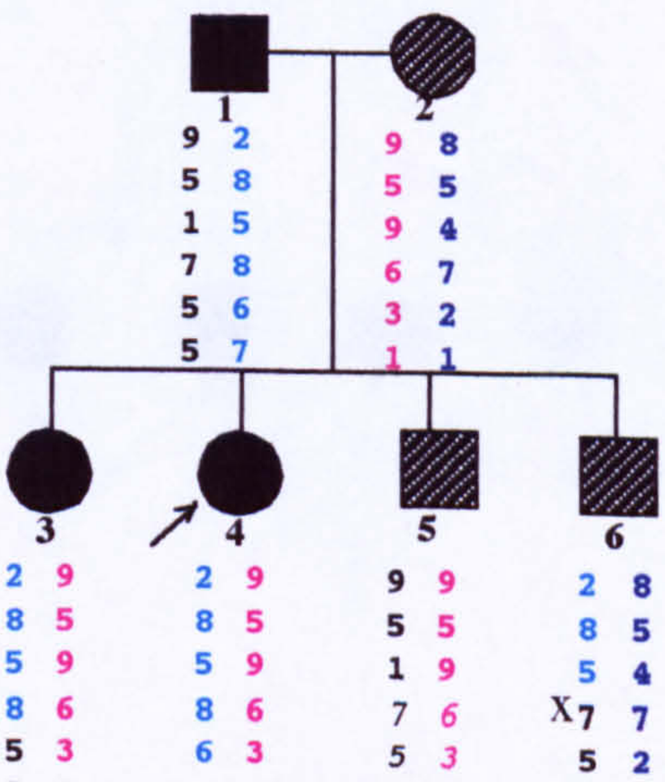
Pedigree 10



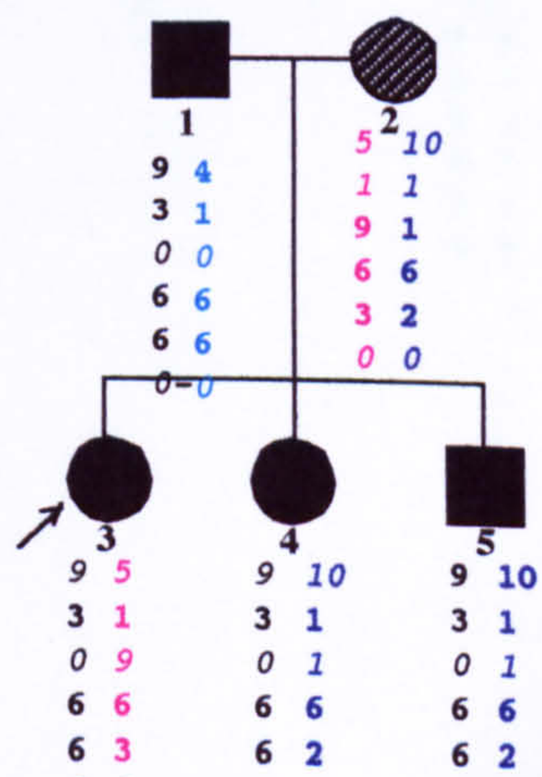
Pedigree 12



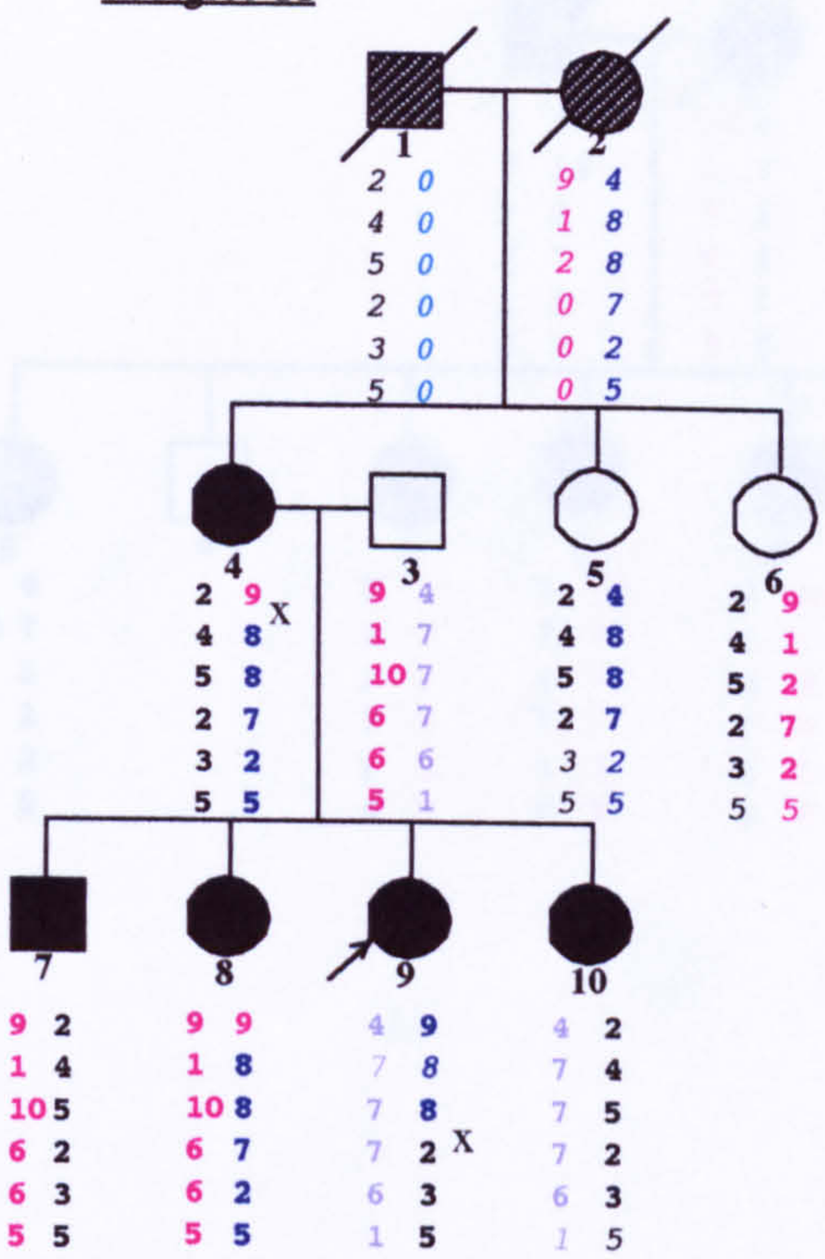
Pedigree 13



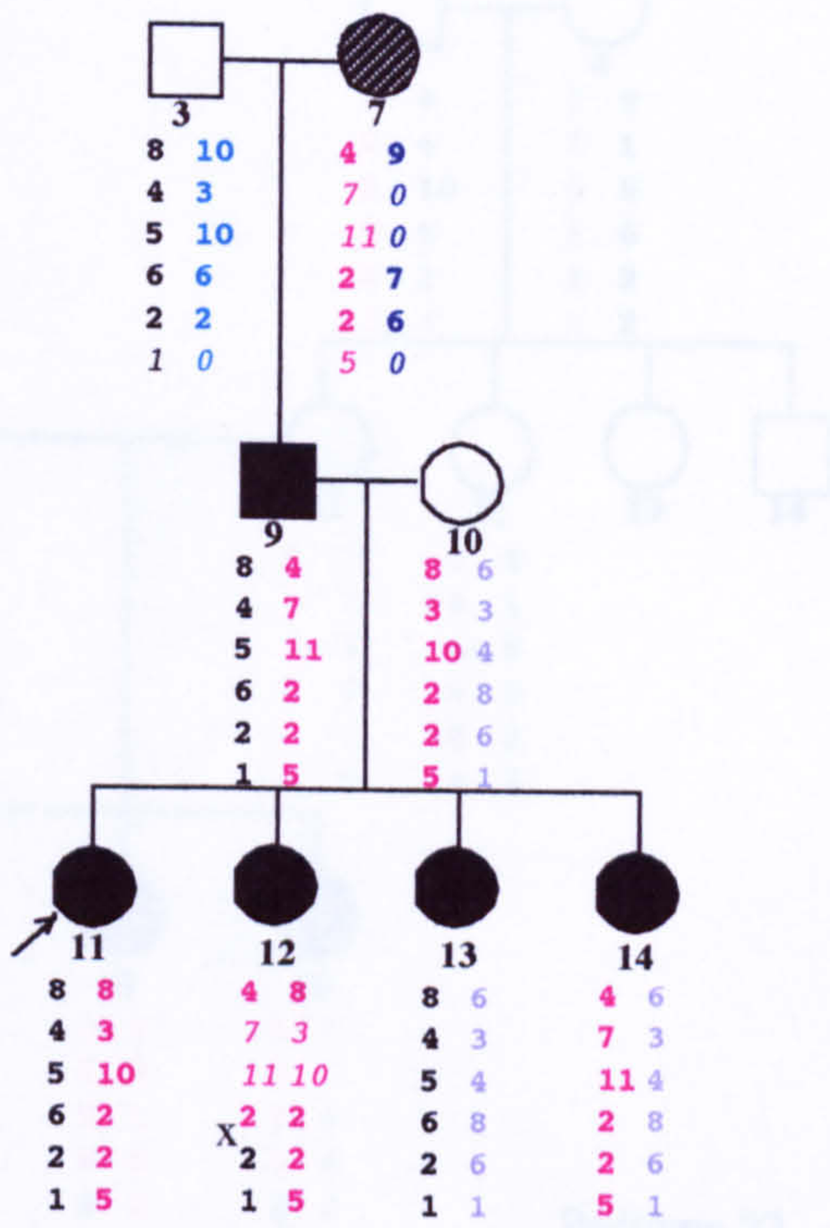
Pedigree 16



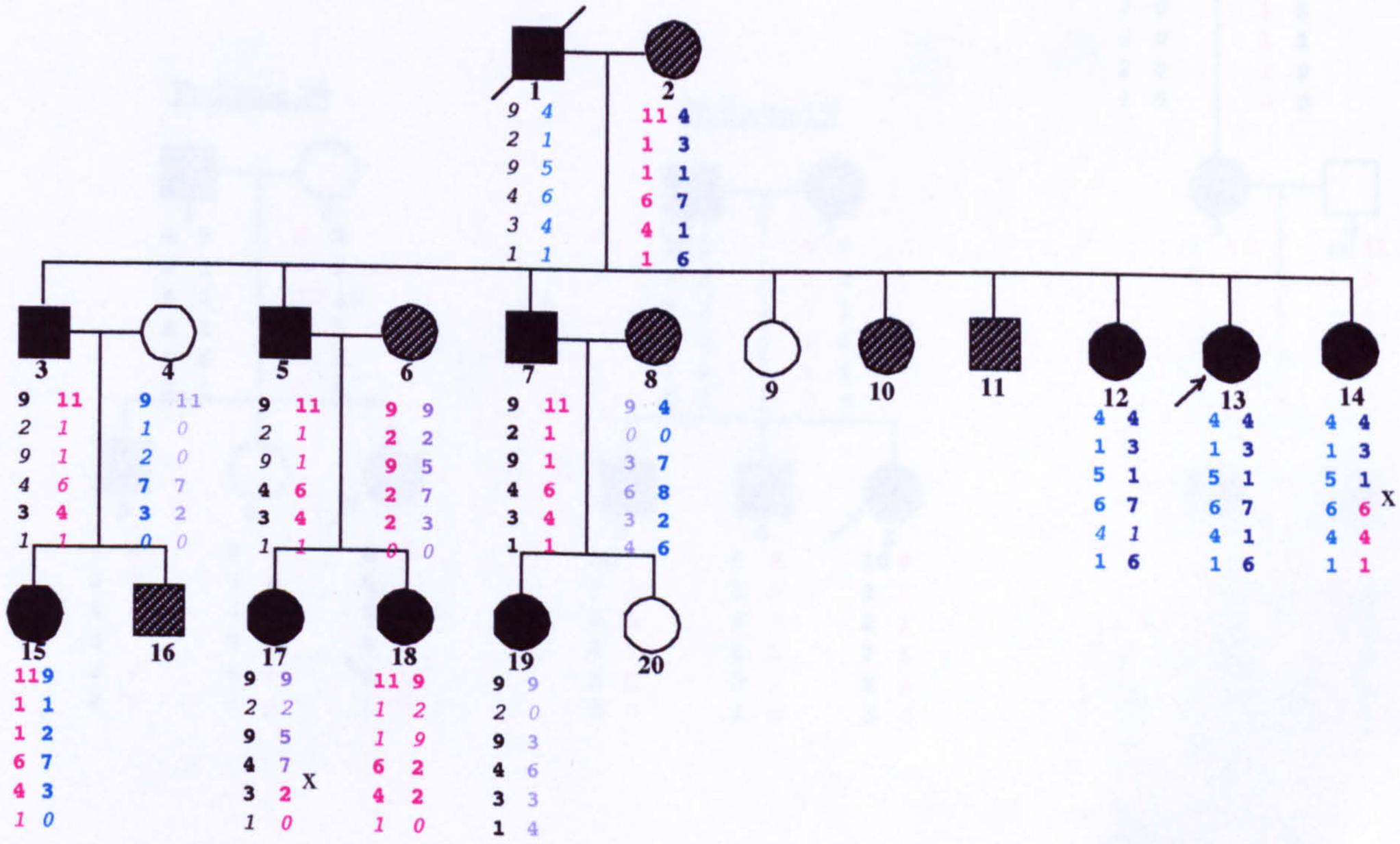
Pedigree 11



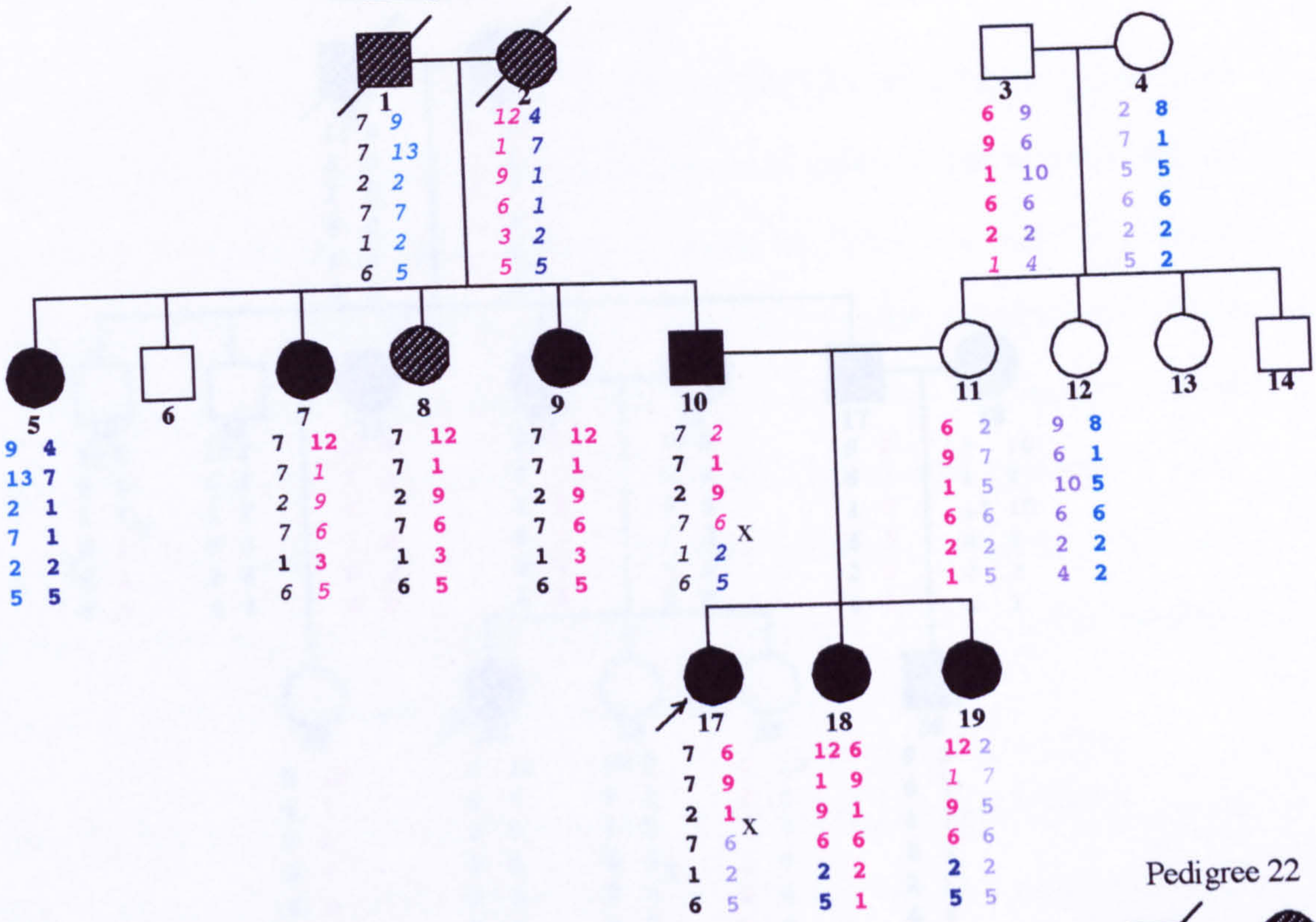
Pedigree 14



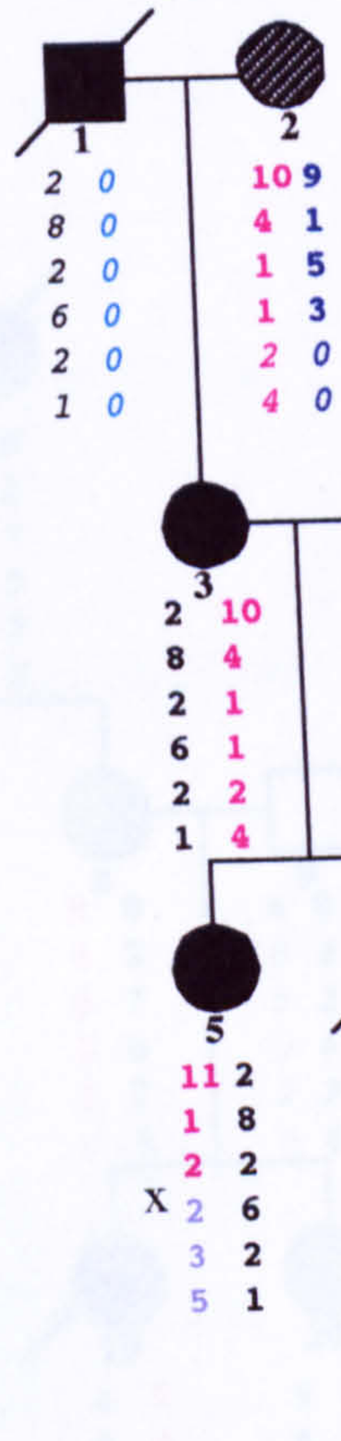
Pedigree 17



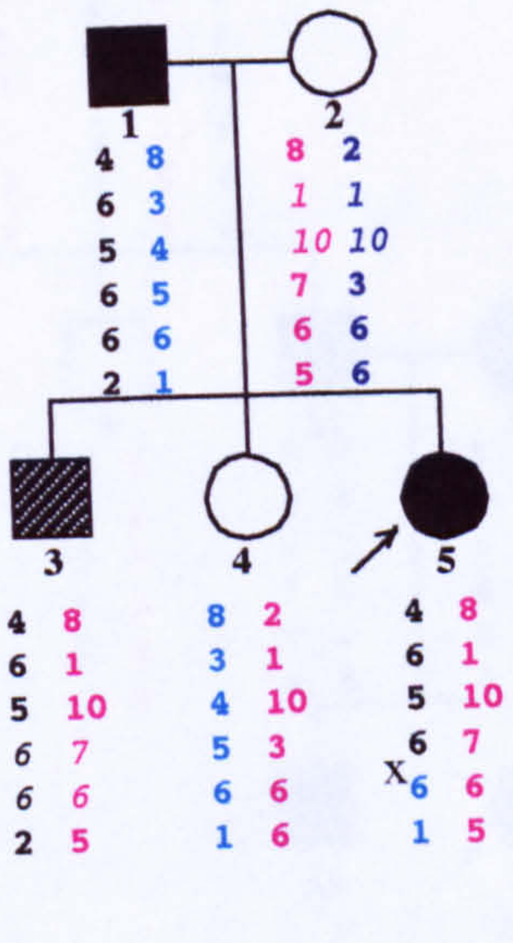
Pedigree 20



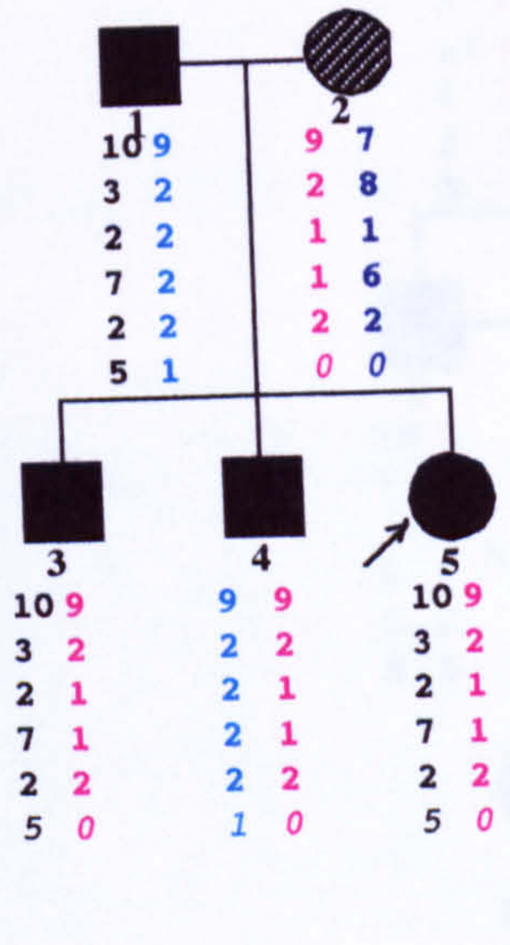
Pedigree 22



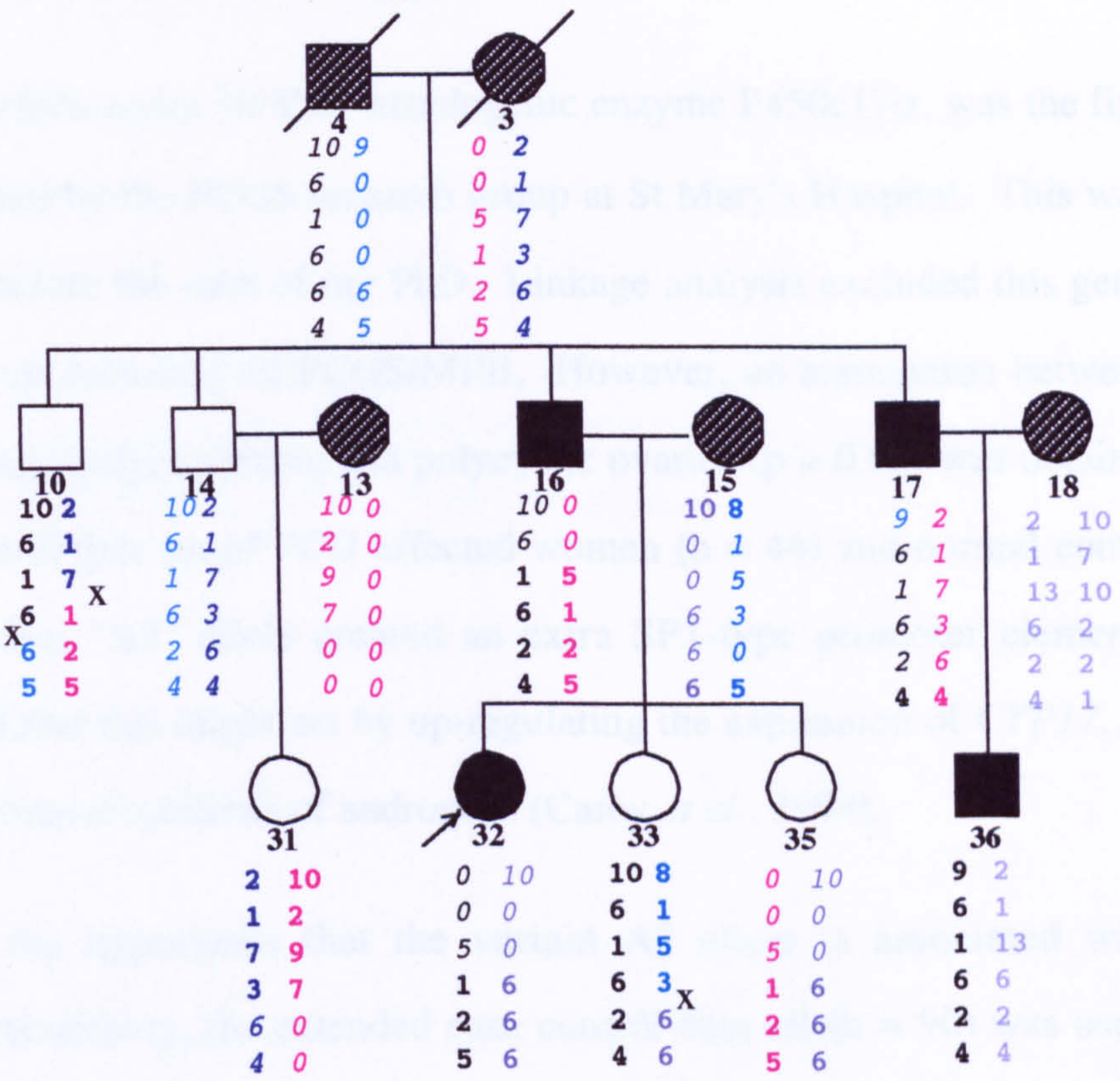
Pedigree 18



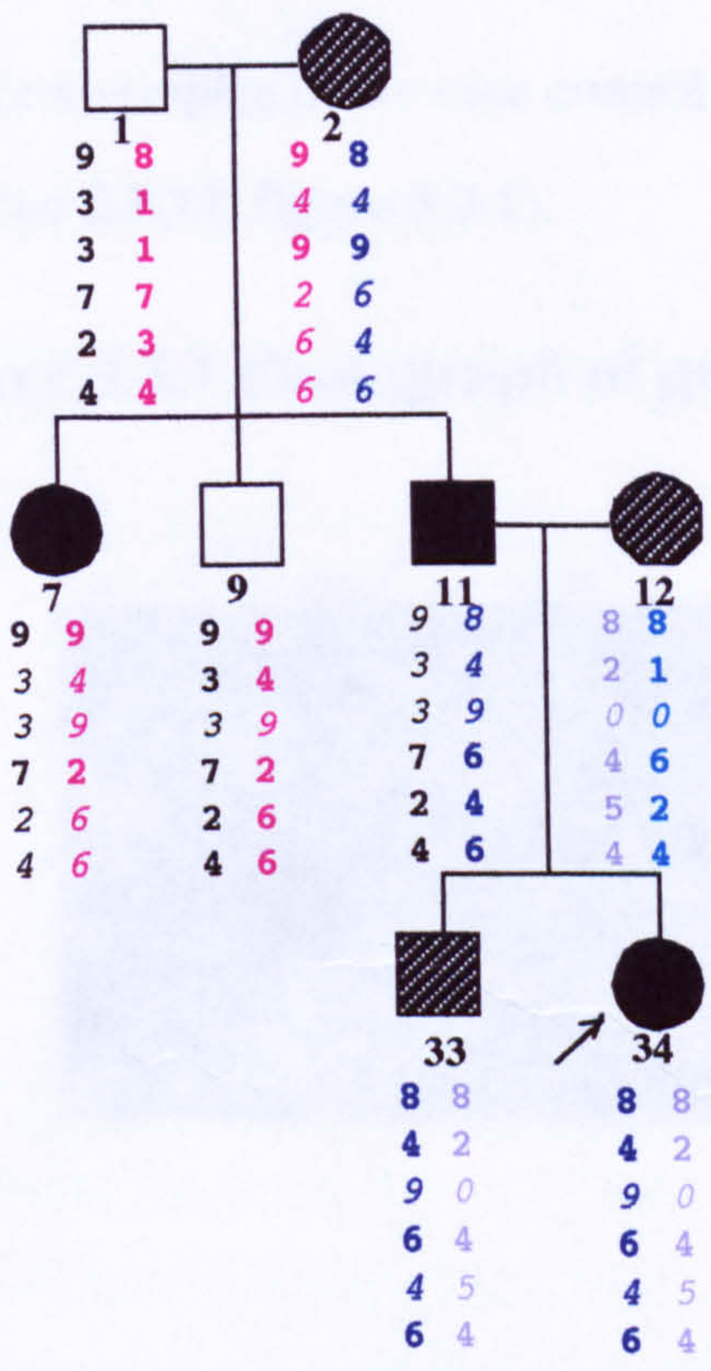
Pedigree 19



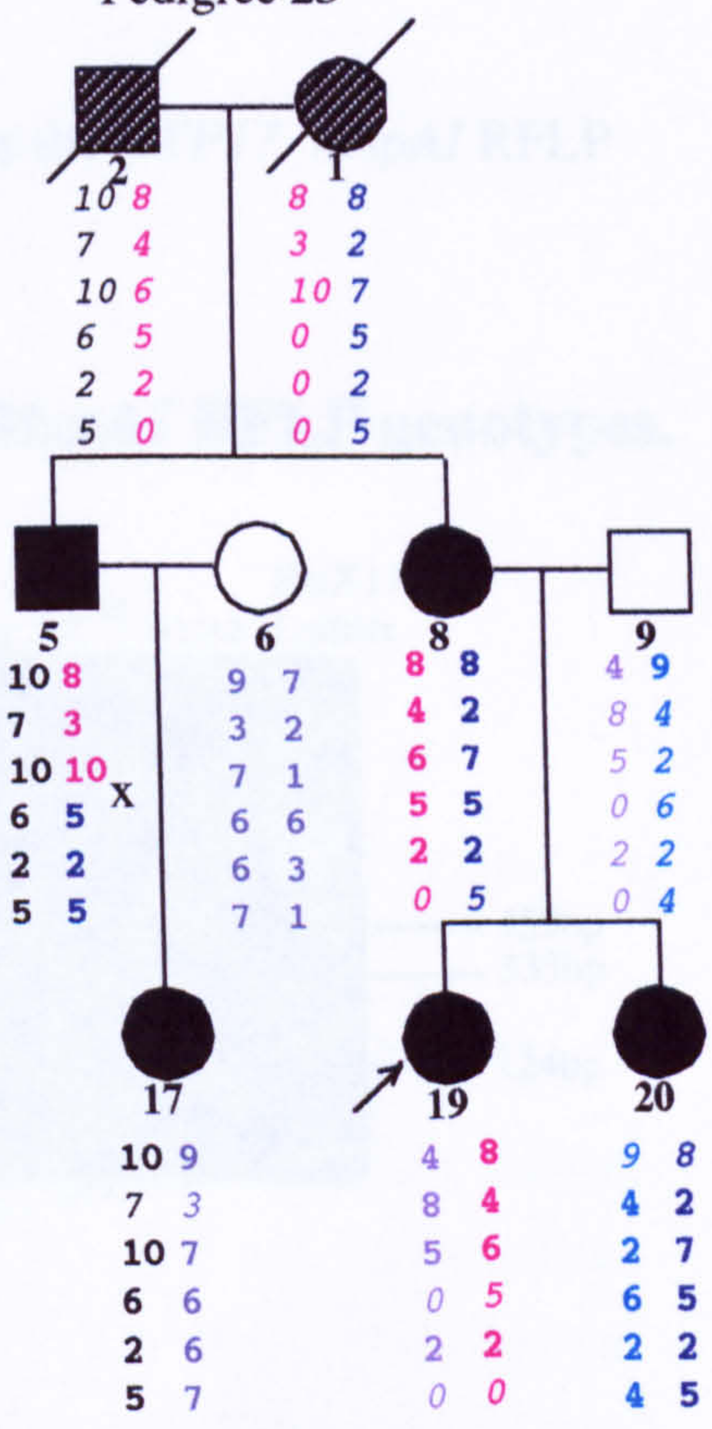
Pedigree 23



Pedigree 24



Pedigree 25



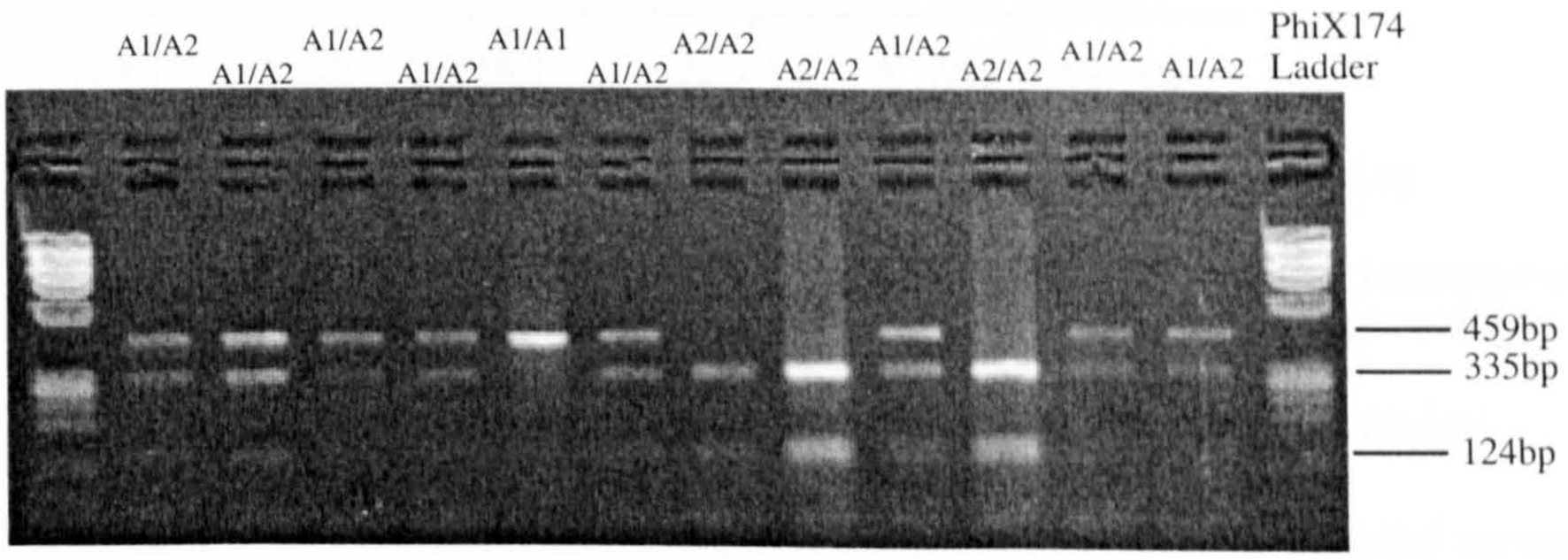
3. *CYP17* association study

CYP17, which codes for the steroidogenic enzyme P450c17α, was the first gene to be investigated by the PCOS research group at St Mary's Hospital. This was carried out in 1994 before the start of my PhD. Linkage analysis excluded this gene as a major locus in the aetiology of PCOS/MPB. However, an association between a 5' single base change polymorphism and polycystic ovaries (p = 0.03) was obtained in a small case control data set of PCO affected women (n = 44) and normal controls (n =24). This variant "A2" allele created an extra SP1-type promoter element and it was proposed that this might act by up-regulating the expression of *CYP17*, thus resulting in an increased synthesis of androgens (Carey *et al.*, 1994).

To test the hypothesis that the variant A2 allele is associated with increased P450c17α activity, the extended case control data set (n = 96) was used to examine the association of the A2 allele with serum testosterone levels (Gharani *et al.*, 1996). These results have been further up-dated and are presented below.

The new samples in the case control data set were typed for the *CYP17 MspAI* RFLP (**section 2.3.12**, **figure 3.3.1**).

Figure 3.3.1 Photograph of gel showing *CYP17 MspAI* RFLP genotypes.

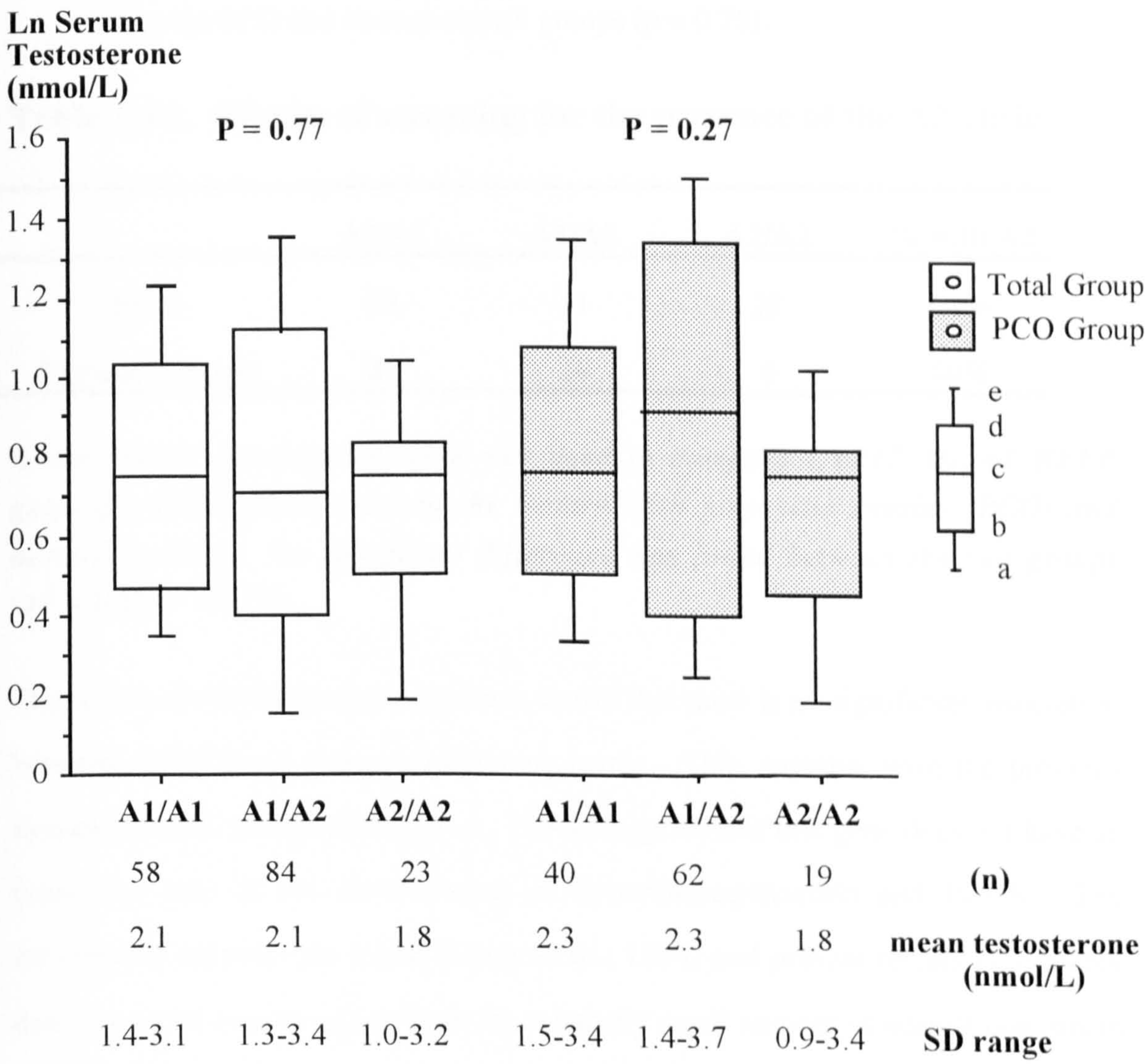


- All individuals (affecteds and controls) were allocated to three groups according to their genotypes, A1/A1, A1/A2 and A2/A2. The Kruskal Wallis test was used

to compare the geometric means of total serum testosterone levels in the different genotype groups (**figures 3.3.2**).

- No significant difference was found between the means of the genotype groups either when the entire data set was examined (p = 0.77), or when women with polycystic ovaries were considered alone (p = 0.27).

Figure 3.3.2 Comparison of the geometric mean of total serum testosterone levels in the three *CYP17* genotype groups for the "Total group" and for the "PCO group" alone.



The Kruskal Wallis test was used to compare the geometric means of total serum testosterone levels between the three CYP17 genotype groups. The p-values obtained for both analyses were shown to be not significant and are given above. "a", "b", "c", "d" and "e" are the 10th, 25th, 50th, 75th and 90th percentiles of serum testosterone levels respectively. The mean and standard deviation (SD) ranges of serum testosterone levels for each group are also given above.

- Consequently, this extended data set was used to re-evaluate the association of the A2 allele with PCO. Previously, 74% of 44 PCO cases were found to have the A2 allele (A1/A2 or A2/A2) compared to 38% of 24 normal controls (Carey *et al.*, 1994).

- A chi square contingency table was used to test for any significant difference in genotype distribution between the groups. The results are summarised in **table 3.3.1.** There was no longer a significant difference in the genotype distributions between the PCO and normal control groups ($p = 0.75$).

Table 3.3.1. Results of screening for the presence of the A2 allele.

	A1/A1	A1/A2	A2/A2	% with A2
PCO	**59**	**81**	**25**	**64%**
Normal control	**21**	**26**	**6**	**60%**

A chi square contingency table was used to compare CYP17 MspA1 RFLP genotype distributions between the women with polycystic ovaries (PCO) and normal controls. No significant difference was found between the two groups ***($\chi^2 = 0.58, p = 0.75$).***

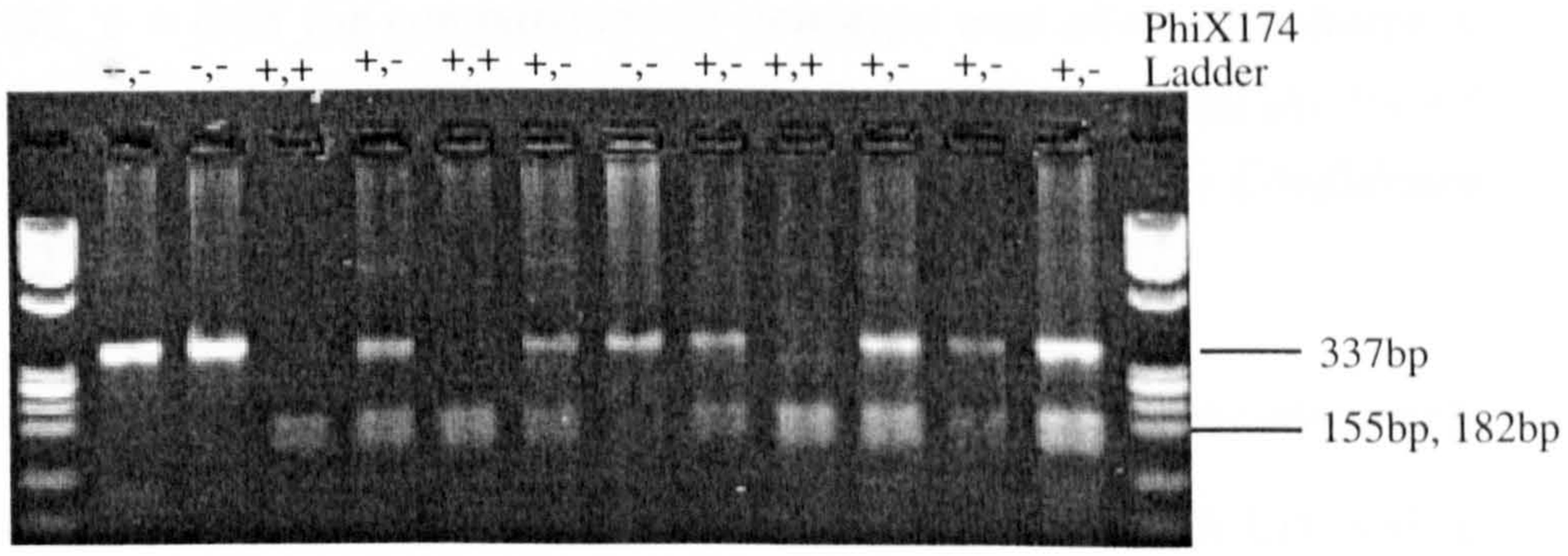
The results of this follow-up study have shown that there is no significant association between *CYP17* and serum testosterone levels. This, together with the previous linkage analysis results (Carey *et al.*, 1994), suggests that this gene does not have an important role in the development of hyperandrogenaemia and PCOS. The discrepancy between the initial (Carey *et al.*, 1994) and present results is probably due to a type I error resulting from the relatively small number of normal controls in the original data set. Therefore, it may be concluded that this base-pair change identified in the promoter region of the *CYP17* gene is a common polymorphism with no obvious role in the aetiology of PCOS or hyperandrogenaemia.

3.4 *SRD5A1* association results

In the following section the role of the gene *SRD5A1*, which codes for the enzyme 5 alpha-reductase type 1 has been investigated as a candidate for involvement in the development of hirsutism.

Two single base change polymorphisms have previously been identified in the *SRD5A1* gene, a *Hinf I* RFLP in exon 1 and a *Nsp I* RFLP in exon 2 (Jenkins *et al.*, 1991). Both polymorphisms occur in the third position of a codon and neither one alters the amino acid sequence of the protein (**section 2.3.13**). Individuals in the case control data set who had a definite assignment of either hirsute or non-hirsute (**section 2.1.2**) were typed for the two *SRD5A1* RFLPs (**section 2.3.13**), and comparisons were made between the genotype and allele distributions in the two groups. The first RFLP to be evaluated was the *Nsp1* RFLP (**figure 3.4.1**, **table 3.4.1**).

Figure 3.4.1 Photograph of gel showing *SRD5A1 Nsp I* RFLP genotypes.



The results of the *NspI* RFLP analysis have shown:

- A significant difference in allele distributions (p = 0.05) between the hirsute and non-hirsute groups (**table 3.4.1**) was observed. There was also a significant odds ratio of 2.4 for being hirsute given the "+,+" genotype (p = 0.02 and $P_c = 0.04$, after a Bonferroni correction was applied (since a similar comparison could be made for the "-,-" genotypes).

Table. 3.4.1 *SRD5A1 Nsp I* RFLP genotype distributions in the hirsute and non-hirsute groups.

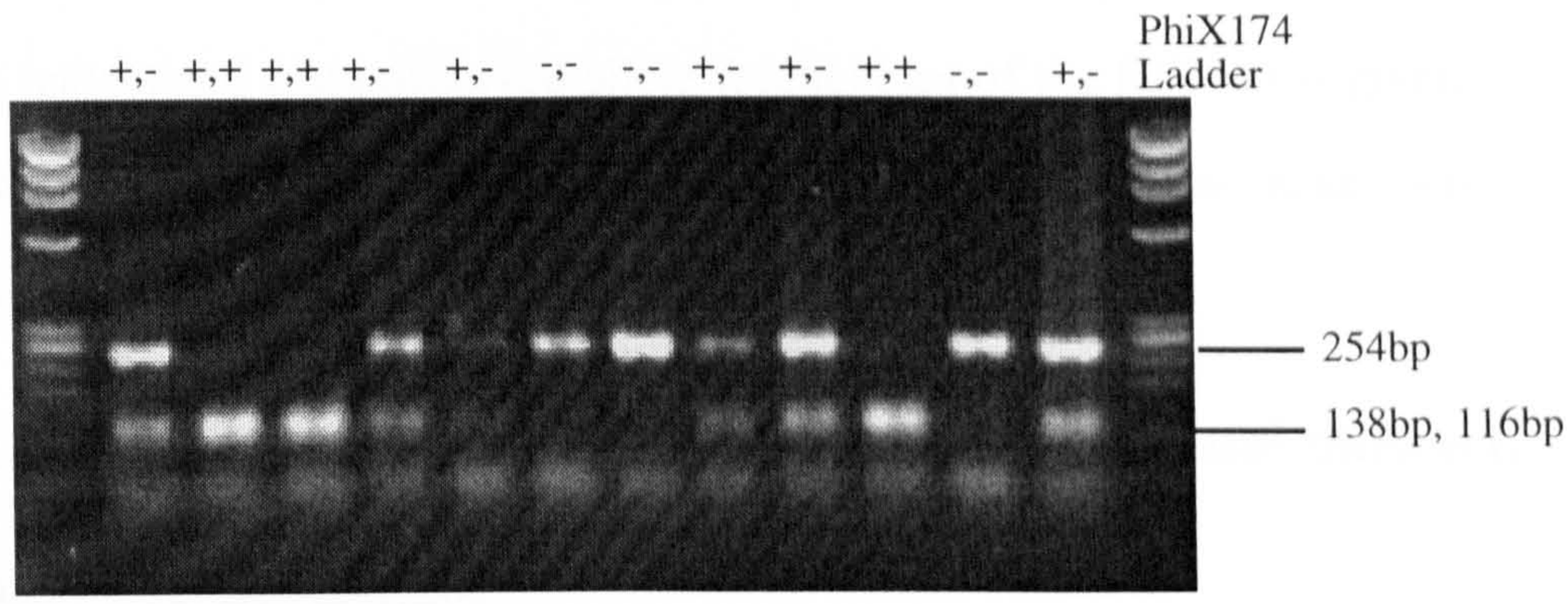
	NspI genotypes		
	-,-	+,-	+,+
Hirsute	29	31	18
Non-Hirsute	53	59	14

A Chi square contingency table was used to compare genotype and allele distributions between the hirsute and non-hirsute group ($\chi^2 = 5.23$, p = 0.07 and $\chi^2 = 2.91$, p = 0.05 for comparisons of genotype and allele distributions respectively). A significant odds ratio of 2.4 for being hirsute given the "+,+" genotype was obtained (Fisher's exact p = 0.02 ($P_c = 0.04$), 95% Confidence Intervals 1.12 - 5.16).

- These results suggest that the *Nsp I* RFLP "+" allele is associated with hirsutism, either by having a direct functional role or through LD with a putative *SRD5A1* "mutation" located elsewhere in the gene.

Subsequently, the women in this data set were typed for the *Hinf I* RFLP (**section 2.3.13**, **figure 3.4.2**) and a comparison between genotype distributions in the hirsute and non-hirsute groups was made (**table 3.4.2**).

Figure 3.4.2 Photograph of gel showing *SRD5A1 Hinf I* RFLP genotypes.



- No significant difference in genotype distributions between the two groups was observed (**table 3.4.2**). These results indicate that the *SRD5A1 Hinf I* RFLP is not associated with hirsutism.

Table. 3.4.2 *SRD5A1 Hinf I* RFLP genotype distributions in the hirsute and non-hirsute groups.

	Hinf I genotypes		
	-,-	+,-	+,+
Hirsute	**10**	**24**	**18**
Non-Hirsute	**24**	**59**	**31**

A Chi square contingency table was used to compare genotype distributions between the hirsute and non-hirsute groups ($\chi^2 = 0.95$, $p = 0.62$). ***No significant difference was observed. The smaller number of individuals typed for the HinfI RFLP compared to the NspI RFLP (table 3.4.1) is due to a number of "dropouts" (older samples of poor quality DNA).***

To further investigate the discrepancy between the association results for the two *SRD5A1* RFLPs, haplotype analysis was also carried out. Initially, the frequency of the four possible haplotypes in the total data set of women (hirsutes and non-hirsutes) who were typed for both RFLPs was established (excluding the double heterozygotes for whom haplotype status could not be ascertained). The haplotype distribution observed in the total data set and the expected haplotype distribution (calculated from observed allele frequencies of the RFLPs) is given below in **table 3.4.3.** A test for the presence of LD between these two polymorphic loci was then carried out (**section 2.4.9**).

Table 3.4.3 *Hinf I/Nsp I* haplotype distribution in the total data set.

Hinf I	*Nsp I*	Observed Haplotype no.	Expected Haplotype no.
+	+	84	49
-	+	5	39
-	-	128	120
+	-	85	94
Total number of haplotypes		**302**	

"Observed haplotype numbers" were ascertained from the total data set of hirsute and non-hirsute women for whom genotype information for both RFLP loci was available (minus the double heterozygotes). The "Expected Haplotype numbers" were calculated from the observed allele frequencies of the two RFLPs. The allele frequencies were obtained from the total data set (hirsutes and controls) typed for each RFLP (i.e. 260 chromosomes for the Nsp I and 176 chromosomes for the Hinf I RFLPs). The allele frequencies were found to be Nsp I "+" (0.29), Nsp I "-" (0.71), Hinf I "+" (0.56) and Hinf I "-" (0.44). Assuming that the two loci are in linkage equilibrium in the population then haplotype frequencies can be derived as the product of the allele frequencies of the two loci. The LD test (section 2.4.9) has demonstrated LD between the "+" alleles of the Hinf I and Nsp I loci (Δ = 0.45, Chi square = 27.1, p = < 0.0001).

- Significant evidence for LD between the *Hinf I* and the *Nsp I* RFLPs was observed ($\Delta = 0.45$, Chi square = 27.1 with a p-value of < 0.0001).

- The *NspI* "+" allele (which is associated with hirsutism) is in LD with the "+" allele of the *Hinf I* RFLP, such that, of the 89 haplotypes observed with a *NspI* "+" allele, only five haplotypes show co-inheritance of a *Hinf I* "-" allele (which is considerably less than the number expected under the null hypothesis of no LD).

Subsequently, the haplotype distributions were compared between the hirsute and the non-hirsute groups (**table 3.4.4**).

- A significant difference in haplotype frequencies ($p = 0.009$) was obtained. There were significantly more "*HinfI(+)/NspI*(+)" haplotypes in the hirsute (38%) compared to the non-hirsute group (24%).

Table 3.4.4 *Hinf I/Nsp I* haplotype distributions in the hirsute and non-hirsute groups.

		Observed Haplotype no.	
Hinf I	*Nsp I*	Hirsute	Non-Hirsute
+	+	30	37
-	+	3	0
-	-	27	71
+	-	20	46
Total number of haplotypes		80	154

A chi square contingency table was used to compare haplotype distributions between the hirsute and non-hirsute groups ($\chi^2 = 11.5$, $p = 0.009$). There are significantly more "HinfI(+)/NspI(+)" haplotypes in the hirsute (38%) compared to the non-hirsute group (24%).

In summary, the results of this preliminary study have shown association between the "+" allele of the *SRD5A1* exon 2 *Nsp I* RFLP and hirsutism. This suggests that variation in *SRD5A1* plays a role in the aetiology of hirsutism. This polymorphism does not alter the amino acid sequence of the protein and is hence unlikely to be functional. However, examples of apparently silent mutations affecting mRNA splicing and stability have previously been reported in other genes (Liu *et al.*, 1997a; Milland *et al.*, 1996). Therefore, a functional role for this polymorphism cannot be ruled out. Alternatively, this polymorphism may be in LD with a putative aetiological mutation located elsewhere in the gene.

Haplotype analysis suggests that the aetiological mutation is located on *HinfI*(+)/*NspI*(+) chromosomes. Such a mutation may act by upregulating the activity of 5 alpha-reductase type 1 enzyme, leading to an increased production of DHT in the skin.

4.0 DISCUSSION

4.1 EVALUATION OF LINKAGE RESULTS

The ability to detect a disease-susceptibility locus depends on the contribution the locus makes to the genetic variation of the trait. This may be measured in terms of the relative risk ratio, λ_R, defined as the recurrence risk for a relative of an affected person (where the subscript R denotes the type of relation) divided by the risk for the general population (Risch, 1990a). The estimated sibling relative risk λs (risk to a sibling of an affected proband) ranges from about 4 for symptomatic PCOS (based on a sibling recurrence rate of about 40% and a 10% population prevalence) to a λs of 2.5 for PCO morphology (sibling recurrence rate of 50% and 20% PCO prevalence). This λs is the overall risk ratio of the collective effect of all the disease loci. In general, the higher the λs, the greater the genetic effect. The power to detect individual loci contributing to the overall λs will depend on the genetic architecture of PCOS (i.e. the number of susceptibility loci involved and how they interact with each other), which is not known. The power to detect linkage will also depend on the methodological approach taken as well as on the size and structure (i.e. method of ascertainment) of the data set used.

4.1.1 Assessment of the PCOS families

In this study twenty three PCOS pedigrees have been used to carry out linkage analysis. These pedigrees vary in size from small nuclear (4 individuals) to large extended families with over twenty members. All the pedigrees were ascertained on the basis of a symptomatic proband who presented with either anovulatory infertility and or hirsutism and who therefore had polycystic ovary syndrome. Pedigrees were selected from probands who had at least one affected sister, but in most cases there was more than one. Given the high sibling recurrence rate, sparsely affected pedigrees will be fairly rare and, depending on the true disease phenocopy rate, may

have an increased chance of representing non-genetic cases (McCarthy *et al.*, 1998). Therefore, the selection of families with multiple affected members is useful as it will increase the likelihood that there is a genetic basis for PCOS within the families.

The price that is paid for this type of ascertainment bias is that it may lead to inaccurate specification of segregation parameters which could reduce the power of subsequent parametric linkage analysis (**section 1.8.1**). The increased chance of genetic heterogeneity and parental homozygosity within such pedigrees may also reduce the power. Some compensation may be made for this by allowing for phenocopies in the parametric analysis and an arbitrary 10% phenocopy rate has been used for this reason. In general, it is impossible to quantify the positive or negative effects (in terms of power to detect loci) associated with any particular ascertainment scheme. However, the use of non-parametric linkage analysis provides some insurance against the most negative effects when compared to a parametric analysis performed under the wrong model.

Another consideration with these PCOS pedigrees is that the exact relationship between MPB and PCOS is not clear. Most studies have suggested that MPB may be the male phenotype of PCOS and indeed, in our data set, we have identified probable obligate carriers with MPB (e.g. individual 9 in pedigree 20 (see **figure 2.1**)). There is also evidence for both male-to-female (e.g. pedigrees 20 and 16) and female-to-male transmission (e.g. pedigrees 7 and 11). An additional support for a causative link between PCO and MPB is that both are likely to have an underlying androgenic aetiology. However, what is not clear, is whether all PCO susceptibility genes will also predispose to MPB and vice versa. Since most of these pedigrees were ascertained on the basis of having affected sisters, there are relatively few affected men in this data set (only 28 compared to 64 affected women), and the phenocopy rate and reduced penetrance used for the parametric models are likely to make some allowances for these ambiguities.

Simulation studies aimed at estimating the size of the data set (sibpairs) required for the study of PCOS have shown great variation depending on the underlying model tested (McCarthy *et al.*, 1998). Although it is hard to estimate the power of the pedigree data set used in the linkage studies in this thesis, it does appear to have been sufficiently powerful to provide suggestive evidence for linkage to two different loci (the *CYP11a* locus (results presented in this study) and the *INS VNTR* (Waterworth *et al.*, 1997a)) for which independent evidence for association has also been obtained in additional data sets.

4.1.2 Assessment of the linkage approach taken

In the genetic analysis of common complex disorders, in which the underlying genetic model is unknown, arguments have been presented for taking both parametric and non-parametric approaches (see **section 1.8.1**). Therefore, in the linkage studies performed in this thesis both parametric and non-parametric methods have been employed to evaluate three candidate genes (*CYP19*, *CYP11a* and the *LHR*) for a role in the aetiology of PCOS/MPB. The GENEHUNTER program (**sections 1.8.1** and **2.4.3**), which is capable of carrying out both multipoint parametric and non-parametric linkage analyses simultaneously, was used to perform all the analyses. This program was selected as it is particularly suited to our PCOS/MPB family data set (small to moderately sized families), and has been shown to have advantages over other non-parametric approaches (**section 1.8.1**).

The results of the linkage studies have shown:

- Good agreement and comparable sensitivity between the parametric and non-parametric linkage results.
- Evidence for suggestive linkage at the *CYP11a* locus, by both the dominant parametric and the non-parametric methods.

- GENEHUNTER was also used to test for the presence of genetic heterogeneity at the *CYP11a* locus. The results showed significant evidence for this, and indicated that about 60% of the pedigrees may be linked to the *CYP11a* locus.

- The *CYP19* and *LHR* loci were excluded under all the linkage analyses, indicating that these genes are not likely to play major roles in the aetiology of PCOS/MPB.

Any evidence for linkage must be confirmed in other data sets. This may be done by replicating the linkage results in another pedigree data set; or by carrying out association studies in either a case control data set or, preferably, in a family-based association data set (which unlike case control studies, are not vulnerable to the possible effects of population stratification (Thomson, 1995)).

The evidence for linkage to the *CYP11a* locus in the family data set has been supported by the association results obtained for this gene using the case control data set. Significant evidence for an association between allelic variants of this gene and hyperandrogenaemia has been obtained. Furthermore, other more recent studies, which have used both linkage (Urbanek *et al.*, 1999) and association methods (Diamanti-Kandarakis *et al.*, 1999; Pugeat *et al.*, 1999; Schulze *et al.*, 1999), have also shown agreement with our result (see later **sections 4.3.1** and **4.3.3**).

PCOS is a common disorder, which most likely has a complex aetiological basis. The high risk to relatives observed within PCOS pedigrees, indicated that there may be an oligogenic basis for the inherited susceptibility, where a relatively small number of major genes are thought to be involved (refer to **section 1.5.2**). Therefore, despite the general limitations of linkage analysis in common complex disorders (refer to **section 2.4.4**), and the power (size) limitations of the PCOS/MPB data set, the results presented in this thesis have shown the linkage approach taken to be appropriate for the study of PCOS.

4.2 *CYP19*

4.2.1 Evaluation of linkage results

Both the dominant and recessive parametric and the non-parametric linkage results for *CYP19* show clear exclusion of this gene as a major causative locus for PCOS/MPB. A LOD score of -2 or less was obtained across the whole region for both the parametric analyses and there was no evidence for excess allele sharing in the nonparametric analysis. The inclusion of the *CYP19* intragenic marker in the analysis has allowed the position of this gene to be accurately placed within the framework of markers, thus adding greater confidence to the results obtained.

4.2.2 Significance of the *CYP19* results

Although a minor role for this gene cannot be ruled out (eg. genetic heterogeneity in a minority of the pedigrees) given the relatively small size of the pedigree data set, the results of the linkage analysis demonstrate that *CYP19* is not a major locus in the development of PCOS/MPB.

If sequence variation in *CYP19* is not involved in the up-regulated activity of aromatase observed in the FSH-hyperresponsive granulosa cells from anovulatory PCO (Mason *et al.*, 1994), then what is ? The most likely explanation for these observations comes from the *in vitro* study of Willis *et al.* who demonstrated that preincubation of cultured granulosa cells (from polycystic and normal ovaries) with insulin resulted in a subsequent enhancement of LH action on these cells (Willis *et al.*, 1996). They hypothesised that the effect of hyperinsulinaemia in anovulatory women was to promote premature granulosa cell maturation such that these cells respond to LH at an earlier stage of development than normal. Therefore, the cultured ganulosa cells from the anovulatory PCO (Mason *et al.*, 1994) are likely to be responding to FSH at a level comparable to granulosa cells obtained from a larger (more mature) follicle from a normal ovary. A direct comparison between the

anovulatory, and the ovulatory PCO/ normal cells (from follicles of the same size) should therefore not be made.

Total aromatase deficiency has been shown to be associated with hyperandrogenism. However, it is also associated with other distinctive phenotypes such as ambiguous external genitalia at birth, primary amenorrhea and virilization (Ito *et al.*, 1993). Hence, this is not the type of mutation expected to be involved in PCOS. Instead, a more subtle down regulation of aromatase activity may be predicted to lead to an accumulation of excess androgens. However the lack of significant difference in oestradiol production, observed between the granulosa cells derived from ovulatory polycystic ovaries and those from normal ovaries (Mason *et al.*, 1994), indicates that the activity of aromatase is unlikely to be altered in most cases of polycystic ovaries.

4.3 *CYP11a*

4.3.1 Evaluation of linkage results

Linkage analysis of the *CYP11a* locus was carried out using six polymorphic microsatellite markers, including a *CYP11a* intragenic marker (*CYP11a* $(ac)_n$), that spanned a 32cM region. The intragenic marker is useful as it has allowed the position of *CYP11a* to be anchored within the framework of markers. However, this is not a very informative marker (heterozygosity of 0.63). To increase the information at this locus, marker *D15S169*, which is at a distance of less than 0.5 cM from *CYP11a* $(ac)_n$, was included in the analysis. This increased the overall information content of the combined locus to greater than 0.95.

The results of the non-parametric linkage analysis has provided convincing evidence for linkage to this locus. A peak NPL score of 2.95 ($p = 0.0005$) was obtained at the *CYP11a* $(ac)_n$/*D15S169* locus. This p-value exceeds the genomewide scan threshold of $p < 0.001$, required for suggestive evidence of linkage (Lander and Kruglyak,

1995) (**section 2.4.3**). The guidelines set out by Lander and Kruglyak are designed to take into account the multiple testing aspect of a genomewide scan, but since only a small number of candidate genes (which therefore have a strong *a priori* reason for involvement) has been examined, these are likely to be overly stringent for our studies.

The parametric linkage analysis, under the dominant model, has also provided some evidence for linkage. A peak positive LOD score^{D} of 1.92 was obtained at the *CYP11a* $(ac)_n$/*D15S169* locus. This score has reached the criteria of a minimum LOD score of +1.9, required for suggestive linkage (refer to **section 2.4.3**). Having obtained evidence for suggestive linkage, a test for heterogeneity was also carried out. A maximum HLOD of 2.84 was obtained suggesting that, under the dominant model used, there may be genetic heterogeneity at this locus with about 60% of the pedigrees demonstrating linkage.

These results provide significant evidence for the involvement of *CYP11a* in the aetiology of PCOS/MPB, making it an important gene that deserves further investigation.

Recently, support for our results has been provided by another group who obtained nominal evidence for linkage to the *CYP11a* locus (Urbanek *et al.*, 1999). Urbanek and coworkers used a relatively small data set of 39 sibpairs (from 28 PCOS pedigrees) to carry out non-parametric linkage analysis on 37 candidate genes, using the ASP test. Nominal linkage to *CYP11a* (67% IBD, $\chi^2 = 5.34$, $p = 0.02$) was obtained, but after applying a correction for multiple testing (33 tests) this was no longer significant. This correction is likely to be overly stringent (since linkage to this locus has already been obtained in our data set). Therefore this results may be regarded as confirmation of the linkage obtained in our data set.

4.3.2 Mutation screening of *CYP11a* promoter

Most of the regulatory elements within the *CYP11a* promoter are believed to be contained within a 1.85kb region 5' of the start of translation (Moore *et al.*, 1990). Therefore, this was the sequence that was screened for mutations. Direct sequencing analysis was used to screen PCR amplified fragments of the *CYP11a* promoter from affected and unaffected members of the cosegregating pedigrees.

The results of this investigation revealed the polymorphic nature of the pentanucleotide repeat sequence $(tttta)_n$ at position -466 from the start of transcription (**figure 3.1.11**). However, apart from variation in the number of repeat units of this and the other polymorphic microsatellite marker (the dinucleotide $(ac)_n$ marker, at positions -1314, used in the linkage analysis), no other alterations to the published sequence were found in any individual. These results indicate that there are no single base changes within the region examined that could affect the regulation of this gene. It is still possible that there may be a mutation located outside this region, either within the coding (or intronic) sequence or perhaps within an, as yet unidentified, enhancer element located at a distance from the putative promoter of this gene.

The pentanucleotide polymorphism identified in this region was then evaluated in the case control data set to establish the range, frequency and distribution of alleles in PCOS cases and normal controls.

4.3.3 *CYP11a* $(tttta)_n$ evaluation in the case control data set

The results of this study have shown a significant association between the *CYP11a* $(tttta)_n$ polymorphism and serum testosterone levels ($P_c = 0.0028$ (for the total group), and $P_c = 0.017$ (for the PCOS group)), with the 216- genotype being associated with higher mean testosterone levels (**figure 3.1.13**). Although a similar trend in testosterone distribution was seen for the aPCO and normal control groups, these were not significant. The failure to demonstrate association in these groups may

reflect the small sample sizes (only 4 and 7 individuals with the 216- genotypes for the aPCO and normal control groups respectively). Alternatively, it may indicate that this locus acts epistatically, interacting with other genetic and/or environmental factors, to modify the phenotypic expression of PCOS. Overall, these results imply that variation in *CYP11a* may have a role in increased androgen production in women with PCOS, with the higher repeat length alleles of the *CYP11a* $(tttta)_n$ polymorphism being associated with this elevation.

Recently, a number of other studies have provided support for a role for *CYP11a* in the aetiology of hyperandrogenism and PCOS. Three separate groups have carried out association studies using case control data sets of PCOS and/or hirsute patients and normal controls and found significant associations between the *CYP11a* $(tttta)_n$ polymorphism and both serum testosterone levels and PCOS status (Diamanti-Kandarakis *et al.*, 1999; Pugeat *et al.*, 1999; Schulze *et al.*, 1999).

4.3.4 What is the role of *CYP11a* $(tttta)_n$?

Having obtained significant evidence for the involvement of *CYP11a* in the development of hyperandrogenaemia and PCOS, the questions that arise are: what is the aetiological 'mutation' responsible for this up-regulation of activity and where is it located in the *CYP11a* gene ?

Sequencing analysis of 1.85kb of the promoter region of *CYP11a* only showed allelic variation in the *CYP11a* $(tttta)_n$ polymorphism and did not reveal any other mutation likely to have direct pathogenic effect. The significant association between the *CYP11a* $(tttta)_n$ polymorphism and serum testosterone levels indicates either: (1) that this polymorphism itself is functional and has a direct role in the regulation of expression of the *CYP11a* gene; or (2) that this association may be due to the presence of LD between this marker and a putative aetiological mutation located

elsewhere, either within the *CYP11a* gene or possibly even within another gene in the region.

Is the aetiological mutation elsewhere in the CYP11a gene ?

In an attempt to address the possibility of a mutation within the *CYP11a* coding region, I recently started to screen all 9 exons of the *CYP11a* gene in 12 individuals (comprising probands from the co-segregating pedigrees and the case control data set). *CYP11a* consists of 9 exons (ranging in size from 80bp to 270bp) and 8 introns spread over 20kb of DNA. Direct sequencing of PCR amplified exonic fragments (including exon-intron boundaries) has been carried out using fluorescent cycle sequencing (using the Thermo Sequenase dye terminator kit, Amersham) and an ABI Prism 373 sequencer. Sequencing data has so far only been obtained for exons 1 to 6, and no variants have been identified in any individual (data not presented in this thesis).

Recently, a study has been carried out in which the coding regions of 106 human genes (relevant to cardiovascular disease, endocrinology and neuropsychiatry) have been screened in an attempt to identify and catalogue common single-nucleotide polymorphisms (SNPs) that may eventually be used in the study of susceptibility to common diseases (Cargill *et al.*, 1999). *CYP11a* is one of the genes examined. The entire coding sequence of the *CYP11a* gene was screened in 114 chromosomes. It was estimated that this sample should provide sufficient power to detect alleles with a frequency of >5% over 99% of the time and there should be 65% power to detect an allele with a frequency of <1%. PCO has a high population frequency of about 20% (with an estimated 10% disease allele frequency) and since *CYP11a* is likely to be a major contributing locus, it is likely that there is sufficient power in this analysis to detect the putative PCOS mutation. The results of this survey identified two SNPs in the coding sequence of *CYP11a*. One is a silent mutation in exon 4 (a C to T transition), and the second is an exon 5 G to A transition that changes a glutamic acid residue to a glycine. Although this changes an acidic amino acid to a small neutral

one, this is not an amino acid that is conserved between the bovine and human proteins (Morohashi *et al.*, 1987). Both SNPs have a reported allele frequency of less that 5%.

These variant alleles are not present in any of the samples that I have sequenced. Since these samples were selected with a high chance of carrying the putative mutation, it seems unlikely that these variants are involved in PCOS. However, it is sensible and necessary to at least exclude them by screening our patient samples. Also, once again for the sake of prudence, the sequencing analysis of the remaining exons (7-9) should be completed. This work is presently underway.

Therefore, so far, most of the coding sequence of this gene and the part of promoter region thought to contain the important regulatory elements has been excluded with no obvious other candidate mutation identified. This does not preclude the possibility that the mutation may be within an intron or an enhancer located at a distance from *CYP11a* or even within another nearby gene. No other strong candidate genes have been mapped very close to *CYP11a*. Therefore, this lack of evidence for another mutation brings the focus back to the *CYP11a* (tttta)n polymorphism and a potential role for this in the regulation of *CYP11a*.

Is there a function for CYP11a (tttta)$_n$?

The *CYP11a* (tttta)$_n$ polymorphism is embedded within the 5' regulatory region of the *CYP11a* gene. Significant association between serum testosterone levels and genotypes at this polymorphism have been obtained. However, no regulatory function has, as yet, been assigned to this polymorphism.

Recently, a number of gene-associated VNTRs (variable number tandem repeats) have been implicated in the regulation of genes, and in the susceptibility to common disease (Krontiris, 1995). These include the insulin gene VNTR and the *HRAS1* minisatellite which have been shown to bind transcription factors (in an allele-specific and cell-type specific manner) and modulate the transcriptional activity of

their respective genes. The insulin VNTR has been associated with increased risk to PCOS (Waterworth *et al.*, 1997a) as well as type 1 (Bennett *et al.*, 1995) and type 2 diabetes mellitus (Huxtable *et al.*, 2000), and the *HRAS1* VNTR is associated with increased risk of multiple forms of cancer (Krontiris *et al.*, 1993).

Another VNTR of interest is the $(tttta)_n$ repeat polymorphism, located 1.3kb 5' of the first exon of the apolipoprotein(a) gene (apo(a)). Over 90% of variability in plasma levels of lipoprotein(a) (Lp(a)) have been attributed to sequences at, or closely-linked to, the apo(a) locus. Most of this variability is accounted for by variation in the number of repeats of a transcribed highly polymorphic tandemly repeated cystein-rich motif called kringle 4 (K4), with an inverse relationship between repeat number and plasma Lp(a) levels (Utermann, 1989). However K4 does not account for all the variability, and *in vivo* turnover studies have suggested that variation in plasma Lp(a), observed between individuals with apo(a) K4 alleles of the same length, are due to differences in the rate of apo(a) production (Rader *et al.*, 1994). Subsequently it has been demonstrated that approximately 10-14% of variability (not accounted for by K4) is associated with variation in the number of repeats of the $(tttta)_n$ repeat polymorphism (Trommsdorff *et al.*, 1995). Direct association between alleles of this polymorphism and plasma levels of Lp(a) has been demonstrated (Mooser *et al.*, 1995; Trommsdorff *et al.*, 1995), with alleles with 11 repeats having the lowest plasma Lp(a) concentrations (Mooser *et al.*, 1995). In addition, expression studies have shown a 5-fold higher transcriptional activity in constructs containing eight copies of the apolipoprotein(a) (tttta) repeat compared to those with nine copies (Wade *et al.*, 1993).

Clearly there is a similarity between the apo(a) (tttta)n polymorphism and the *CYP11a* (tttta)n. Firstly, the repeat elements are both pentanucleotide tandem repeats of identical composition; secondly, they are present within the 5' regulatory region of their respective genes; and thirdly both have demonstrated association with serum levels of the product of their genes. Although LD cannot be ruled out as an explanation

for these associations, the possibility of a direct regulatory function for these polymorphisms should also be considered.

4.3.5 Expression studies on *CYP11a* promoter

The hormonal regulation and developmental patterns of expression of P450scc are specific to each steroidogenic tissue. Increased steroidogenesis and accumulation of P450scc mRNA are stimulated by LH and FSH in the ovary, by LH and human chorionic gonadotropin (hCG) in the placenta and testicular Leydig cells and by ACTH in the adrenals. Each of these hormones binds a cell surface receptor that activates a G protein to increase intracellular cyclic AMP (cAMP), which in turn increases the transcription of *CYP11a* (Moore *et al.*, 1990).

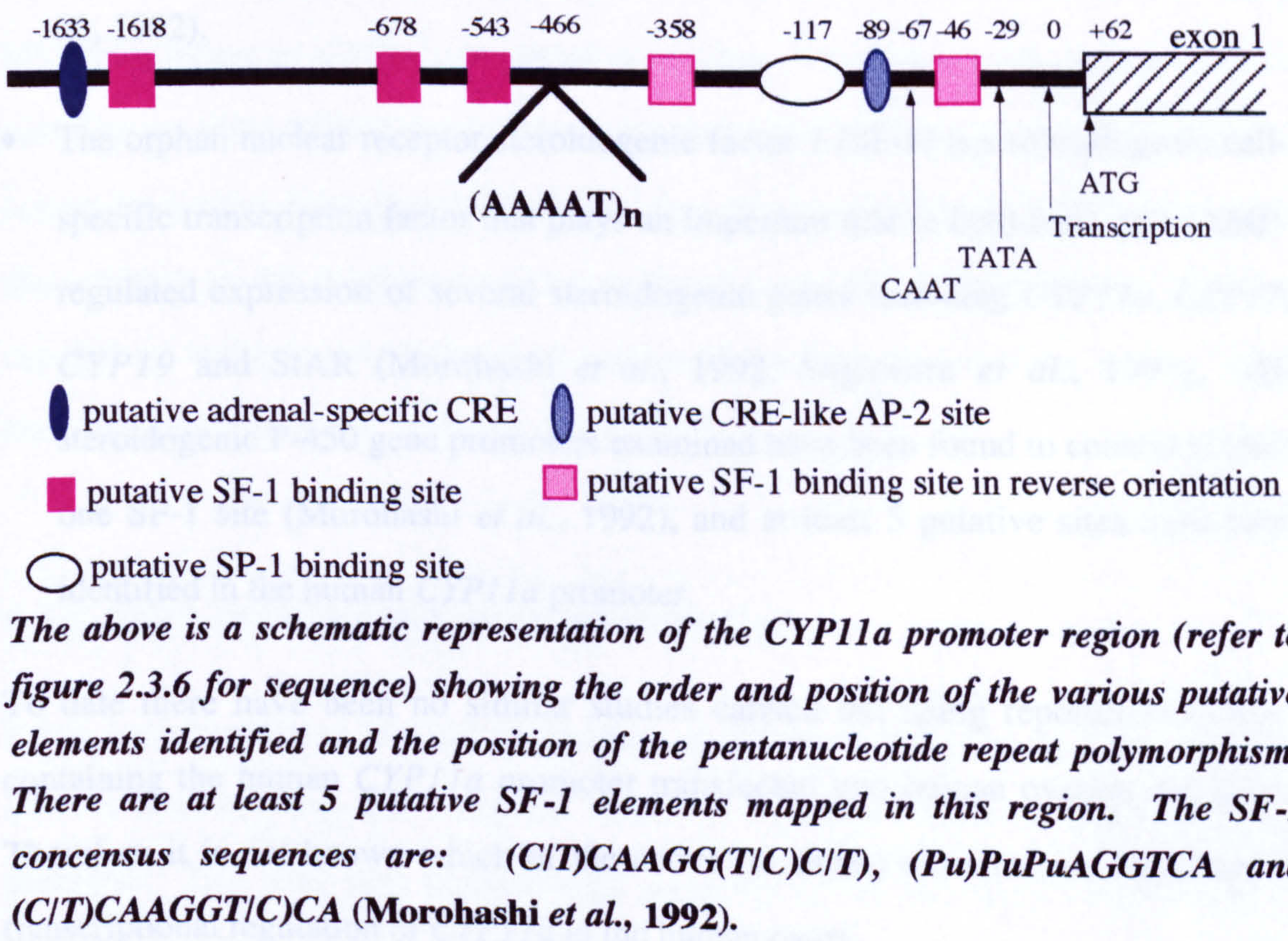
Since the mid-1980s, after the cloning of the human *CYP11a* gene (Chung *et al.*, 1986; Morohashi *et al.*, 1987), many groups have attempted to characterise the 5' regulatory region and to elucidate the activation mechanism that controls its expression (Chung *et al.*, 1989; Moore *et al.*, 1990; Inoue *et al.*, 1991; Moore *et al.*, 1992; Hum *et al.*, 1993; Takayama *et al.*, 1994; Rodriguez *et al.*, 1997; Monté *et al.*, 1998).

Most of these studies were based on the transfection of a series of expression vectors, containing deletion constructs of the 5' upstream sequences of human *CYP11a* (from -76bp to -2.5kb from the start of transcription) fused to either the bacterial gene for chloramphenicol acetyltransferase (CAT) or the firefly luciferase gene, into a variety of steroidogenic cell lines. The cell lines that have been used included mouse adrenal Y-1 (Chung *et al.*, 1989; Moore *et al.*, 1990; Inoue *et al.*, 1991; Takayama *et al.*, 1994), mouse testicular MA-10 (Hum *et al.*, 1993) and I-10 (Takayama *et al.*, 1994) cells, human NCI-H295 adrenal cells (Monté *et al.*, 1998; Rodriguez *et al.*, 1997) and human placental JEG-3 cells (Moore *et al.*, 1992). These constructs were then used to identify sequences important for both basal activation as well as cAMP responsiveness.

As data from various studies accumulated, it soon became clear that there may be both tissue and species-specific differences in the regulation of *CYP11a*. For example, the location of elements identified in constructs transfected into human NCI-H295 adrenal cells are substantially different from those that appeared functionally equivalent when the human gene promoters were transfected into mouse adrenal Y-1, mouse testicular MA-10 or human placental JEG-3 cells (Rodriguez *et al.*, 1997). Clemens *et al.* also found substantial differences between the rat and mouse *CYP11a* promoter regulatory elements expressed in respective murine granulosa cell lines (Clemens *et al.*, 1994).

However, despite the discrepancies between specific element usage in the various cells types, a general pattern has emerged (refer to **figure 2.3.6** for *CYP11a* 5' sequence). A summary of the major promoter elements identified is illustrated in **figure 4.1** and is as follows:

Figure 4.1 Schematic representation of the *CYP11a* promoter region.



***The above is a schematic representation of the CYP11a promoter region (refer to figure 2.3.6 for sequence) showing the order and position of the various putative elements identified and the position of the pentanucleotide repeat polymorphism. There are at least 5 putative SF-1 elements mapped in this region. The SF-1 concensus sequences are: (C/T)CAAGG(T/C)C/T), (Pu)PuPuAGGTCA and (C/T)CAAGGT/C)CA* (Morohashi *et al.*, 1992).**

- The basal promoter elements lie within the region -90bp from start of transcription. This region is known to contain the putative CAAT and TATA motifs as well as a SF-1 (Steroidogenic factor 1) elements at position -46 (Takayama *et al.*, 1994; Liu and Simpson, 1997b; Monté *et al.*, 1998).

- There is also an element at position -117 (GGGGAGGAGC) which matches 9 of 10 bases of the SP-1 concensus G/TGGGCGGG/AG/AC/T. This has been shown to increase basal transcription by up to 600% (Moore *et al.*, 1990).

- The main adrenal-specific region involved in cAMP regulation is at position -1626 and -1633 (TGATGTCA). This sequence matches at 7 of 8 bases with the concensus CRE (cAMP-response element) TGACGTCA (Chung *et al.*, 1989; Moore *et al.*, 1990; Inoue *et al.*, 1991; Takayama *et al.*, 1994; Monté *et al.*, 1998).

- A CRE-like element has also been identified at position -89 (GAGCCTG) in transfections using JEG-3 cells (i.e. possibly placenta specific). This matches the AP-2 site of simian virus-40 enhancer (which in known to confer cAMP regulation), and was shown to increase basal expression by up to 750% (Moore *et al.*, 1992).

- The orphan nuclear receptor steroidogenic factor 1 (SF-1) is a steroidogenic cell-specific transcription factor that plays an important role in both basal and cAMP-regulated expression of several steroidogenic genes including *CYP11a*, *CYP17*, *CYP19* and StAR (Morohashi *et al.*, 1992; Sugawara *et al.*, 1997). All steroidogenic P-450 gene promoters examined have been found to contain at least one SF-1 site (Morohashi *et al.*, 1992), and at least 5 putative sites have been identified in the human *CYP11a* promoter.

To date there have been no similar studies carried out using reporter constructs containing the human *CYP11a* promoter transfected into human ovarian cell lines. Therefore it is not known which of the above (or other) elements are important in transcriptional regulation of *CYP11a* in the human ovary.

Interaction between CYP11a promoter elements

As with other nuclear receptors (Tijan, 1995), SF-1 is likely to function as a component of a multiprotein complex in the regulation of transcriptional activity. In fact, several studies have demonstrated that there are interactions between the various *CYP11a* putative promoter elements, and that protein complexes are involved in this interaction (Takayama *et al.*, 1994; Liu and Simpson, 1997b; Monté *et al.*, 1998). Takayama and colleagues demonstrated that the distal elements between -1.8kb and -1.5kb (SF-1 and CRE) interact with the basal promoter (including the -46 SF-1 element) to induce transcriptional activation in Y-1 cells. They also showed that this interaction is only functional in steroidogenic cells (indicating the significance of SF-1). Furthermore, activation by the distal promoter is dependent on the presence of the basal promoter, which was shown to bind a nuclear protein by gel shift analysis. Lui and coworkers demonstrated that SP-1 and SF-1 function co-operatively (through their respective promoter elements at -46 (SF-1) and -117 (SP-1) (**figure 4.1**)) in the transactivation of the bovine *CYP11a* promoter expressed in Y-1 and bovine luteal cells (Liu and Simpson, 1997b). They also used a mammalian two-hybrid system to show that SP-1 and SF-1 can associate *in vivo*. Finally, Monté *et al.* have shown that SF-1 regulation of *CYP11a* expression involves interaction with the ubiquitous coactivators CBP (CREB-binding protein) and p300 (Monté *et al.*, 1998). These coactivators have been proposed to play key roles in modulating the initiation of transcription by RNA polymerase II (Lundblad *et al.*, 1995). Furthermore, they are capable of bridging transcriptional activators and the components of the basal transcriptional apparatus.

4.3.6 DNA bending and gene transcription

It has been postulated that the interaction of activators with enhancers may cause DNA to bend in a way that brings the enhancers closer to one another and to the core promoter. This may then facilitate the assembly of the complete basal complex, which in turn may distort the underlying DNA in a way that enables RNA polymerase

to begin its journey along the coding region (Tijan, 1995). Sequence directed and protein-induced DNA bending has been found in both prokaryotic and eukaryotic promoters, and in some cases, this appears to modulate the transcriptional initiation rate (Lamond and Travers, 1983; Gourse *et al.*, 1986; McAllister and Archberger, 1988).

X-ray crystallography or NMR has shown DNA bending in several structures of DNA-transcription factor complexes. DNA is severely bent in the complexes with CAP (Schultz *et al.*, 1991), TATA-box binding protein TBP (Kim *et al.*, 1993), and the integration host factor IHF (Rice *et al.*, 1996). *Escherichia coli* RNA polymerase has been shown to induce bending at the *gal* and bacteriophage T7 A1 promoters and modulate DNA bending at the *lac* promoter. The strongest known promoters share a common known bending sequence of five to six contiguous adenosines centred near position -43 (Koo *et al.*, 1986).

Furthermore, properly phased A tract-mediated bends have been shown to functionally replace the CAP binding/bending site in the *E. coli gal* promoter *in vivo* (Bracco , 1989). Similarly, appropriately phased (dA)n·(dT)n tracts, replacing the CAP binding site upstream from the *lac* promoter, were shown to increase the rate of specific transcriptional initiation by roughly 10 fold (Gartenberg and Crothers, 1991). This suggests that the effect of DNA bending modulates transcription initiation at the lac promoter (Gartenberg and Crothers, 1991). Two models were proposed to explain the role of CAP-induced bending in transcriptional activation. The first is that energy stored in the bend may facilitate mechanical processes at some stage during the transcription program (Zinkel and Crothers, 1991), and the second is that the bend may promote essential protein-DNA and/or protein-protein contacts that would not be sterically feasible with a straight promoter (Wu and Crothers, 1984).

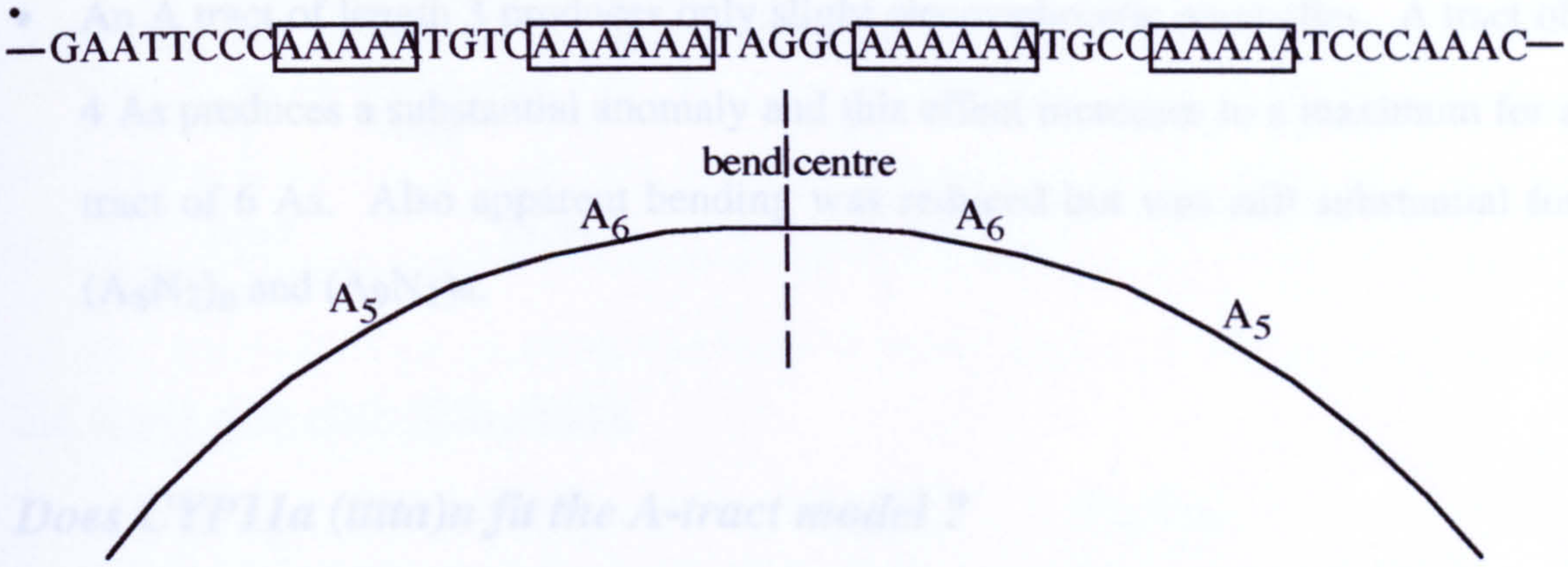
Thus it may be proposed that mechanisms that facilitate the bending of the DNA at critical sites within the transcription complex may influence the rate of transcription and therefore the regulation of the gene.

4.3.7 Is there a role for *CYP11a* $(tttta)_n$ in DNA bending ?

As yet no regulatory function has been assigned to *CYP11a* $(tttta)_n$. However, there are two features associated with this polymorphism that beg the question as to whether it may indeed have a functional role. Firstly, this repeat element is embedded in the centre of the *CYP11a* regulatory region (**figure 4.1**), flanked by many putative control elements that have been shown to bind transcription factors, which then interact directly with each other. Secondly, there is some evidence that the sequence composition of this polymorphism may induce bending in the DNA helix (as discussed below).

During the early 1980s, Wu and Crothers (Wu and Crothers, 1984) made an intriguing discovery. They identified the sequence that was responsible for the characteristic slow gel electrophoretic mobility of double-stranded (ds) DNA fragments isolated from the kinetoplast body of tropical parasites. By using the variation in gel mobility, they mapped a bending locus within a *Sau*3A restriction fragment from *Leishmania tarentolae* kinetoplast DNA. Sequencing analysis revealed a striking feature which was a regular repeat of the sequence element $CA_{5-6}T$ with a 10bp periodicity around the centre of the bend. It was presumed that each A_n tract produces a small bend in the DNA helix axis: repetition of these elements in phase with the helix screw then results in their coherent addition to form a large overall bend (**figure 4.2**).

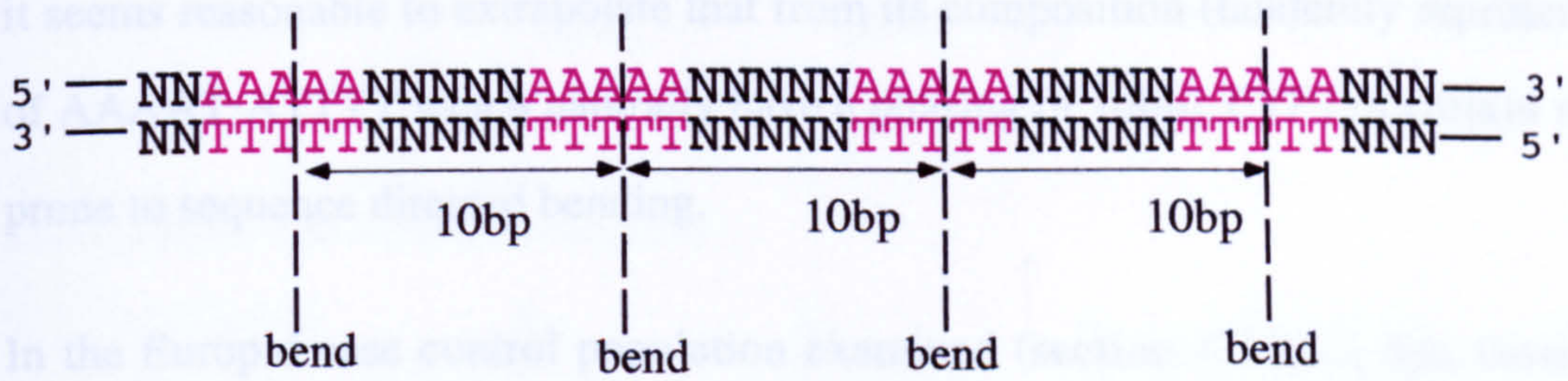
Figure 4.2 Bend site at the *Leishmania tarentolae* kinetoplast DNA



Adapted from (Wu and Crothers, 1984). ***Only the top strand of the ds DNA has been shown.***

Subsequently, non-denaturing polyacrylamide gel mobility experiments were designed to test this model (Koo *et al.*, 1986). They found that:

- Minimum mobility (maximum bending) is induced by the series $(A_5N_5)_n$, in which the 10bp phasing



nearly matches the expected helix screw of about 10.3bp per turn (calculated from the average of 10.5 for B-DNA and 10.1 for poly(dA)·poly(dT) DNA in solution).

- Phasing of less or greater than 10bp leads to a virtual normalisation of the mobility.

- Interrupting the A_5 tract with another nucleotide, i.e. $(A_2\underline{N}A_2N_5)n$, causes the gel mobility to revert nearly back to normal (i.e. bending was abolished). Therefore,

a continuous run of A residues is fundamental for the bending phenomenon. Interestingly the base that is least disruptive is a T.

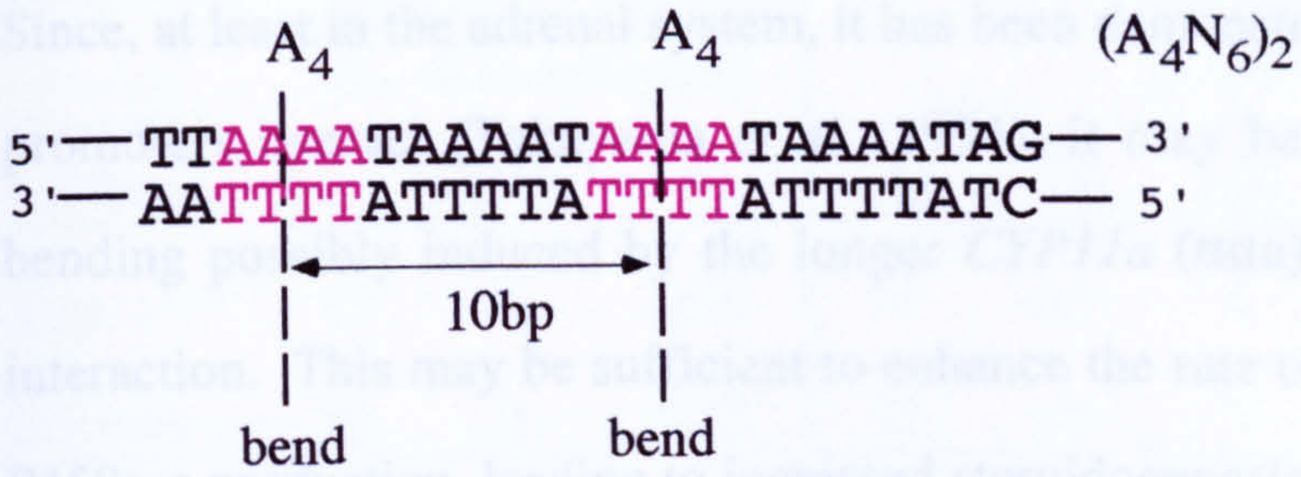
- An A tract of length 3 produces only slight electrophoretic anomalies. A tract of 4 As produces a substantial anomaly and this effect increases to a maximum for a tract of 6 As. Also apparent bending was reduced but was still substantial for $(A_8N_2)_n$ and $(A_9N_1)_n$.

Does CYP11a (tttta)n fit the A-tract model ?

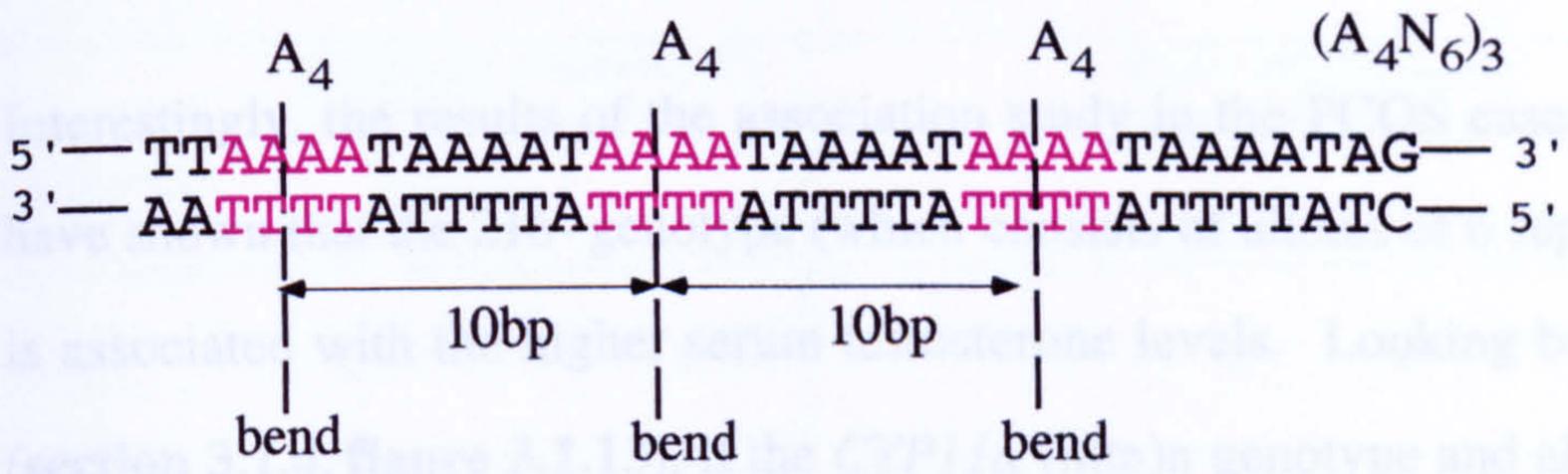
Although a limited composition of $(A_5N_k)_n$ oligonucleotides were tested (where Ns were predominantly Cs and Gs), in their experiments (Koo *et al.*, 1986) Koo and colleagues demonstrated that for maximum bending an A tract of at least 4 is required with a phasing of 10bp. Furthermore when Cs or Gs flanked both the 3' and 5' ends of the A tract (i.e. 5'-CAAAAAC-3' or 5'-GAAAAAG-3') bending was reduced, and bending was maximal when there was a C at the 5' and a T at the 3' ends (a T at both ends was not tested). Lastly, they demonstrated that a T residue was the least disruptive when placed in the middle of an A tract. Therefore, although a sequence with the exact composition as the *CYP11a* (tttta)n was not tested in their experiments, it seems reasonable to extrapolate that from its composition (tandemly repeated units of AAAAT·ATTTT which naturally have a phasing of 10bp), *CYP11a* (tttta)n may be prone to sequence directed bending.

In the Europid case control population examined (**section 3.1.4**) in this thesis, four alleles were identified with 4, 6, 8 and 9 repeats. The predicted phasing of these dA·dT tracts is as follows:

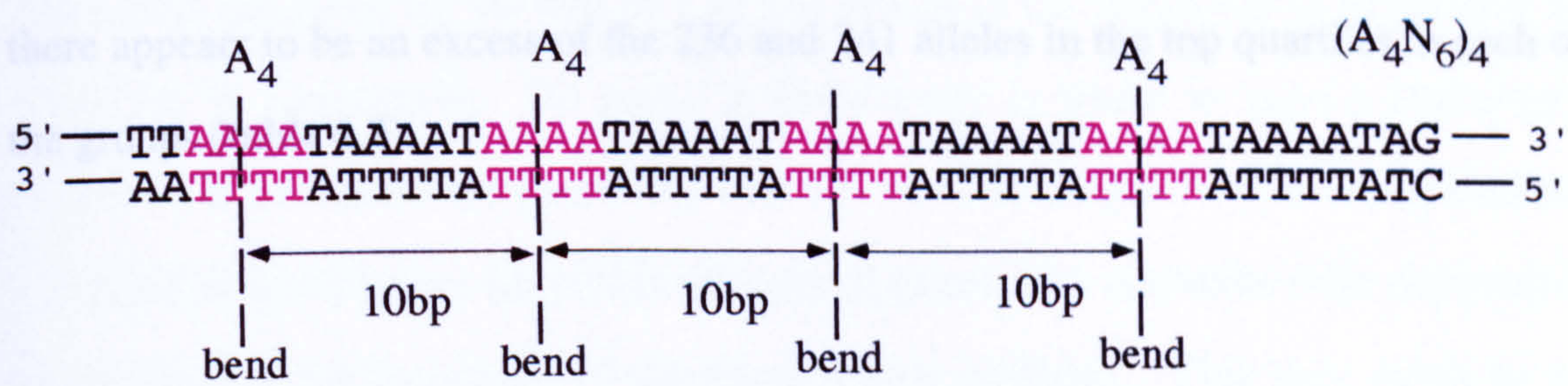
Four repeat units allele (216 allele):



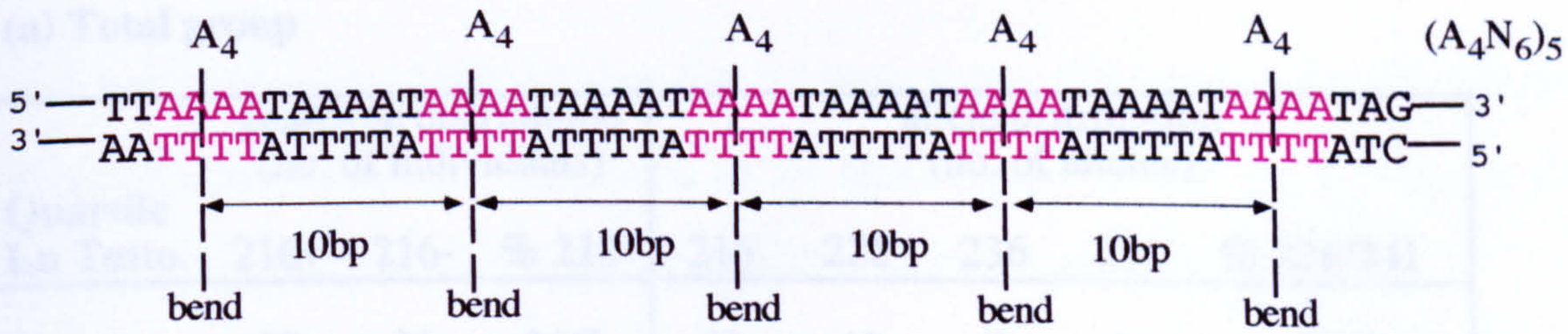
Six repeat units allele (226 allele):



Eight repeat units allele (236 allele):



Nine repeat units allele (241 allele):



Therefore, if this locus is involved in sequence-directed bending then clearly the longer alleles will bend to a greater degree than the shorter alleles.

Proposed role of CYP11a (tttta)n in transcriptional regulation

Since, at least in the adrenal system, it has been demonstrated that the distal and basal promoters interact (Takayama *et al.*, 1994), it may be postulated that the greater bending possibly induced by the longer *CYP11a* (tttta)n alleles may facilitate this interaction. This may be sufficient to enhance the rate of transcription, and therefore P450scc production, leading to increased steroidogenesis. Alternatively, this bending may be involved in the interaction of more distantly located activators with the core transcription complex.

Interestingly, the results of the association study in the PCOS case control data set have shown that the 216- genotype (which consists of alleles of 6 repeats and longer) is associated with the higher serum testosterone levels. Looking back at these data (**section 3.1.4, figure 3.1.13**), if the *CYP11a* (tttta)n genotype and allele distributions of top and bottom quartiles of the Ln testosterone levels of each of the groups (except the aPCO group as there were too few individuals for comparison) are compared, there appears to be an excess of the 236 and 241 alleles in the top quartiles in each of the groups (**table 4.3**).

Table 4.3 *CYP11a* (tttta)n genotype and allele distributions in top and bottom quartiles of Ln total serum testosterone.

(a) Total group

Quartile Ln Testo.	genotype distribution (no. of individuals) 216+	216-	% 216-	allele distribution (no. of alleles) 216	226	236	241	% 236/241
Top	37	21	**36%**	52	41	9	14	**20%**
Bottom	42	7	**14%**	62	29	1	6	**7%**

(b) PCOS group

Quartile Ln Testo.	genotype distribution (no. of individuals) 216+	216-	% 216-	allele distribution (no. of alleles) 216	226	236	241	% 236/241
Top	16	11	**41%**	21	20	4	9	**24%**
Bottom	23	3	**12%**	33	15	2	2	**8%**

(c) Normal control group

Quartile Ln Testo.	genotype distribution (no. of individuals) 216+	216-	% 216-	allele distribution (no. of alleles) 216	226	236	241	% 236/241
Top	12	2	**14%**	18	3	5	2	**25%**
Bottom	13	1	**7%**	20	7	0	1	**4%**

The model presented for the role of *CYP11a* (tttta)n in the regulation of expression of *CYP11a,* is speculative. To prove it definitively is likely to pose a challenge. However, it has raised the possibility that this class of simple VNTRs when situated upstream or downstream (or within introns) of genes may modulate their expression by facilitating the formation of the transcription complex. This may result in an enhancement of transcription (if activators are involved) or may even lead to a repression of transcription if repressors are involved.

The apo(a) gene may indeed be another example where DNA bending at the (tttta)n polymorphism influences expression of the apo(a) gene. Interestingly, there is an inverse relationship here between the allele repeat length and transactivation.

4.3.8 Evidence for general upregulation of steroidogenic enzymes

A recent *in vitro* study by a group in Pennsylvania has provided further evidence for a role for *CYP11a* in the hypersecretion of androgens in polycystic ovaries (Nelson *et al.*, 1999). Nelson and coworkers examined steroid hormone production, steroidogenic enzyme activity and mRNA expression in normal and PCOS theca cells propagated in long-term culture. They found a marked increase in progesterone, 17 α-hydroxyprogesterone and testosterone production by theca cells from PCO compared to normal theca cells (refer to **figure 1.4**). This supports the earlier study by Gilling-Smith *et al.*, who showed hypersecretion of progesterone and androstenedione in primary cultures of PCO derived theca cells compared to normal theca cells (Gilling-Smith *et al.*, 1994). However, in the more recent study, the long-term propagation of these cells in culture demonstrates that this increased steroidogenesis is an intrinsic abnormality in PCOS theca cells and not a result of residual effects of the hormonal milieu to which the cells were exposed *in vivo*.

They further demonstrated that this increased steroidogenesis occurred as a result of increased *CYP11a* and *CYP17* mRNA expression and increased 3 β-hydroxysteroid dehydrogenase and 17 β-hydroxysteroid dehydrogenase activity in PCOS theca cells compared to normal cells (Nelson *et al.*, 1999), suggesting that this upregulation of expression is a stable and intrinsic property of PCOS theca cells. They also showed that in contrast to the *CYP11a*/*CYP17* upregulation, the magnitude of steroidogenic acute regulatory protein (StAR) mRNA expression levels did not differ between the PCOS and normal theca cells.

The results of this study have important implications. It had been argued that the rate-limiting step in steroidogenesis is the transport of cholesterol into mitochondria where it is subsequently converted to pregnenolone by P450scc. The protein responsible for this transport is StAR. The observed normal StAR mRNA levels in the PCOS theca compared to normal theca indicate that the increased steroidogenesis in polycystic ovaries is not due to an up-regulation of StAR. Furthermore, in the

recent PCOS linkage study by Urbanek and coworkers, StAR was excluded as a major susceptibility locus (Urbanek *et al.*, 1999). Therefore both the StAR gene and its regulators have been excluded as having a significant aetiological role in PCOS.

Since the acute regulation of StAR, like that of *CYP11a* and *CYP17*, is cAMP and SF-1 regulated, this suggests that there is no overall difference in cAMP or SF-1 regulation in PCOS theca compared to normal theca. Instead it suggests that there may be a selective alteration in steroidogenic enzyme expression.

The selective upregulation of these enzymes may reflect an intrinsic abnormality within the genes of the particular enzymes (e.g. a genetic abnormality in *CYP11a*). However, this is unlikely to be the case for *CYP17* as no evidence for an intrinsic defect in this gene has been obtained. Alternatively, it may involve a component of a signal transduction pathway that selectively influences the expression of multiple genes in the steroidogenic machinery. It may even be a combination of the two mechanisms. Linkage and association data suggest that, at least for *CYP11a*, an intrinsic genetic defect may be present. This may act synergistically with the putative *trans*-acting factor to further increase the expression of this gene.

The acute biosynthesis of steroid hormones, in response to trophic hormone stimulation, occurs in the order of minutes to tens of minutes. In contrast, the chronic regulation occurs in the order of hours (Stocco, 1999). In their experiments, Nelson and coworkers cultured PCOS and normal theca cells with forskolin (a cAMP agonist) for 48 hours before measuring *CYP11a* and *CYP17* mRNA levels (Nelson *et al.*, 1999). It may be possible that after the acute activation of steroidogenesis, the intrinsically increased steroidogenic activity of PCOS cells (which may be mediated by a single component, e.g. *CYP11a* upregulation) might then lead to a secondary general upregulation of steroidogenesis by an autocrine mechanism. This may be mediated by testosterone, which has been shown to stimulate increased granulosa and theca cell proliferation in an androgen-receptor mediated manner (Vendola *et al.*, 1998).

4.4 *LHR*

4.4.1 Evaluation of linkage results

A summary of the linkage results obtained for the *LHR* locus are as follows:

- Radiation hybrid mapping localised the *LHR* gene to the interval *D2S119-LHR-D2S123*, at a distance of about 1-2cM from *D2S123*.

- Subsequently, parametric and non-parametric linkage analyses were carried out using six polymorphic microsatellite markers spanning the *LHR* locus.

- The *LHR* was excluded as a major susceptibility locus for PCOS/MPB by both the parametric (dominant and recessive) and the non-parametric analyses.

- The parametric analysis under the recessive model was unable to provide sufficient evidence for proof or rejection of linkage for the region encompassing the markers *D2S119* to *D2S352*. This reflects inadequate power to exclude at this locus and may, in a larger data set, indicate the presence of a minor locus.

4.4.1 Significance of *LHR* gene results

Although a minor role for this gene cannot be ruled out, the results of the *LHR* locus linkage analysis suggest that this gene does not have an important role in the pathogenesis of PCOS/MPB. These results are supported by two other recent studies. In the study by Urbanek *et al.*, linkage analysis was carried out on 37 candidate genes (one of which was the *LHR* gene) using a relatively small PCOS sibpair data set (mentioned earlier in **section 4.3.1** and again later in **section 4.7.2**). The results of this study were in agreement with the results obtained in our analysis: no evidence for excess allele sharing was found (Urbanek *et al.*, 1999). The second study to lend support for the exclusion of the *LHR* as a major contributing locus in PCOS is the work carried out by Nelson *et al.* (Nelson *et al.*, 1999) (mentioned above in the

previous section (**section 4.3.8**)). In the ovary, the acute regulation of many steroidogenic genes, such as *CYP11a*, *CYP17*, *CYP19* and *StAR* etc., is under the control of LH (acting through the LH receptor) via the cAMP regulated pathway. The evidence for up-regulation of expression of selected steroidogenic genes (eg. *CYP11a* and *CYP17*) and not the expression of others, such as the *StAR* gene, in PCOS theca cells compared to normal theca cells, therefore suggests that this up-regulation does not involve the cAMP pathway. These results provide indirect evidence that cAMP regulation of steroidogenesis (i.e. from the LH receptor signal transduction down to the transactivation of the appropriate target steroidogenic genes) may be normal in PCOS.

4.5 *CYP17*

4.5.1 Objective of the follow-up study

The purpose of this follow-up study was to extend the previous results obtained in our laboratory (Carey *et al.*, 1994) and to test the hypothesis that the A2 allele (through an up-regulation of the expression of *CYP17*) may be associated with an increased synthesis of androgens.

4.5.2 Outcome of this extended study

The results of this larger follow-up study have shown that there is no association between genotypes of this variant and serum testosterone levels. Furthermore, there is no longer evidence for association of the A2 allele with PCOS. These results suggest that this gene does not have an important role in the aetiology of hyperandrogenaemia in PCOS women. They further indicate that the presence of the variant A2 allele (and therefore a proposed extra SP-1 element) is unlikely to have a significant role in up-regulation of the expression of *CYP17* (which would be

expected to lead to increased enzyme activity and thus higher androgen production). Since the publication of the original findings, other groups have also investigated this gene by both association (Pugeat *et al.*, 1996) and linkage (Urbanek *et al.*, 1999) methods and have similarly found no evidence for a role for this variant in the aetiology of hyperandrogenaemia and PCOS.

The discrepancy between the initial and present results is probably due to a type I error resulting from the relatively small number of normal controls in the original data set. It may therefore be concluded that this base-pair change identified in the promoter region of the *CYP17* gene is likely to be a common polymorphism with no obvious role in the aetiology of PCOS or hyperandrogenaemia. However, definitive proof that this variant allele is not functional may only be obtained by carrying out expression studies on the promoter of this gene.

4.5.3 Significance of the CYP17 results

The overall results obtained for *CYP17* indicate that sequence variation in this *gene* is unlikely to have a significant role in the development of PCOS. However, recent data from the study by Nelson *et al.* (mentioned above in **section 4.3.8**) have shown that *CYP17* mRNA levels are elevated in PCOS theca cells and this appears to be a stable and intrinsic property of PCOS theca. This indicates that, as proposed previously (Barnes *et al.*, 1989; Rosenfield *et al.*, 1990; Gilling-Smith *et al.*, 1994), the enzyme P450c17α plays a significant role in the hypersecretion of androgens in PCOS. This upregulation of *CYP17* expression does not appear to be as a result of variation in the *CYP17* gene but is most likely due to an upstream *trans*-acting factor. This factor may be a primary defect (i.e. may be a susceptibility locus) or may be under the regulation of yet another upstream component.

4.6 *SRD5A1*

4.6.1 Summary of association results

The *SRD5A1* gene was evaluated in the case control data set for a potential role in the development of hirsutism. This has been done by using the two previously identified RFLPs that are single base silent substitutions in exons 1 (*Hinf*I RFLP) and exon 2 (*Nsp*I RFLP) of the *SRD5A1* gene (Jenkins *et al.*, 1991). Each RFLP was assessed individually by comparing the genotype distributions between the hirsute and non-hirsute groups of women. Haplotype analysis was carried out to test for the presence of $\overset{LD}{\curlywedge}$ between these polymorphisms and subsequently haplotype distributions were compared between the hirsute and non-hirsute subjects.

The results of this study were as follows:

- An association between the *NspI* RFLP "+,+" genotype and hirsutism was obtained with an odds ratio of 2.4 for being hirsute given the "+,+" genotype (P_c = 0.04).
- In contrast to the *NspI* RFLP results, there was no significant association between the *HinfI* RFLP and hirsutism.
- Haplotype analysis revealed the presence of $\overset{LD}{\curlywedge}$ between the *NspI* and *HinfI* RFLPs ($p < 0.0001$), with a significant excess of "HinfI(+)/NspI(+)" haplotypes and significantly fewer "HinfI(-)/NspI(+)" haplotypes than would be expected by chance.
- Comparison of haplotype distributions between the hirsute and the non-hirsute groups has shown a significant difference in haplotype frequencies ($P_c = 0.017$). There were significantly more "*HinfI(+)/NspI*(+)" haplotypes in the hirsute (38%) compared to the non-hirsute group (24%).

4.6.2 Interpretation of the data

Using a relatively small case control data set, an association between *SRD5A1* and hirsutism has been demonstrated. The results of the *NspI* RFLP analysis suggest that the "+" allele is associated with hirsutism. This may act either by having a direct functional role or through LD with a putative aetiological "mutation", located elsewhere in the gene.

The lack of association observed for the *Hinf I* "+" allele with hirsutism may be due to a type II error (i.e. a lack of power of the data set). The power (sample size required) to detect association between a variant (aetiological mutation) and disease depends on many factors including the frequency of the variant and the increase in susceptibility that it confers as well as the magnitude of LD between the variant and the nearby marker and the frequency of the disease associated marker allele (Cox and Bell, 1989; Risch and Merikangas, 1996; Kruglyak, 1999). In general, as the allele frequency of the disease associated marker increases, then so does the size of the data set required to detect any significant deviations from the frequency expected under the assumption of no association (Cox and Bell, 1989). Moreover, as the difference between the frequencies of the variant and the associated marker allele increases then the maximal level of LD between them decreases, which will in turn decrease the power to detect association (Kruglyak, 1999). Therefore, the lack of observed association of the *HinfI* RFLP "+" allele may partly be explained by the fact that it is a relatively common allele (frequency of 0.56) compared to the NspI RFLP "+" allele (0.29).

Haplotype analysis is more useful for the detection of association than single-marker analysis as it will increase the effective level of LD with the variant (the higher informativeness of a haplotype being a consequence of its lower frequency on non-variant chromosomes) (Kruglyak, 1999). This has been demonstrated by the above results, where a greater significance has been obtained for the haplotype analysis.

Furthermore, the results suggest that the variant may be located on HinfI(+)/NspI(+) chromosomes.

Alternatively, the results obtained may have arisen by chance, given the small size of the data set (i.e. a type I error), or may be an indication of a difference in the population substructure between the cases and control (e.g. due to population stratification). Although the chances of ethnic mismatching was minimised by carefully selecting only Europid individuals, in reality Europeans are quite a heterogeneous group and therefore the possibility of the latter situation cannot be ruled out.

Given the small size and type of the data set used in this study, these results need to be confirmed in a larger, preferably family-based association data set (e.g. consisting of trios of affected cases and their parents). The main advantage of family-based association methods such as the transmission disequilibrium test (TDT) is that controls are derived internally and therefore are not vulnerable to errors associated with differences in population substructure (Spielman *et al.*, 1993).

4.6.3 Assessment of the role of *SRD5A1* in hirsutism

The preliminary results of this study suggest that variation in *SRD5A1* may play a role in the aetiology of hirsutism. This susceptibility locus appears to confer a relatively modest increased risk of developing hirsutism (odds ratio of 2.4 for the *NspI* "+,+" genotype and OR of 1.9 for the *HinfI*(+)/*NspI*(+) haplotype). The results of both the haplotype and the single-marker analyses indicate that the aetiological variant is more strongly associated with the *NspI* "+" allele. This polymorphism does not alter the amino acid sequence of the protein and is hence unlikely to be functional. However, examples of apparently silent mutations affecting mRNA splicing and stability have previously been reported in other genes (Liu *et al.*, 1997a; Milland *et al.*, 1996). Therefore a functional role for this polymorphism cannot be ruled out.

Alternatively this polymorphism may be in LD with a putative aetiological mutation located elsewhere in the gene (which probably resides on *HinfI*(+)/*NspI*(+) chromosomes). In this case, the actual increased risk conferred by this susceptibility locus may be higher than indicated by these results. If the strength of LD between the NspI "+" allele and the aetiological variant is strong then the observed risk is closer to the true risk. If the LD between these loci is weak, then a much larger data set is required to detect this increased risk.

Any mutation in *SRD5A1*, that is associated with hirsutism, may act by upregulating the activity of 5 alpha-reductase type 1 enzyme, leading to an increased production of DHT in the skin.

4.7 FUTURE WORK

In this thesis, a candidate gene approach, using two methodological strategies (linkage analysis in a pedigree data set and association studies in a case control data set), has been used to investigate the genetic basis of PCOS. The results have revealed evidence for the involvement of one gene, *CYP11a*, as a susceptibility locus in the development of hyperandrogenaemia and PCOS and another locus, *SRD5A1*, in the aetiology of hirsutism. However, since both the data sets used were relatively small, these results need to be confirmed by replication in separate, larger data sets.

Linkage analysis and association studies are important complementary approaches to the analysis of complex traits. Both non-parametric (NPL method of GENEHUNTER) and parametric (two simple SL approach) linkage analysis methods have been proposed as appropriate and valid methods for the study of complex traits (see **section 1.8.1**) (Kruglyak *et al.*, 1996; Greenberg *et al.*, 1998; Durner *et al.*, 1999). The main limitation of linkage analysis is the power to detect loci of smaller effect. Risch and Merikangas have suggested (by using an ASP approach) that

linkage analysis is really only feasible for the detection of loci of relatively high effect, e.g. those that confer a genotype relative risk (GRR) (the increased chance that an individual with a particular genotype has the disease) of 4 or greater (Risch and Merikangas, 1996). They have also shown that, in contrast to linkage analysis, association studies have greater power to detect genes of more modest effect (GRR of < 2). The main limitation of association analysis is, that the disease locus itself, or a nearby marker that is in LD with it, needs to be tested. Therefore this, for the present time at least, limits the use of the latter approach to the investigation of candidate genes.

Eventually, it is hoped that, once the limitations of current technological capabilities have been overcome, whole genomewide LD association studies may be carried out using single nucleotide polymorphisms (SNPs) that cover the entire genome. It has been estimated that between 100,000 to 500,000 such SNPs would be required to allow sufficient power to detect significant associations (Risch and Merikangas, 1996; Kruglyak, 1999).

4.7.1 Sample collection

Sibling-pairs

As mentioned above, the most ideal data set for the study of complex traits is one that allows both linkage and association analyses to be performed. For this reason our group at St Mary's has already started to collect a data set of affected sibling-pairs (sibpairs) (i.e. a PCOS proband plus at least one affected sister), with the addition of DNA samples form both parents, wherever possible. This data set may be used as a pedigree data set for linkage analysis and can also be used as a parent-offspring trio data set to carry out family-based association studies, such as the transmission disequilibrium test (TDT), on candidate gene polymorphisms. Further reasons for collecting sibpairs, as opposed to larger families (as we had done before), is that, being smaller, they are easier to collect. They also avoid the uncertainties associated

with assignment of the postmenopausal women and the issues regarding the male phenotype of MPB. DNA samples for additional siblings (male or unaffected sisters) may also be collected wherever possible as it may aid genotyping where parents are missing and, in the case of male siblings, may allow further study of the male phenotype of PCOS.

Simulation studies aimed at estimating the size of the sibpair data set required for the study of PCOS have shown great variation depending on the underlying model tested (McCarthy *et al.*, 1998). However, it has been estimated that 400 sibpairs, used in a genomewide scan, should provide 80% power to detect loci with λs of 1.5 at a LOD score of 2.6 (and 95% power at a LOD of 1.5).

Parent-offspring trios

In addition to the sibpairs, a parent-offspring trio data set is being collected which will be used to carry out family-based association studies (eg. TDT) to evaluate candidate polymorphisms. This data set consists of PCOS affected probands and DNA from both parents. Family-based association methods have several advantages over traditional case control association studies. The main advantage is that controls are derived internally and therefore are not vulnerable to false positive errors associated with differences in population substructure (Spielman *et al.*, 1993). This type of data set also allows the assessment of parent-of-origin effects such as those that have already been observed at the *INS*-VNTR locus for PCOS (Bennett *et al.*, 1997; Waterworth *et al.*, 1997b), type 1 (Bennett *et al.*, 1996) and type 2 diabetes (Huxtable *et al.*, 2000).

Detailed phenotypic and biochemical characterisation

Detailed clinical, ultrasound, endocrine and metabolic data will be gathered from probands and siblings, but not from parents. Ascertainment of parents is solely for genotypic, rather than phenotypic, information. Clinical assessment of probands and

siblings will include ovarian ultrasound, detailed menstrual history, body mass index (BMI), waist hip ratio, Ferriman-Gallwey score (hirsutism), assessment of the presence or absence of acne and alopecia. Biochemical data will include, at minimum, serum measurements of testosterone, androstenedione, SHBG, LH, FSH, DHT, lipids, fasting glucose and specific insulin. In addition, glucose tolerance will be tested (where possible) by a 75g oral glucose tolerance test (OGTT). Male siblings will also be assessed for the presence or absence of premature MPB and similar detailed biochemical data will be gathered for further characterisation of the male phenotype.

Once sufficient numbers have been collected, these measurements will allow sub-phenotyping and sub-classification of women with PCOS (e.g. with regards to ovulatory status, hirsutism or obesity). This may aid the genetic analysis by enriching (increasing the λs for a particular phenotype), within the subgroup, for susceptibility genes relevant to that phenotype.

4.7.2 Experimental approaches

So far, approximately 80 sibpair pedigrees (some of which have more than one affected sister) and 150 parent-offspring trios (not including the sibpairs) have been collected. Initially these new patient resources will be used to:

- Confirm positive results already obtained for previously investigated candidate genes such as *CYP11a*, *INS*-VNTR and *SRD5A1* (also see below **section 4.7.3** and **4.7.4**).

- Investigate other pertinent candidate genes which are relevant to androgen biosynthesis and / or metabolism, follicular development and function and to the related nutritional and metabolic pathways (e.g. insulin secretion and action).

- Investigate any other candidates identified by other studies of PCOS and related disorders such as NIDDM and obesity. For example, in the recent study by

Urbanek *et al.* (Urbanek *et al.*, 1999) (mentioned previously in **section 4.3.1**), a relatively small data set of 39 PCOS affected sibpairs was used to carry out non-parametric linkage analysis (using the ASP approach) on 37 candidate genes from 33 chromosomal regions. Apart from the nominal linkage to *CYP11a* (mentioned earlier **section 4.3.1**), evidence for linkage to another locus, the follistatin gene, was obtained. Significant excess allele sharing (72% IBD, p = 0.0003) was observed for the follistatin locus, and this remained significant after correction for the 33 tests (P_c = 0.01). Follistatin, an activin binding protein, plays a role in ovarian follicular development by neutralising the activity of activin *in vivo* and *in vitro* (Mather *et al.*, 1997; Guo *et al.*, 1998). Activin is a member of the transforming growth factor-β superfamily and promotes ovarian follicular development, theca cell androgen production and increases pituitary FSH secretion (Shibata *et al.*, 1996; Mather *et al.*, 1997). Therefore, follistatin is an important candidate for investigation in our PCOS data sets.

- Wherever possible, the TDT data set may be used to carry out LD mapping of candidate loci.

- The TDT data set may also be used in the (near?) future, once a whole genome SNP map is available and the limitations of current technological capabilities have been overcome, to carry out LD mapping of the loci that confer more modest effects on disease susceptibility.

Once the sibpair data set is large enough, this resource may be used in the following ways:

- To carry out a genomewide scan, depending on how much of the major genetic component of PCOS remains to be elucidated.

- To subclassify the data set according to phenotypic (using clinical and/or biochemical data) and genetic information (e.g. evidence for linkage to a particular locus such as *CYP11a* or the *INS.* VNTR). This may allow refinement

of the aetiology of PCOS in each subgroup and allow the role of modifying genes to be investigated within each group. It is also likely to be important in reducing factors that are known to confound genetic analysis in cases such as genetic heterogeneity. Genetic classification of the data set may also allow the interaction of susceptibility loci to be investigated (e.g. *CYP11a* and *INS*. VNTR or *SRD5A1*). This may reveal information about the mode of interaction, i.e. whether they behave independently, as in the case of genetic heterogeneity, or whether they interact epistatically.

The positive results obtained in this thesis, for the genes *CYP11a* and *SRD5A1*, will also be followed up (see below **sections 4.7.3** and **4.7.4**).

4.8.3 CYP11a

Future work on CYP11a and its role in the aetiology of PCOS can be divided into four main areas:

- Replication of the linkage and association results using the recently-acquired sibpair and TDT data sets (refer to section 4.7.1 above). This will allow confirmation of the CYP11a locus as a PCOS susceptibility gene.

- Completion of the mutation screening of the remaining CYP11a exons (7 to 9) should also be carried out. Although no SNPs were identified in these regions by the Cargill et al. study (Cargill et al., 1999), it is still important to finish this work in our data set.

- Evaluation of the two SNPs recently identified in the *CYP11a* coding region (Cargill *et al.*, 1999), as well as any others that may be present in exons 7 to 9, to be carried out using all the different PCOS data sets (i.e. the pedigree and case control data sets used in this thesis as well as the recently collected samples). The presence or absence of LD between these SNPs and the *CYP11a* (tttta)n

polymorphism may also be assessed. If LD is found to be present then it may prove useful in carrying out LD mapping of the aetiological mutation.

- Exploration of the possibility of a role for *CYP11a* (tttta)n in both sequence-directed DNA bending and in the regulation of expression of the *CYP11a* gene. The former may be carried out by evaluating the mobility of DNA fragments containing different *CYP11a* (tttta)n alleles in non-denaturing polyacrylamide gels (see below). The latter may be investigated by performing expression studies on a variety of constructs carrying the *CYP11a* promoter region (containing different *CYP11a* (tttta)n alleles) fused to a reporter gene (see below).

Polyacrylamide gel mobility analysis

Polyacrylamide gel mobility analysis may be used to test the hypothesis that the *CYP11a* (tttta)n results in sequence-induced DNA curvature. This experimental approach is based on the fact that curved double-stranded (ds) DNA migrates slower than straight ds DNA in non-denaturing polyacrylamide gels, with increased bending (which results in a decrease in DNA end-to-end distance) leading to a decrease in gel-electrophoretic mobility (Lerman and Frisch, 1982; Koo *et al.*, 1986; Diekmann, 1992).

This type of gel assay is relatively simple to perform and has the following advantages:

- Since only small amounts of DNA are needed for analysis, PCR amplified DNA fragments can easily be tested.

- There is no need to purify test DNA fragments since the mobility of DNA molecules is not affected by the presence of other DNA fragments and a variety of chemical compounds present in small quantities.

- Data analysis is relatively simple. One approach is to construct a standard curve from the migration of normal marker fragments (various standard DNA fragment

ladders such as *Hae*III digest of ϕX or *BsuRI* digest of pBR322) and use this to determine the apparent length of the test DNA fragment. The apparent length divided by the actual sequence length defines the relative value, the *k* factor, which is plotted against the actual fragment sequence length. Fragments with normal mobility in gels will have a *k* factor of 1.0.

- For considerably curved DNA, the sensitivity of bend detection is quite high since the gel migration anomaly increases with the square of the degree of DNA curvature.

- The effects of factors known to influence the gel migration of bent DNA can also be analysed, e.g. running of the gel in the presence of ethidium bromide will ameliorate the migration anomaly.

The experimental design to be used is as follows:

- DNA from individuals homozygous for each of the four *CYP11a* (tttta)n alleles (i.e. 216,216 (4 repeats), 226,226 (6 repeats), 236,236 (8 repeats) and 241,241 (9 repeats)) will be used to PCR amplify test DNA fragments (using the same primer set as before **section 2.3.11**).

- The migration of each of these alleles will be analysed by running test DNA fragments on a 8% non-denaturing polyacrylamide gel together with a set of standard marker ladders.

- After electrophoresis is complete, the gel will be stained with ethidium bromide and then viewed and photographed using a UV transilluminator.

- A standard curve (logarithm of base pairs versus distance (cm) migrated) will be constructed and the *k* factor for each test DNA sample obtained. This may then be used to assess the presence and relative degree of DNA bending for each of the test samples.

- This experiment can also be repeated with ethidium bromide present during the running of the gel. This will reduce any migration anomaly resulting from DNA bending. The use of ethidium bromide in this way will also distinguish between gel migration anomalies arising from DNA curvature and those arising from other structural variations of DNA (e.g. a DNA four-way junction). While the migration abnormality of the former is substantially reduced, that of the latter will be increased further.

Expression studies

The objective of carrying out expression work on the promoter of *CYP11a* is to assess the role of the *CYP11a* (tttta)n in the regulation of transcription. There are several reasons why the results of previous expression studies **(section 4.3.5)** are inconclusive as far as evaluating the function of the *CYP11a* (tttta)n is concerned. Firstly, in each study all deletion constructs tested were prepared from the same fragment of promoter DNA sequence and therefore only one *CYP11a* (tttta)n allele was tested. In fact in five out of six studies the same human *CYP11a* genomic clone (containing the 4 repeat *CYP11a* (tttta)n allele) was used to construct expression plasmids (Chung *et al.*, 1989; Moore *et al.*, 1990; Moore *et al.*, 1992; Rodriguez *et al.*, 1997; Monté *et al.*, 1998). The sixth study used a clone which contained a 6 repeat unit allele (Takayama *et al.*, 1994).

Secondly, the design of the experiments was aimed at identifying regions and, specifically, elements that bind trans-acting transcription factors. Therefore, serial deletion constructs would not necessarily reveal a function for the *CYP11a* (tttta)n sequence *per se* since it is not thought to bind a transcription factor. Furthermore, if it really does have a role in facilitating the interaction of flanking promoter elements then any construct that removes a putative distal element would indirectly affect the function of *CYP11a* (tttta)n. Moreover a so-called "distal element" may actually be located at a considerable distance from the basal promoter region *in vivo* and therefore such expression studies would not reveal the function of the *CYP11a*(tttta)n.

The study by Takayama *et al.* showed that the human *CYP11a* distal (-1.8kb to -1.5kb) and basal (up to -57bp) promoters interact when expressed in mouse adrenal Y-1 cells (Takayama *et al.*, 1994). Since these elements flank the *CYP11a* (tttta)n polymorphism (which is present at position -466 (refer to **figure 4.1**)), this system may be used to test any differences in transactivation between expression vector constructs containing different *CYP11a* (tttta)n alleles. The proposed experimental outline is as follows:

- DNA from individuals homozygous for each of the 4 *CYP11a* (tttta)n alleles will be selected and used to PCR amplify a 2kb fragment of the *CYP11a* 5' promoter region (eg. from +49bp to -1950bp from start of transcription).

- A variety of expression vector constructs will be generated that carry the different PCR amplified *CYP11a* upstream regions fused to the Luciferase reporter gene.

- As a control, a reporter construct will also be generated that contains the *CYP11a* upstream region but with the *CYP11a* (tttta)n polymorphism absent. This may be generated by excising the *CYP11a* (tttta)n polymorphism from one of the above expression vectors using unique flanking restriction enzyme sites.

- The transcriptional activity of each construct will be assessed by transfecting into mouse adrenal Y-1 tumour cells. A direct comparison between the activity of the different constructs may then be made.

Since both tissue and species-specific differences in *CYP11a* promoter element usage have been demonstrated (Rodriguez *et al.*, 1997), it is also important to evaluate the role of the *CYP11a* (tttta)n polymorphism in human ovarian cells. One approach would be to repeat the above experiments using a human ovarian tumour cell-line such as the human ovarian thecal-like tumour (HOTT) cells (Rainey *et al.*, 1996). This cell-line has been shown to be responsive to forskolin (but not to LH), producing high levels of progesterone and 17 α-hydroxyprogesterone as well as C19 steroids. Forskolin treatment has also been shown to induce the expression of the genes of a

variety of steroidogenic enzymes including *CYP11a*. A limitation of such an experiment is that a putative ovarian specific (distal) promoter element may be located outside the 2kb region to be tested.

4.7.4 *SRD5A1*

Future work on *SRD5A1* will initially involve confirmation of the preliminary results indicating a role for this gene in the development of hirsutism (i.e. its action on hair follicles in the skin). Also the possibility of a role for this gene in the aetiology of PCO *per se* needs to be investigated.

5 α-reductase type 1 and 2 isozymes are both expressed in ovarian granulosa and thecal cells. It has recently been demonstrated that, although there was no difference in the pattern or relative amount of mRNA expression between ovarian follicles from PCOS women compared with those from normal controls, the total 5 α-reductase activity (predominantly the 5 α-reductase type 1) was 4-fold higher in the PCOS follicles (Jakimiuk *et al.*, 1999). Therefore, a role for 5 α-reduced androgens in the pathogenesis of PCOS has been proposed.

Ideally these studies will be carried out using a family-based association data set (eg. a TDT data set).

- To confirm a role for this gene in the development of hirsutism, a hirsute TDT data set is required. The PCOS TDT/sibpair data set (**section 4.7.1**) may be used for this, if sufficient numbers of hirsute subjects have been collected. Otherwise, a separate hirsute data set will need to be collected.

- The TDT and sibpair data sets can also be used to evaluate the role of this gene in the aetiology of PCO/PCOS.

- Mutation screening of the *SRD5A1* gene may also be carried out in hirsute PCOS subjects carrying the HinfI(+)/NspI(+) haplotype. The *SRD5A1* contains five exons and four introns spanning over 35kb of DNA. The length of exons varies from 102bp to 1359bp and those of the introns vary from 4.1kb to over 14kb of DNA. The sequence of the exons and exon-intron boundaries as well as the 5'-flanking and the 3'-untranslated regions have been published (Jenkins *et al.*, 1991). This should allow for the design of a PCR based direct sequencing assay, to screen each of the five exons for mutations. Identification of any base changes may then be used to extend the haplotype analysis and may allow LD mapping of the aetiological mutation.

4.8 CONCLUSIONS

In this thesis, a candidate gene approach has been taken to study the molecular genetic basis of PCOS, a common complex trait disorder. This, in combination with the different methodological approaches used (linkage and association strategies), has proved successful in yielding positive results. Linkage analysis has been carried out using both parametric and non-parametric methods, which were implemented using the GENEHUNTER program. This approach has provided convincing evidence for the involvement of *CYP11a* in the aetiology of PCOS and has excluded the genes *CYP19* and *LHR* as major susceptibility loci. Furthermore the linkage results have indicated the presence of genetic heterogeneity, with about 60% of the pedigree demonstrating linkage to the *CYP11a* locus.

Association studies are important complementary approaches to linkage analysis in the study of complex traits and are considered more appropriate than linkage for the study of genes of modest effect (Lander and Kruglyak, 1995; Risch and Merikangas, 1996). Case control association studies were used to assess the role of two genes as

modifiers of the phenotypic expression of PCOS. The results revealed that the *CYP17* gene is unlikely to have a direct role in modulating serum androgen levels. However, preliminary results have provided some evidence for the involvement of the *SRD5A1* gene, which encodes the enzyme 5 alpha reductase, in the aetiology of hirsutism in PCOS.

Mutation screening of the *CYP11a* promoter region revealed the polymorphic nature of the *CYP11a* (tttta)n pentanucleotide repeat sequence. Case control association studies were subsequently used to confirm a role for *CYP11a* in the aetiology of hyperandrogenaemia in women with PCOS. The association results have indicated that the higher repeat length alleles of the *CYP11a* (tttta)n are associated with elevated mean serum testosterone levels. This polymorphism is present in the 5' regulatory region of the *CYP11a* gene and is flanked by numerous promoter elements involved in the transactivation of the *CYP11a* gene. Although it is most likely that the association results obtained are as a consequence of LD between this polymorphism and another (causative) mutation located elsewhere in the *CYP11a* gene, a direct role for this VNTR in the up-regulated expression of the *CYP11a* gene has been proposed.

The *in vitro* study of Nelson and co-workers has confirmed that increased steroidogenesis is an intrinsic abnormality in PCOS theca cells (Nelson *et al.*, 1999). Up-regulation of a number of steroidogenic enzymes, including *CYP11a*, has been demonstrated. This may indicate either that variation in the *CYP11a* gene is the primary defect (which subsequently leads to a secondary general up-regulation of steroidogenic genes) or it may be acting epistatically with a putative *trans*-acting factor, possibly a component of a signal transduction pathway, which is a separate susceptibility locus.

In conclusion, *CYP11a,* encoding the enzyme P450scc, the first enzymatic step in steroid hormone biosynthesis, has been shown to be an inherited susceptibility locus in the aetiology of hyperandrogenism and PCOS. This is one of the first

susceptibility loci to be identified for PCOS. Other genes and loci are now beginning to emerge from studies carried out in our laboratory (*INS* VNTR (Waterworth *et al.*, 1997a)) and by the work of others (e.g. the chromosome 5p14 region containing the follistatin gene (Urbanek *et al.*, 1999)). It is hoped that once a powerful enough patient resource has been collected and the appropriate statistical and technological capabilities are available, that the interaction of these and other loci may be evaluated. This should allow the elucidation of the complex aetiological basis of this disorder. Ultimately, this is hoped to lead to better therapy for the symptomatology of PCOS and intervention of the long-term health risks that are associated with this common disorder.

5.0 REFERENCES

Abdel-Gadir, A., Mowafi, R.S., Alnaser, H.M., Alrashid, A.H., Alonezi, O.M. and Shaw, R.W. (1990). Ovarian electrocautery versus human menopausal gonadotrophins and pure follicle stimulating hormone therapy in the treatment of patients with polycystic ovarian disease. *Clin. Endocrinol. (Oxf)*, **33,** 585-592.

Adams, J., Franks, S., Polson, D.W., Mason, H.D., Abdulwahid, N., Tucker, M., Morris, D.V., Price, J. and Jacobs, H.S. (1985). Multifollicular ovaires:clinical and endocrine features to pulsatile gonadotrophin releasing hormone. *Lancet*, **2,** 1375-1379.

Adams, J., Polson, D.W., Franks,S. (1986). Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *BMJ*, **293,** 355-359.

Baird, D.T., Corker, C.S., Davidson, D.W., Hunter, W.M., Michie, E.A., Van Look, P.F.A. (1977). Pituitary-ovarian relationships in polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **45,** 798-801.

Bardin, C.W and Lipsett, M.B. (1967). Testosterone and androstenedione blood production rates in normal women and women with idiopathic hirsutism or polycystic ovaries. *J. Clin. Invest.*, **46,** 891-902.

Barnes, R.B., Rosenfield, R.L., Burstein, S., Ehrmann, D.A. (1989). Pituitary-ovarian responses to nafarelin testing in the polycystic ovary syndrome. *N. Eng. J. Med.*, **320,** 559-565.

Barnes, R.B., Rosenfield, R.L., Ehrmann, D.A., Cara, J.F., Cuttler, L., Levitsky, L.L. and Rosenthal, I.M. (1994). Ovarian hyperandrogenism as a result of congential adrenal virilizing disorders: evidence for perinatal masculinization of neuroendocrine function in women. *J. Clin. Endocrinol. Metab.*, **79,** 1328-1333.

Barth, J.H. (1988). Alopecia and Hirsutism: current concepts in pathogenesis and management. *Drugs*, **35,** 83-91.

Beckmann, J.S., Tomfohrde, J., Barnes, R.I., Williams, M., Broux, O., Richard, I., Weissenbach, J., Bowcock, A.M. (1993). A linkage map of human chromosome 15 with an average resolution of 2cM and containing 55 polymorphic microsatellites. *Hum. Mol. Genet.*, **2,** 2019-2030.

Bennett, S.T., Lucassen, A.M., Gough, S.C.L., Powell, D.E., Undlien, D.E., Pritchard, L.E., Merriman, M.E., Kawaguchi, Y., Dronsfield, M.J., Pociot, F., Nerup, J., Bouzekri, N., Cambon-Thomsen, A., Ronnigen, K.S., Barnett, A.H., Bain, S.C., Todd, J.A. (1995). Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature Genet.*, **9,** 284-292.

Bennett, S.T., Todd, J.A., Waterworth, D.M., Franks, S. and McCarthy, M.I. (1997). Association of insulin gene VNTR polymorphism with polycystic ovary syndrome (letter). *Lancet*, **349,** 1771-1772.

Bennett, S.T, and Todd, J.A. (1996). Human type I diabetes and the insulin gene: principles of mapping polygenes. *Annu. Rev. Genet.*, **30.**

Bland, J.M, and Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *BMJ.*, **310,** 170.

Blankstein, J., Rabinovici, J., Goldenberg, M., Shaley, J., Mehta, A., Serr, D.M. and Mashiach, S. (1987). Changing pituitary reactivity to follicle-stimulating hormone and luteinizing hormone-releasing hormone after induced ovulatory cycles and after anovulation in patients with polycystic ovarian disease. *J. Clin. Endocrinol. Metab.*, **65,** 1164-1167.

Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. and Bue, H. (1989). Synthetic curved DNA sequences can act as transcriptional activators in Escherichia coli. *EMBO J.*, **8,** 4289-4296.

Brown, J.B. (1978). Pituitary control of ovarian function - concepts derived from gonadotrophin therapy. *Aust. N. Z. J. Obstet. Gynaecol.*, **18,** 46-54.

Bunker, C.B., Newton, J.A., Conway, G.S., Jacobs, H.S., Greaves, M.W. and Dowd, P.M. (1991). The hormonal profile of women with acne and polycystic ovaries. *Clin. Exp. Dermatol.*, **16,** 420-423.

Carey, A.H., Chan, K.L., Short, F., White,D., Williamson, R. & Franks, S. (1993). Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin. Endocrinol.*, **38,** 653-658.

Carey, A.H., Waterworth, D.M., Patel, K., White, D., Little, J., Novelli, P., Franks, S., Williamson, R. (1994). Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum. Mol. Genet.*, **3,** 1873-1876.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalkanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q. and Lander, E.S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.*, **22,** 231-238.

Castello, R., Tosi, F., Perrone, F., Negri, C., Muggeo, M. and Moghetti, P. (1996). Outcome of long-term treatment with the 5 alpha-reductase inhibitor finasteride in Idiopathic hirsutism:clinical and hormonal effects during a 1-year course of therapy and 1-year follow up. *Fertil. Steril.*, **66,** 734-740.

Chadha, S., Pache, T.D., Huikeshoven, J.M., Brinkmann, A.O. and van der Kwast, T.H. (1994). Androgen receptor expression in human ovarian and uterine tissue of long-term androgen-treated transsexual women. *Hum. Pathol.*, **25**.

Chang, R.J., Nakamura, R.M., Judd, H.L., Kaplan, S.A. (1983a). Insulin resistance in nonobese patients with polycystic ovary syndrome *J. Clin. Endocrinol. Metab.*, **57,** 356-359.

Chang, R.J., Laufer, L.R., Meldrum, D.R., DeFazio, J., Lu, J.K., Vale, W.W., Rivier, J.E. and Judd, H.L. (1983b). Steroid secretion in polycystic ovarian disease after ovarian suppression by a long-acting gonadotropin-releasing hormone agonist. *J. Clin. Endocrinol. Metab.*, **56,** 897-903.

Chung, B.-C., Matteson, K.J., Voutilainen, R., Mohandas, T.K. and Miller, W.L. (1986). Human cholesterol side-chain cleavage enzyme, P450scc: cDNA cloning, assignment of the gene to chromosome 15, and expression in the placenta. *Proc. Natl, Acad. Sci.*, **83,** 8962-8966.

Chung, B.-C., Hu, M-C, Lai, C-C and Lin, C-H. (1989). The 5'-region of the P450XIA1 (P450scc) gene contains a basal promoter and an adrenal-specific activating domain. *Biochem. Biophys. Res. Comm.*, **160,** 276-281.

Ciaraldi, T.P., El-Roeiy, A., Madar, Z., Reichart, D., Olefsky, J.M., Yen, S.S.C. (1992). Cellular mechanisms of insulin resistance in polycystic ovarian syndrome. *J. Clin. Endocrinol. Metab.*, **75,** 577-583.

Clayton, R.C., Ogden, V., Hodgkinson, J.,. (1992). How common are polycystic ovaries in normal women and what is their significance for the fertility in the general population ? *Clin. Endocrinol.*, **37,** 127-134.

Clemens, J.W., Lala, D.S., Parker, K.L. and Richards, J.S. (1994). Steroidogenic factor-1 binding and transcriptional activity of the cholesterol side-chain cleavage promoter in rat granulosa cells. *Endocrinol.*, **134,** 1499-1508.

Clerget-Darpoux, F., Bonaiti-Pellie, C. and Hochez, J. (1986). Effects of misspecifying genetic parameters in lod score analysis. *Biometrics*, **42,** 393-399.

Conte, F.A., Grumbach, M.M., Ito, Y., Fisher, C.R. and Simpson, E.R. (1994). A syndrome of female pseudohermaphroditism, hypergonadotropic hypogonadism, and multicystic ovaries associated with missense mutations in the gene encoding aromatase (P450arom). *J. Clin. Endocrinol. Metab.*, **78,** 1287-1292.

Conway, G.S., Honour, J.W. and Jacobs H.S. (1989). Heterogeneity of the polycystic ovary syndrome: clinical, endocrine and ultrasound features in 556 patients. *Clin. Endocrinol. (Oxf)*, **30,** 459-470.

Conway, G.S., Jacobs, H.S., Holly, J.M.P., and Wass, J.A.H. (1990). Effects of luteinizing hormone, insulin, insulin-like growth factor-1 and insulin-like growth factor small binding protein 1 in the polycystic ovary syndrome. *Clin. Endocrinol.*, **33,** 593-603.

Conway, G.S., Agrawal, R., Betteridge, D.J., and Jacobs, H.S. (1992). Risk factors for coronary artery disease in lean and obese women with the polycystic ovary syndrome. *Clin. Endocrinol.*, **37,** 119-125.

Cooper, H.E., Spellacy, W.N., Prem, K.A. and Cohen, W.D. (1968). Hereditary factors in the Stein-Leventhal syndrome. *Am. J. Obstet. Gynecol.*, **100,** 371-387.

Cottingham, R.W., Idury, R.M., and Schaffer, A.A. (1993). Faster sequential genetic linkage computations. *Am. J. Hum. Genet.*, **53,** 252-263.

Couzinet, B., Le Strat, N., Brailly, S. and Schaison, G. (1986). Comparative effects of cyproterone acetate or a long-acting gonadotropin-releasing hormone agonist in plycystic ovarian disease. *J. Clin. Endocrinol. Metab.*, **63,** 1031-1035.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. and Myers, R.M. (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, **250,** 245-250.

Cox, N.J. and Bell, G.I. (1989). Disease associations, Chance, Artifact, or susceptibility genes? *Diabetes*, **38,** 947-950.

Cusan, L., Dupont, A., Gomez, J.L., Tremblay, R.R. and Labrie, F. (1994). Comparison of flutamide and spironolactone in the treatment of hirsutism: a randomized controlled trial. *Fertil. Steril.*, **61,** 281-287.

Dahlgren, E., Janson, P.O., Johansson, S., Lapidus, L., Oden, A. (1992a). Polycystic ovary syndrome and risk for myocardial infarction: evaluated from a risk factor model based on a prospective study of women. *Acta Obstet Gynecol Scand*, **71,** 599-604.

Dahlgren, E., Johannson, S., Lindstedt, G. (1992b). Women with polycystic ovary syndrome wedge resected in 1956 to 1965: a long term follow up focusing on natural history and circulating hormones. *Fertil. Steril.*, **57,** 505-513.

Day, C. (1999). Thiazolidinediones: a new class of antidiabetic drugs. *Diabet. Med.*, **16,** 179-192.

Diamanti-kandarakis, E., Kouli, C., Tsianateli, T. and Bergiele, A. (1998). Therapeutic effects of metformin on insulin resistance and hyperandrogenism in polycystic ovary syndrome. *Eur. J. Endocrinol.*, **138,** 269-274.

Diamanti-Kandarakis, E., Bartzis, M., Bergiele, A., Tsianateli, T. and Kouli, C. (1999). Association of a gene CYP11a microsatellite polymorphism with polycystic ovary syndrome (PCOS). In *81st Annual Meeting of The Endocrine Society*: San Diego, California.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380,** 152-154.

Diekmann, S. (1992). Analyzing DNA curvature in polyacrylamide gels. *Methods Enzymol.*, **212,** 30-46.

Dizier, M.-H., Bonaiti-Pellie, C. and Clerget-Darpoux, F. (1993). Conclusions of segregation analysis for family data generated under two-locus models. *Am. J. Hum. Genet.*, **53,** 1338-1346.

Donesky, B.W. and Adashi, E.Y. (1995). Surgically induced ovulation in the polycystic ovary syndrome: wedge resection revisited in the age of laparoscopy. *Fertil. Steril.*, **63,** 439-463.

Dunaif, A., Scully, R.E., Andersen, R.N., Chapin, D.S. and Crowley, W.F. Jr. (1984). The effects of continuous androgen secretion on the hypothalamic-pituitary axis in woman: evidence from a luteinized thecoma of the ovary. *J. Clin. Endocrinol. Metab.*, **59,** 389-393.

Dunaif, A., Graf, M., Mandeli, J., Laumas, V., and Dobrjansky, A. (1987). Characterization of groups of hyperandrogenic women with acanthosis nigricans, impaired glucose tolerance, and/or hyperinsulinaemia. *J. Clin. Endocrinol. Metab.*, **65,** 499-507.

Dunaif, A., Segal, K.R., Futterweit, W., and Dobrjansky, A. (1989). Profound peripheral insulin resistance, independent of obesity, in polycystic ovary syndrome. *Diabetes*, **38,** 1165-1174.

Dunaif, A., Segal, K.R., Shelley, D.R., Green, G., Dobrjansky, A., and Licholai, T. (1992a). Evidence for distinctive and intrinsic defects in action in polycystic ovary syndrome. *Diabetes*, **41,** 1257-1266.

Dunaif, A. (1992b). Insulin resistance and ovarian hyperandrogenism. *The Endocrinologist*, **2,** 248-260.

Dunaif, A., Xia, J., Book, C.B., Schenker, E., and Tang, Z. (1995). Excessive insulin receptor serine phosphorylation in cultured fibroblasts and skeletal muscle: a potential

mechanism for insulin resistance in the polycystic ovary syndrome. *J. Clin. Invest.*, **96,** 801-810.

Dunaif, A., Scott, D., Finegood, D., Quintana, B. and Whitcomb, R. (1996a). The insulin sensitizing agent troglitazone improves metabolic and reproductive abnormalities in the polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **81,** 3299-3306.

Dunaif, A. (1997). Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr. Rev.*, **18,** 774-800.

Dunaif, A. and Finegood, D.T. (1996b). β-cell dysfunction independent of obesity and glucose intolerance in the polycystic ovary syndrome *J. Clin. Endocrinol. Metab.*, **81,** 942-947.

Durner, M., Vieland, V.J. and Greenberg, D. (1999). Further evidence for the increased power of LOD scores compared with nonparametric methods. *Am. J. Hum. Genet.*, **64,** 281-289.

Ehrmann, D.A., Sturis, J., Byrne, M.M., Karrison, T., Rosenfield, R.L. and Polonksky, K.S. (1995). Insulin secretory defects in polycystic ovary syndrome. *J. Clin. Invest.*, **96,** 520-527.

Ehrmann, D.A., Cavaghan, M.K., Imperial, J., Sturis, J., Rosenfield, R.L. and Polonsky, K.S. (1997a). Effects of metformin on insulin secretion, insulin action, and ovarian steroidogenesis in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **82,** 524-530.

Ehrmann, D.A., Schneider, D.J., Sobel, B.E., Cavanagh, M.K., Imperial, J., Rosenfield, R.L. and Polonsky, K.S. (1997b). Troglitazone improves defects in insulin action, insulin secretion, ovarian steroidogenesis, and fibrinolysis in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **82,** 2108-2116.

Erickson, G.F., Wang, C. and Hsueh, A.J. (1979). FSH induction of LH receptors in granulosa cells cultured in a chemically defined medium. *Nature*, **279,** 336-338.

Evans, D., Hoffmann, R.G., Kalkhoff, R.K. and Kissebah, A.H. (1984). Relationship of body fact topography to insulin sensitivity and netabolic profiles in premenopausal women. *Metabolism*, **33,** 68-75.

Fasetti, L., De Fusco, D., Leleftheriou, G. and Rosina, B. (1997). Treatment of hirsutism by finasteride and flutamide in women with polycystic ovary syndrome. *Gynecol. Endocrinol.*, **11,** 251-257.

Fauser, B.C., Pache, T.D., Hop, W.C., de Jong, F.H. and Dahl, K.D. (1992). The significance of a single serum LH measurement in women with cycle disturbances: discrepancies between immunoreactive and bioactive hormone estimates. *Clin. Endocrinol. (Oxf)*, **37,** 445-452.

Ferriman, D. and Gallwey, J. (1961). Clinical assessment of hair growth in women. *J. Clin. Metab.*, **21,** 1440-1445.

Ferriman, D. and Purdie, A.W. (1979). The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin. Endocrinol. (Oxf)*, **11,** 291-300.

Flier, J.S., Eastman, R.C., Minaker, K.L., Matteson, D. and Rowe, J.W. (1985). Acanthosis nigricans in obese women with hyperandrogenism: characterization of an insulin-resistant state distinct from type A and B syndromes. *Diabetes*, **34,** 101-107.

Fox, R., Corrigan, E., Thomas, P.A. and Hull, M.G.R. (1991). The diagnosis of polycystic ovaries in women with oligo-amenorrhoea: predictive power of endocrine tests. *Clin. Endocrinol.* (Oxf), **34,** 127-131.

Franks, S., Adams, J., Mason, H. and Polson, D. (1985). Ovulatory disorderd in women with polycystic ovary syndrome. *Clin. Obstet. Gynecol.*, **12,** 605-632.

Franks, S. (1989). Polycystic ovary syndrome: a changing perspective. *Clin. Endocrinol.*, **31,** 87-120.

Franks, S. (1991). The ubiquitous polycystic ovary. *J. Endocrinol.*, **129,** 317-319.

Franks, S. (1995). Polycystic ovary syndrome. *N. Eng. J. Med.*, **333,** 853-861.

Franks, S., Gharani, N., Waterworth, D., Batty, S., White, D., Williamson, R. and McCarthy, M. (1997). The genetic basis of polycystic ovary syndrome. *Hum. Reprod.*, **12,** 2641-2648.

Futterweit, W., Dunaif, A., Yeh, H.C. and Kingsley, P. (1988). The prevalence of hyperandrogenism in 109 consecutive female patients with diffuse alopecia. *J. Am. Acad. Dermatol.*, **19,** 831-836.

Futterweit, W. and Deligdisch, L. (1986). Histopathological effects of exogeneously administered testosterone in 19 female to male transsexuals. *J. Clin. Endocrinol. Metab.*, **62,** 16-21.

Gartenberg, M.R and Crothers, D.M. (1991). Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. *J. Mol. Biol.*, **219,** 217-230.

Gharani, N., Waterworth, D.M., Williamson, R. and Franks, S. (1996). 5' polymorphism of the CYP17 gene is not associated with serum testosterone levels in women with polycystic ovaries [letter]. *J. Clin. Endocrinol. Metab.*, **81,** 4174.

Gharani, N., Waterworth, D.M., Batty, S., White, D., Gilling-Smith, C., Conway, G.S., McCarthy, M.I., Franks, S. and Williamson, R. (1997). Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum. Mol. Genet.*, **6,** 397-402.

Gilling-Smith, C., Willis, D.S., Beard, R.W. and Franks, S. (1994). Hypersecretion of androstenedione by isolated thecal cells from polycystic ovaries. *J. Clin. Endocrinol. Metab.*, **79,** 1158-1165.

Gilling-Smith, C., Story, E.H. and Franks, S. (1997). Evidence for a primary abnormality of thecal cell steroidogenesis in the polycystic ovary syndrome. *Clin. Endocrinol.*, **47,** 93-99.

Givens, J.R. (1988). Familial polycystic ovarian disease. *Endocrinol. Metab. Clin. North Am.*, **17,** 771-783.

Gormley, G.J. (1995). Finasteride: a clinical review. *Biomed. Pharmacother.*, **49,** 319-324.

Gougeon, A. (1986). Dynamics of follicular growth in the human: a model from preliminary results. *Hum. Reprod.*, **1,** 81-87.

Gourse, R.L., de Boer, H.A. and Nomura, M. (1986). DNA determinants of rRNA synthesis in E. coli: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell*, **44,** 197-205.

Graf, M.J., Richards, C.J., Brown, V., Meissner, L. and Dunaif, A. (1990). The independent effects of hyperandrogenaemia, hyperinsulinaemia and obesity in lipid and lipoprotein profiles in women. *Clin. Endocrinol.*, **33,** 119-131.

Greenberg, D.A. (1989a). Inferring mode of inherirtance by comparison of lod scores. *Am. J. Med. Genet.*, **34,** 480-486.

Greenberg, D.A., Abreu, P. and Hodge, S.E. (1998). The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am. J. Hum. Genet.*, **63,** 870-879.

Greenberg, D.A and Hodge, S.E. (1989b). Linkage analysis under "random" and "genetic" reduced penetrance. *Genetic Epidemiol.*, **6,** 259-264.

Guo, Q., Kuma, T.R. Woodruff, T., Hadsell, L.A., DeMayo, F.J. and Matzuk, M.M. (1998). Overexpression of mouse follistatin causes reproductive defects in transgenic mice. *Mol. Endocrinol.*, **12,** 96-106.

Gyapay, G., Morrisette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994). The 1993-1994 Genethon human genetic linkage map. *Nature Genet.*, **7,** 246-339.

Hague, W.M., Adams, J., Reeders, S.T., Peto, T.E.and Jacobs, H.S. (1988). Familial polycystic ovaries: a genetic disease? *Clin. Endocrinol. (Oxf)*, **29,** 593-605.

Halkerston, I.D.K., Eichhorn, J. and Hechter, O. (1961). A requirement for reduced triphosphopyridine nucleotide for cholesterol side-chain cleavage by mitochondrial fractions of bovine adrenal cortex. *J. Biol. Chem.*, **236,** 374.

Harada, n., Ogawa, H., Shozu, M. and Yamada,K. (1992). Genetic studies to characterise the origin of the mutation in placental aromatase deficiency. *Am. J. Hum. Genet.*, **51,** 666-667.

Hill, W.G. and Weir, B.S. (1994). Maximum-likelihood estimation of gene location by linkage disequilibrium. *Am. J. Hum. Genet.*, **54,** 705-714.

Hillier, S., Whitelaw, P. and Smyth, C. (1994). Sex steroid metabolism in the ovary. *The Biochemist*, 22-26.

Hillier, S.G. (1991). 'Cellular basis of follicular endocrine function'. In *Ovarian Endocrinology*, Hillier, S.G. (ed) pp. 1-24. Blackwell Scientific Publications: Oxford.

Hodge, S.E. and Elston, R.C. (1994). Lods, wrods, and mods: The interpretation of lod scores calculated under different models. *Genet. Epidemiol.*, **11,** 329-342.

Holte, J., Bergh, T., Berne, C., Berglund, L. and Lithell, H. (1994a). Enhanced early insulin response to glucose in relation to insulin resistance in women with polycystic

ovary syndrome and normal glucose tolerance. *J. Clin. Endocrinol. Metab.*, **78,** 1052-1058.

Holte, J., Bergh, T., Gennarelli, G. and Wide, L. (1994b). The independent effects of polycystic ovary syndrome and obesity on serum concentrations of gonadotrophins and sex steroids in premenopausal women. *Clin. Endocrinol.*, **41,** 473-481.

Holte, J., Bergh, T., Berne, C., Wide, L. and Lithell, H. (1995). Restored insulin sensitivity but persistently increased early insulin secretion after weight loss in obese women with polycystic ovary syndrome *J. Clin. Endocrinol. Metab.*, **80,** 2586-2593.

Holte, J. (1996). Disturbance in insulin secretion and sensitivity in women with polycystic ovary syndrome. *Bailliere's Clin. Endocrinol. Metab.*, **10,** 221-247.

Hughesdon, P.E. (1982). Morphology and morphogenesis of the Stein-Leventhal ovary and of so-called 'hyperthecosis'. *Obstet. Gynecol. Surv.*, **37,** 59-77.

Hull, M.G. (1987). Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. *Gynecol. Endocrinol.*, **1,** 235-245.

Hull, M.G. (1992). The causes of infertility and relative effectiveness of treatment. In *Infertility*, Templeton A.A., Drife, J.O., (ed) pp. 33-62. Springer-Verlag: Lomdon

Hum, D.W., Staels, B., Black, S.M. and Miller, W.L. (1993). Basal transcriptional activity and cyclic AMP responsiveness of the human P450scc promoter transfected into mouse Leydig MA-10 cells. *Endocrinol.*, **132,** 546-552.

Huxtable, S.J., Saker, P.J., Haddad, L., Walker, M., Frayling, T.M., Levy, J.C., Hitman, G.A., O'Rahilly, S., Hattersley, A.T. and McCarthy, M.I. (2000). Analysis of parent-offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally-transmitted class III variable number tandem repeat alleles. *Diabetes*, **49,** in press.

Inoue, H., Watanabe, N., Higashi, Y. and Fujii-Kuriyama, Y. (1991). Structure of regulatory regions in the human cytochrome P-450scc (desmolase) gene. *Eur. J. Biochem.*, **195,** 563-569.

Ito, Y., Fisher, C.R., Conte, F.A., Grumbach, M.M. and Simpson, E.R. (1993). Molecular basis of aromatase deficiency in an adult female with sexual infantilism and polycystic ovaries. *Proc. Natl. Acad. Sci. USA*, **90,** 11673-11677.

Jacobs, H.S. (1994). Pulsatile GnRH in multifollicular and polycystic ovary patients. In *Ovulation induction: basic science and clinical advances.*, Filicori, M., Flamigni, C. (ed) pp. 267-274. Excerpta Medica: Amsterdam.

Jakimiuk, A.J., Weitsman, S.R. and Magoffin, D.A. (1999). 5 α-reductase activity in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **84,** 2414-2418.

Jenkins, E.P., Hsieh, C-L., Milatovich, A., Normington, K., Berman, D. M., Francke, U., and Russell, D. W. (1991). Characterization and chromosomal mapping of a human steroid 5 α-reductase gene and pseudogene and mapping of the mouse homologue. *genomics*, **11,** 1102-1112.

Jialil, I., Naiker, P., Reddi, K., Moodley, J. and Joubert, S.M. (1987). Evidence for insulin resistance in nonobese patients with polycystic ovarian disease. *J. Clin. Endocrinol. Metab.*, **64,** 1066-1069.

Kennedy, G.C., German, M.S., Rutter, W.J. (1995). The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nature Genetics*, **9,** 293-298.

Kiddy, D.S., Hamilton-Fairly, D., Seppala, M., Koistinen, R., James, V.H.T., Reed, M.H. and Franks, S. (1989). Diet-induced changes in sex hormone binding globulin and free testosterone in women with polycystic ovaries: correlation with serum insulin and insulin-like growth factor I. *Clin. Endocrinol. (Oxf)*, **31,** 757-763.

Kiddy, D.S., Sharp, P.S., White, D.M., Scanlon, M.F., Mason, H.D., Bray, C.S., Polson, D.W., Reed, M.J. and Franks, S. (1990). Differences in clinical and endocrine features between obese and non-obese subjects with polycystic ovary syndrome: An analysis of 263 consecutive cases. *Clin. Endocrinol.*, **32,** 213-220.

Kiddy, D.S., Hamilton-fairley, D., Bush, A., Short, F., Anyaoku, V., Reed, M.J., Franks, S. (1992). Improvement in endocrine and ovarian function during dietary treatment of obese women with polycystic ovary syndrome. *Clin. Endocrinol.*, **36,** 105-111.

Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P. (1993). Crystal of a yeast TBP/TATA-box complex. *Nature*, **365,** 512-520.

Kirschner, M.A., Zucker, I.R. and Jespersen,D. (1976). Idiopathic hirsutism - an ovarian abnormality. *N. Engl. J. Med.*, **294,** 637-640.

Kirschner, M.A., Samojlik, E. and Silber, D. (1983). A comparison of androgen production and clearance in hirsute and obese women. *J. Steroid Biochem.*, **19,** 607-614.

Kirschner, M.A., and Bardin, C.W. (1972). Androgen production and metabolism in normal and virilized women. *Metabolism*, **21,** 667-688.

Koo, H.-S., Wu, H-M. and Crothers D.M. (1986). DNA bending at adenine.thymine tracts. *Nature*, **320,** 501-506.

Kremer, H., Kraaij, R., Toledo, S.P.A., Post, M., Fridman, J.B., Hayashida, C.Y., van Reen, M., Milgrom, E., Ropers, H.H., Mariman, E., Themmen, A.P.N., Brunner, H.G. (1995). Male pseudohermaphoditism due to a homozygous missense mutation of the luteinizing hormone receptor gene. *Nature Genetics*, **9,** 160-164.

Krontiris, T.G., Devlin, B., Karp, D.D., Robert, N.J. and Rich, N. (1993). An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N. Engl. J. Med.*, **329,** 517-.

Krontiris, T.G. (1995). Minisatellites and human disease. *Science*, **269,** 1682-1683.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996). Parametric and non-parametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58,** 1347-1363.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.*, **22,** 139-144.

Lalouel, J.M. and Yee, S. (1980). POINTER: a computer program for complex segregation analysis pointers. Tech rep, Population Genetics Laboratory, University of Hawaii, Honolulu.

Lamond, A.T. and Travers, A.A. (1983). Requirement for an upstream element for optimal transcription of a bacterial tRNA gene. *Nature*, **305**, 248-250.

Lander, E. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.*, **11,** 241-247.

Lander, E.S. and Green, P. (1987). Construction of multilocus genetic maps in humans. *Proc. Natl. Acad. Sci. USA*, **84,** 2363-2367.

Lander, E.S. and Schork, N.J. (1994). Genetic dissection of complex traits. *Science*, **265,** 2037-2048.

Lathrop, G.M., Lalouel, J.M. (1984a). Easy calculations of LOD scores and genetic risks on small computers. *Am. J. Hum. Genet.*, **36,** 460-465.

Lathrop, G.M., Lalouel, J.M., Julier, C., Ott, J. (1984b). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci.*, **81,** 3443-3446.

Lathrop, G.M., Lalouel, J.M., White, R.L. (1986). Construction of human genetic linkage maps: likelihood calculations for multilocus analysis. *Genet. Epidemiol.*, **3,** 39-52.

Laue, L., Chan, W.-Y., Hsueh, A.J.W., Kudo, M., Hsu, S. Y. Wu, S.-M., Blomberg, L. and Cutler,G.B. (1995). Genetic heterogeneity of constitutively activating mutations of human luteinizing hormone receptor in familial male-limited precocious puberty. *Proc. Natl. Acad. Sci.*, **92,** 1906-1910.

Legro, R.S. (1995). The genetics of polycystic ovary syndrome. *Am. J. Med.*, **98,** 9S-16S.

Legro, R.S., Spielman, R., Urbanek, M., Driscoll, D., Strauss III, J.F. and Dunaif, A. (1998). Phenotype and Genotype in polycystic ovary syndrome. *Recent Prog. Horm. Res.*, **53,** 217-256.

Lerman, L.S. and Frisch, H.L. (1982). Why does the electrophoretic mobility of DNA in gels vary with the length of the molecule? *Biopolymers*, **21,** 995-997.

Lesko, S.M., Rosenberg, L. and Sharpiro, S. (1993). A case control study of baldness in relation to myocardial infarction in men. *Proc. Natl. Acad. Sci., USA*, **81,** 3443-3446.

Liu, W., Qian, C. and Francke, U. (1997a). Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nature Genet.*, **16,** 328-329.

Liu, Z. and Simpson, E.R. (1997b). Steroidogenic Factor 1 (SF1) and SP1 are required for regulation of bovine CYP11a gene expression in bovine luteal cells and adrenal Y1 cells. *Mol. Endocrinol.*, **11,** 127-137.

Lobo, R.A. (1991). Hirsutism in polycystic ovary syndrome: current concepts. *Clin. Obstet. Gynecol.*, **34,** 817-826.

Lundblad, J.R., Kwok, R.P.S., Laurance, M.E., Harter, M.L. and Goodman. (1995). Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. *Nature*, **374,** 85-88.

Lunde, O., Magnus, P., Sandvik, L. and Hoglo, S. (1989). Familial clustering in the polycystic ovarian syndrome. *Gynecol. Obstet. Invest.*, **28,** 23-30.

Lunenfeld, B. and Insler,V. (1978). *Diagnosis and treatment of functional infertility.* Gross Verlag.: Berlin.

Mahendroo M.S, M., G.D. Mendelson, C.R. and Simpson, E.R. (1991). Tissue-specific expression of Human $P450_{arom}$:the promoter responsible for expression in adipose is different from that utilized in the placenta. *J. Biol. Chem.*, **266,** 11276-11281.

Mason, H.D., Willis, D.S., Beard, R.W., Winston, R.M.L., Margara, R. and Franks, S. (1994). Estradiol production by Granulosa cells of normal and polycystic ovaries: Relationship to menstrual cycle history and concentrations of gonadotropins and sex steroids in follicular fluid. *J. Clin. Endocrinol. Metab.*, **79,** 1355-1360.

Mather, J.P., Moore, A. and Li, R.H. (1997). Activins, inhibins and follistatins: further thoughts on a growing family of regulators. *Proc. Soc. Exp. Biol. Med.*, **215,** 209-222.

McAllister, C.F. and Archberger, E.C. (1988). Effects of polyadenine-containing curved DNA on promoter utilization in Bacillus subtilis. *J. Biol. Chem.*, **263,** 11743-11749.

McCarthy, M.I., Kruglyak, L. and Lander, E.S. (1998). Sib-pair collection strategies for complex diseases. *Genet. Epidemiol.*, **15,** 317-340.

McNatty, K.P., Hillier, S.G., van den Boogaard, A.M., Trimbos-Kemper, T.C., Reichert, L.E. Jr. and van Hall, E.V. (1983). Follicular development during the luteal phase of the menstrual cycle. *J. Clin. Endocrinol. Metab.*, **56,** 1022-1031.

Means, G.D., Kilgore, M.W., Mahendroo, M.S., Mendelson, C.R. and Simpson, E.R. (1991). Tissue-specific promoters regulate aromatase cytochrome P450 gene expression in human ovary and fetal tissues. *Mol. Endocrinol.*, **5,** 2005-2013.

Milland, J., Christiansen,D., Thorley, B.R., McKenzie, I.F. and Loveland, B.E. (1996). Translation is enhanced after silent nucletide substitutions in A+T-rich sequences of the coding region of CD46 cDNA. *Eur. J. Biochem.*, **238,** 221-230.

Miller, J.A. and Jacobs H.S. (1986). Treatment of hirsutism and acne with cyproterone acetate. *Clin. Endocrinol. Metab.*, **15,** 373-389.

Monté, D., DeWitte, F. and Hum, D.W. (1998). Regulation of the human P450scc gene by Steroidogenic Factor 1 is mediated by CBP/p300. *J. Biol. Chem.*, **273,** 4585 - 4591.

Moore, C.C.D., Brentano, S. T. and Miller, W. L. (1990). Human P450scc gene transcription is induced by cyclic AMP and repressed by 12-O-tetradecanoylphorbol-13-acetate and A23187 through independent cis elements. *Mol. Cell. Biol.*, **10,** 6013-6023.

Moore, C.C.D., Hum, D.W. and Miller, W.L. (1992). Identification of positive and negative placenta-specific basal elements and a cyclic adenosine 3',5'-Monophosphate responce element in human gene for P450scc. *Mol. Endocrinol.*, **6,** 2045-2058.

Mooser, V., Francesco, P.M., Bopp, S., Petho-Schramm, A., Guerra, R., Boerwinkle, E., Muller, Hans-Joachin and Hobbs, H.H. (1995). Sequence polymorphisms in the Apo(a) gene associated with specific levels of Lp(a) in plasma. *Hum. Mol. Genet.*, **4,** 173-181.

Morohashi, K., Sogawa, K., Omura, T. and Fujii-kuriyama Y. (1987). Gene structure of human cytochrome P-450 (SCC), cholesterol desmolase. *J. Biochem.*, **101,** 879-887.

Morohashi, K., Honda, S., Inomata, Y., Handa, H. and Omura, T. (1992). A common *trans*-acting factor, Ad4-binding protein, to the promoters of steroidogenic P-450s. *J. Biol. Chem.*, **267,** 17913-17919.

Morton, N.E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, **7**, 277-318.

Nahum, R., Thong, K.J. and Hillier, S.G. (1995). Metabolic regulation of androgen production by human thecal cells *in vitro*. *Hum. Reprod.*, **10**, 75-81.

Nelson, V.L., Legro, R.S., Strauss III, J.F. and McAllister, J.M. (1999). Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. *Mol. Endocrinol.*, **13**, 946-957.

Nestler, J., and Jakubowicz, D. (1996). Decreases in ovarian cytochrome P450c17α activity and serum free testosterone after reduction of insulin secretion in polycystic ovary syndrome. *N. Engl. J. Med.*, **335**, 617-620.

Norman, R.J., Masters, S. and Hague, W. (1996). Hyperinsulinemia in common in family members of women with polycystic ovary syndrome. *Fert. Steril.*, **66**, 942-947.

Nothen, M.M., Propping, P. and Fimmers, R. (1993). Association versus linkage studies in psychosis genetics. *J. Med. Genet.*, **30**, 634-637.

O'Driscoll, J.B., Mamtora,H., Higginson, J., Pollock, A., Kane, J. and Anderson, D.C. (1994). A prospective study of the prevalence of clear cut endrocrine disorders and polycystic ovaries in 350 patients with hirsutism or androgenic alopecia. *Clin. Endocrinol. (Oxf)*, **41**, 231-236.

O'Meara, N., Blackman, J.D., Ehrmann, D.A., Barnes, R.B, Jaspan, J.B., Rosenfield, R.L. and Polonsky, K.S. (1993). Defects in B-cell function in functional ovarian hyperandrogenism. *J. Clin. Endocrinol. Metab.*, **76**, 1241-1247.

Owen, M.J. and McGuffin, P. (1993). Association and linkage complementary strategies for complex disorders. *J. Med. Genet.*, **30**.

Pache, T.D., Chadha, S., Goorens, L.J.G., Hop, W.C.J., Jaarsma, K.W., Dommerholt, H.B.R. and Fauser, B.C.J.M. (1991). Ovarian morphology in long-term androgen-treated female to male transsexuals. A human model for the study of polycystic ovarian syndrome? *histopatholoy*, **19,** 445-452.

Parisi, L., Tramonti, M., Derchi, L.E., Casciano, S., Zurli, A. and Rocchi, P. (1984). Polycystic ovarian disease:ultrasonic evaluation and correlations with clinical and hormonal data. *J. Clin. Ultrasound*, **12,** 21-26.

Peiris, A.N., Aiman, E.J., Drucker, W.D. and Kissebah, A.H. (1989). The relative contributions of hepatic and peripheral tissues to insulin resistance in hyperandrogenic women. *J. Clin. Endocrinol. Metab.*, **68,** 715-720.

Peserico, A., Angeloni, G., Bertoli, P., Marini, A., Piva, G., Panciera, A. and Suma, V. (1989). Prevalence of polycystic ovaries in women with acne. *Arch. Dermatol. Res.*, **281,** 502-503.

Pierpoint, T., McKeigue, P.M., Isaacs, A.J., Wild, S.H. and Jacobs, H.S. (1998). Mortality of women with polycystic ovary syndrome at long-term follow-up. *J. Clin. Epidemiol.*, **51,** 581-586.

Polson, D.W., Franks, S., Reed M.J., Cheng, R.W., Adams, J. and James, V.H. (1987a). The distribution of oestradiol in plasma in relation to uterine cross-sectional area in women with polycystic or multifollicular ovaries. *Clin. Endocrinol. (Oxf)*, **26,** 581-588.

Polson, D.W., Mason, H.D., Saldahna, M.B. and Franks, S. (1987b). Ovulation of a single dominant follicle during treatment with low-dose pulsatile follicle stimulating hormone in women with polycystic ovary syndrome. *Clin. Endocrinol. (Oxf)*, **26,** 205-212.

Polson, D.W., Adams, J., Wadsworth, J. and Franks, S. (1988). Polycystic ovaries-a common finding in normal women. *Lancet*, **1,** 870-872.

Polymeropoulos, M.H., Xiao, H., Rath, D. S. and Merril, C. R. (1991). Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). *Nucleic Acids Res.*, **19,** 195.

Poretsky, L. (1994). Insulin resistance and hyperandrogenism: Update 1994. *Endocrine Reviews*, **2,** 125-129.

Pugeat, M., Cousin, P., Nicolas, M.H., Mirakian, P., Sayegh, M., Barret, Ch., Ducros, Cl., Dechaud, H., Forest, M.G. and Lejeune, H. (1999). Pentanucleotide repeat polymorphism in the promoter region of the human CYP11a gene and polycystic ovaries in hirsute women. In *81st Annual Meeting of The Endocrine Society*: San Diego, California.

Pugeat, M., Nicolas, M.H., Cousin, P., Mirakian, P., Brebant, C., Dechaud, H., Lejeune, H. and Forest, M.G. (June 1996). Polymorphism of the 5' promoter of the human CYP17 gene encoding P450c17α and adrenal secretion in hirsute women. In *Proceedings of the 10th International Congress of Endocrinology, San Francisco, June 1996.*

Rader, D.J., Cain, W., Ikewaki, K., Talley, G., Zech, L.A., Usher, D. and Brewer, H.B. (1994). The inverse association of plasma lipoprotein(a) concentrations with apolipoprotein(a) isoform size is not due to differences in Lp(a) catabolism but to differences in production rate. *J. Clin. Invest.*, **93,** 2758-2763.

Rainey, W.E., Sawetawan, C., McCarthy, J.L., McGee, E.A., Bird, I.M., Word, R.A. and Carr, B.R. (1996). Human ovarian tumor cells: a potential model for thecal cell steroidogenesis. *J Clin. Endocrinol. and Metab.*, **81,** 257-263.

Rebar, R., Judd, H.L., Yen, S.S., Rakoff, J., Vandenberg, G. and Naftolin, F. (1976). Characterization of the inappropriate gonadotropin secretion in polycystic ovary syndrome. *J. Clin. Invest.*, **57,** 1320-1329.

Rice, P.A., Yang, S., Mizuuchi,K. and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87,** 1295-1306.

Richards, J.S., Jahnsen, T., Hedin, L., Lifka, J., Ratoosh, S.L., Durcia, J.M. and Goldring, N.B. (1987). Ovarian follicular development: from physiology to molecular biology. *Recent Prog. Horm. Res.*, **43,** 231-270.

Richards, J.S. and Midgley, A.R. Jr. (1976). Protein hormone action:a key to understanding ovarian follicular and luteal cell development. *Biol. Reprod.*, **14,** 82-
Richards J.S. (1994). Hormonal control of gene expression in the ovary. *Endocrine Rev.*, **15,** 725-751.

Riley, B.P. (1996). Molecular genetics of familial schizophrenia in southern African Bantu-speaking pedigrees. In *Department of Biochemistry and Molecular Genetics.* Imperial College: London.

Risch, N. (1990a). Linkage strategies for genetically complex traits: 1. Multilocus models. *American Journal of Human Genetics*, **46,** 222-228.

Risch, N. (1990b). Linkage strategies for genetically complex traits: 2. the power of affected relative pairs. *Am. J. Hum. Genet.*, **46,** 229-241.

Risch, N., Giuffra, L. (1992). Model misspecification and multipoint linkage analysis. *J. Hum. Hered.*, **42,** 77-92.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, **273,** 1516-1517.

Robinson, S., Chan, S.P., Spacey, S., Anyaoku, V., Johnston, D.G., Franks, S. (1992). Post-prandial thermogenesis is reduced in polycystic ovary syndrome and is associated with increased insulin resistance. *Clin. Endocrinol. (Oxf)*, **36,** 537-543.

Robinson, S., Kiddy, D., Gelding, S.V., Willis, D., Nithythyananthan, R., Bush, A., Johnston, D.G. and Franks, S. (1993). The relationship of insulin sensitivity to

menstrual pattern in women with hyperandrogenism and polycystic ovaries. *Clin. Endocrinol.*, **39,** 351-355.

Robinson, S., Henderson, A.D., Gelding, S.V., Kiddy, D., Nithythyananthan, R., Bush, A., Richmond, W., Johnston, D.G. and Franks, S. (1996). Dyslipidaemia is associated with insulin resistance in women with polycystic ovaries. *Clin. Endocrinol.*, **44,** 277-284.

Rodin, A., Thakkar, H., Taylor, N. and Clayton, R. (1994). Hyperandrogenism in polycystic ovary syndrome. Evidence of dysregulation of 11 β-hydroxysteroid dehydrogenase. *N. Engl. J. Med.*, **330,** 460-465.

Rodriguez, H., Hum, D.W., Staels, B. and Miller, W.L. (1997). Transcription of the human genes for cytochrome P450scc and P450c17 is regulated differently in human adrenal NCI-H295 cells than in mouse adrenal Y1 cells. *J. Clin. Endocrinol. Metab.*, **82,** 365-371.

Rosenfield, R.L., Barnes, R.B., Cara, J.F. and Lucky, A.W. (1990). Dysregulation of cytochrome P450c 17 alpha as the cause of polycystic ovarian syndrome. *Fertil. Steril.*, **53,** 785-791.

Rosenthal, I.M., Refetoff, S., Rich, B., Barnes, R.B., Sunthornthepvarakul, T., Parma, J. and Rosenfield, R.L. (1996). Response to challenge with gonadotropin-releasing hormone agonist in a mother and her two sons with a constitutively activating mutation of the luteinizing hormone receptor - a clinical research center study. *J. Clin. Endocrinol. Metab.*, **81,** 3802-3806.

Rousseau-Merck, M.F., Misrahi, M., Atger, M., Loosfelt, H., Milgrom, E. and Berger, R. (1990). Localization of the human luteinizing hormone/choriogonadotropin receptor gene (LHCGR) to chromosome 2p21. *Cytogenet. Cell Genet.*, **54,** 77-79.

Russell, D.W. and Wilson, J.D. (1994). Steroid 5 alpha-reductase: two genes/two enzymes. *Annu. Rev. Biochem.*, **63,** 25-61.

Schaffer, A.A., Gupta, S.K., Shriram, K. and Cottingham, R.W. (1994). Recomputation in linkage analysis. *Hum. Hered.*, **44,** 225-237.

Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science*, **253,** 1001-1007.

Schulze, E., Scheffer, S., Luer, S., Maser-Gluth, C., Gerhard, I., Rabe, T. and Schwabe, U. (1999). Genetic analysis of women with hyperandrogenism. In *81st Annual Meeting of the Endocrine Society.*: San Diego, California.

Sharp, P.S., Kiddy, D.S., Reed, M.J., Anyaoku, V., Johnston, D.G. and Franks, S. (1991). Correlation of plasma insulin and insulin-like growth factor-I with indicies of androgen transport and metabolism in women with polycystic ovary syndrome. *Clin. Endocrinol. Metab.*, **35,** 253-257.

Shenker, A., Laue, L., Kosugi, S., Merendino, J.J.,Jr., Minegishi, T. and Cutler, G.B.,Jr. (1993). A constitutively activating mutation of the luteinizing hormone receptor in familial male precocious puberty. *Nature*, **365,** 652-654.

Shibata, H., Kanzaki, M., Takeuchi, T., Miyazaki, J. and Kojima, I. (1996). Two distinct signalling pathways activated by activin A in glucose-responsive pancreatic beta-cell lines. *J. Mol. Endocrinol.*, **16,** 249-258.

Shima, K., Kitayama, S. and Nakano, R. (1987). Gonadotropin binding sites in human ovarian follicle and corpora lutea during the menstrual cycle. *Obstet. Gynecol.*, **69,** 800-806.

Shozu, M., Akasofu, K., Harada, T. and Kubota, Y. (1991). A new cause of female pseudohermaphroditism: placental aromatase deficiency. *J. Clin. Endocrinol. Metab.*, **72,** 560-566.

Simpson, J.L. (1992). *Elucidating the genetics of polycystic ovary syndrome.* Polycystic Ovary Syndrome. Blackwell Scientific Publications: Oxford.

Singh, A., Hamilton-Fairley, D., Koitinen, R., Seppala, M., James, V.H.T., Franks, S. and Reed, M.J. (1990). Effect of insulin-like growth factor type I (IGF-I) and insulin on the secretion of sex hormone binding globulin and IGF-I binding protein (IBP-I) by human hepatoma cells. *J. Endocrinol.*, **124,** R1-R3.

Smith, C.A.B. (1959). Some comments on the statistical methods used in linkage investigations. *Am. J. Hum. Genet.*, **11,** 289-304.

Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, **52,** 506-516.

Spinder, T., Spinjkstra, J.J., van den Tweel, J.G., Burger, C.W., van Kessel, H., Hompes, P.G. and Gooren, L.J. (1989). The effects of long term testosterone administration on pulsatile luteinizing hormone secretion and on ovarian histology in eugonadal female to male transsexual subjects. *J. Clin. Endocrinol. Metab.*, **69.**

Stein, I.F. and Leventhal, M.L. (1935). Amenorrhea associated with bilateral polycystic ovaries. *Am. J. Obstet. Gynecol.*, **29,** 181-191.

Stewart, E.A., McKusick, K.B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., Lee, R., Maratukulam, A., O'Connor, K., Perkins, S., Piercy, M., Qin, F., Reif, T., Sanders, C., She, X., Sun, W.L., Tabar, P., Voyticky, S., Cow les, S., Fan, J.B., Mader, C., Quackenbush, J., Myers, R.M. and Cox, D.R. (1997). An STS-based radiation hybrid map of the human genome. *Genome Research*, **7,** 422-433.

Stocco, D.M. (1999). Steroidogenic acute regulatory (StAR) protein: what's new ? *Bioessays*, **21,** 768-775.

Stone, D. and Hechter, O. (1955). Studies on ACTH action in perfused bovine adrenals: aspects of progesterone as an intermediary in corticosteroidogenesis. *Arch Biochem Biophys*, **54,** 121.

Suarez, B.K. (1978). The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens*, **12,** 87-93.

Sugawara, T., Kiriakidou, M., McAllister, J.M., Holt, J.A., Arakane, F. and Strauss, J.F. Jr. (1997). Regulation of expression of the steroidogenic acute regulatory protein (StAR) gene: a central role for steroidogenic factor 1. *Steroids*, **62,** 5-9.

Swanson, M., Sauerbrei, E.E. and Cooperberg, P.L. (1881). Medical implicationsof ultrasonically detected polycystic ovaries. *J. Clin. Ultrasound*, **9,** 219-222.

Takayama, K., Morohashi, K., Honda, S., Hara, N. and Omura, T. (1994). Contribution of Ad4BP, a steroidogenic cell-specific transcription factor, to regulation of the human CYP11A and bovine CYP11B genes through their distal promoters. *J.Biochem.*, **116,** 193-203.

Thomson, G. (1995). Mapping disease genes: family-based association studies. *American Journal of Human Genetics*, **57,** 487-498.

Tijan, R. (1995). Molecular machines that control genes. *Scientific American*, **272,** 39-45.

Timpatanapong, P. and Rojanasakul, A. (1997). Hormonal profile and prevalence of polycystic ovary syndrome in women with acne. *J. Dermatol.*, **24,** 223-229.

Trommsdorff, M., Kochl, S., Lingenhel, A., Kronenberg, F., Delport, R., Vermaak, H., Lemming, L., Klausen, I.C., Faergeman, O., Utermann, G., et al. (1995). A pentanucleotide repeat polymorphism in the 5' control region of the apolipoprotein(a) gene is associated with lipoprotein(a) plasma concentrations in Caucasians. *J. Clin. Invest.*, **96,** 150-157.

Urbanek, M., Legro, R.S., Driscoll, D.A., Azziz, R. Ehrmann, D.A., Norman, R.J., Strauss III, J.F., Spielman, R.S. and Dunaif, A. (1999). Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. *Proc. Natl. Acad. Sci.*, **96,** 8573-8578.

Utermann, G. (1989). The mysteries of lipoprotein(a). *Science*, **246,** 904-910.

Vafiadis, P., Bennett, S.T., Colle, E., Grabs, R., Goodyer, C.G. and Polychronakos, C. (1996). Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes. *J. Autoimmunity*, **9,** 397-403.

Velazquez, E.M., Mendoza, S., Hamer, T., Sosa, F. and Glueck, C.J. (1994). Metformin therapy in polycystic ovary syndrome reduces hyperinsulinaemia, insulin resistance, hyperandrogenaemia, and systolic blood pressure, while facillitating normal menses and pregnancy. *Metabolism*, **43,** 647-654.

Vendola, K.A., Zhou,J., Adesanya, O.O., Weil, S.J. and Bondy, C.A. (1998). Androgens stimulate early stages of follicular growth in the primate ovary. *J. Clin. Invest.*, **101,** 2622-2629.

Wade, D.P., Clarke, J.G., Lindahl, G.E., Liu, A.C., Zysow, B.R., Meer, K., Shwartz, K.S. and Lawn, R.M. (1993). 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc. Natl. Acad. Sci. USA*, **90,** 1369-1373.

Wajchenberg, B.L., Achando, S.S., Okada, H., Czeresnia C.E., Piexoto, S., Lima, S.S. and Goldman, J. (1986). Determination of the source(s) of androgen overproduction in hirsutism associated with polycystic ovary syndrome by simultaneous adrenal and ovarian venous catheterization. Comparison with the dexamethasone suppression test. *J. Clin. Endocrinol. Metab.*, **63,** 1204-1210.

Waldstreicher, J., Santoro, N.F., Hall, J.E., Filicori, M. and Crowley, W.F. Jr. (1988). Hyperfunction of the hypothalamic-pituitary axis in women with polycystic ovarian

disease: indirect evidence for partial gonadotroph desensitization. *J. Clin. Endocrinol. Metab.*, **66,** 165-172.

Wang, C.F. and Gemzell, C. (1980). The use of human gonadotropins for induction of ovulation in women with polycystic ovarian disease. *Fertil. Steril.*, **33,** 479-486.

Waterworth, D.M., Bennett, S.T., Gharani, N., McCarthy, M., Hague, S., Batty, S., Conway, G.S., White, D., Todd, J.A., Franks, S. and Williamson, R. (1997a). Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. *Lancet*, **349,** 986-990.

Waterworth, D.M. (1997b). Study of the molecular genetics of polycystic ovary syndrome. In *Department of Biochemistry and Molecular Genetics.* Imperial College: London.

Weaver, J.U., Kopelman, P.G. and Hitman, G.A. (1992). Central obesity and hyperinsulinaemia in women are associated with polymorphism in the 5' flanking region of the human insulin gene. *Eur. J. Clin. Invest.*, **22,** 265-270.

Weeks, D.E. and Lange, K. (1988). The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.*, **42,** 315-326.

White, D.M., Polson, D.W., Kiddy, D., Sagle, P., Watson, H., Gilling-Smith, C., Hamilton-Fairley, D. and Franks, S. (1996). Induction of ovulation with low-dose gonadotropins in polycystic ovary syndrome: an analysis of 109 pregnancies in 225 women. *J. Clin. Endocrinol. Metab.*, **81,** 3821-3824.

Whitelaw, P.F., Smyth, C.D., Howles, C.M. and Hillier, S.G. (1992). Cell-specific expression of aromatase and LH receptor mRNA in rat ovary. *J. Mol. Endocrinol.*, **9,** 309-313.

Wild, R.A. and Bartholomew, M.J. (1988). The influence of body weight on lipoprotein lipids in patients with polycystic ovary syndrome. *Am. J. Obstet. Gynecol.*, **159,** 423-427.

Willis, D., Mason, H., Gilling-Smith, C. and Franks, S. (1996). Modulation by insulin of follicle stimulating hormone and luteinizing hormone action in human granulosa cells of normal and polycystic ovaries. *J. Clin. Endocrinol. Metab.*, **81,** 302-309.

Wilson, J.D., Griffin, J.E. and Russell, D.W. (1993). Steroid 5 alpha-reductase 2 deficiency. *Endocrine Rev.*, **14,** 577-593.

Wu, H-W. and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308,** 509-513.

Yano, K., Kohn, L.D., Saji, M., Okuno, A., and Cutler, G.B.Jr. (1997). Phe576 plays an important role in the secondary structure and intracellular signaling of the human luteinizing hormone/chorionic gonadotropin receptor *J. Clin. Endocrinol. Metab.*, **82,** 2585-2591.

Yen, S.S.C., Vela, P. and Rankin, J. (1970). Inappropriate secretion of follicle-stimulating hormone and luteinizing hormone in polycystic ovarian disease. *J. Clin. Endocrinol. Metab.*, **30,** 435-442.

Yen, S.S.C. (1980). The polycystic ovary syndrome. *Clin. Endocrinol.*, **12,** 177-207.

Zawadzki J.K. and Dunaif A. (1992). Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. *In: Dunaif A, Givens JR, Haseltine FP, Merriam GR, eds. Polycystic ovary syndrome. Oxford, England: Blackwell Scientific,*, 377-84.

Zinkel, S.S. and Crothers, D.M. (1991). Catabolite activator protein-induced DNA bending in transcrition initiation. *J. Mol. Biol.*, **219,** 201-215.

BIBL. LONDIN. UNIV.

0021-972X/96/$03 00 0
Journal of Clinical Endocrinology and Metabolism
Copyright © 1996 by The Endocrine Society
Vol. 81, No. 11
Printed in U.S.A.

LETTERS TO THE EDITOR

5′ Polymorphism of the CYP17 Gene is Not Associated with Serum Testosterone Levels in Women with Polycystic Ovaries*

To the editor:

Polycystic ovary syndrome (PCOS) is a highly prevalent endocrine disorder characterized by hyperandrogenemia and is the most common cause of anovulatory infertility and hirsutism (1–3). Despite the variable clinical presentation a consistent finding is the presence of elevated serum androgens (4, 5), suggesting an underlying disorder of androgen biosynthesis in PCOS, for which there appears to be a genetic basis (6, 7).

Previously several studies had suggested that there may be an abnormal regulation of the enzyme P450c17α in women presenting with PCOS. This enzyme is thought to catalyze the rate-limiting step in androgen biosynthesis in the ovaries and the adrenals and was therefore regarded as a good candidate for involvement in the etiology of PCOS (8). In a previous study, we carried out linkage analysis on the gene CYP17, coding for P450c17α, on chromosome 10q24.3, in twenty PCOS/male pattern baldness (MPB) pedigrees (9). Clear exclusion of this locus as a primary genetic defect in PCOS/MPB was obtained. However sequencing analysis of the CYP17 promoter region identified a base pair change that confers an additional SP1-type promoter element. This also created an *Msp*A1 restriction enzyme cleavage site that allowed a simple screening assay to be set up. The frequency of this variant (A2) allele was investigated in the pedigrees and in a small, case-control data-set (n = 68) comprising consecutively identified caucasians with polycystic ovaries and normal controls. Preliminary results showed an association of the A2 allele with PCOS (odds ratio = 3.57, P = 0.03). We proposed that this extra promoter element may up-regulate the expression of CYP17, resulting in an increased synthesis of androgens.

We have now extended our case-control data-set (n = 96) and have obtained data on total serum testosterone levels. To confirm our hypothesis we have used this data-set to examine the association of the A2 allele with serum testosterone levels. New samples were typed for the CYP17 polymorphism as described previously (9). All individuals (affected and control) were then allocated to three groups according to their genotypes, A1/A1, A1/A2, and A2/A2. Analysis of variance (ANOVA) was used to compare mean serum testosterone levels in the different groups (Fig. 1). No significant difference was found between the means of the groups (P = 0.82).

Consequently, we have used this extended data-set to reevaluate the association of the A2 allele with PCOS. A χsquare contingency table was used to test for any significant difference between the groups. The results are summarized in Table 1. We could no longer find an association between the A2 allele and PCOS (P = 0.30). Previously we had found the frequency of individuals with this allele to be 74% and 38% for PCOS affected and normal control respectively, however our results in this larger series of subjects now show frequencies of 67% and 62% respectively.

The results of this follow-up study have shown that there is no significant association between CYP17 and PCOS, suggesting that this gene does not play a major role in the etiology of hyperandrogenemia. These findings are in keeping with data reported in another recent study of hyperandrogenemic women (10). The discrepancy between the previous and present results is probably due simply to the relatively small number of normal controls in the original data-set. We may therefore conclude that this base-pair change identified in the promoter region of the CYP17 gene is a common polymorphism with no obvious role in the etiology of PCOS or hyperandrogenemia.

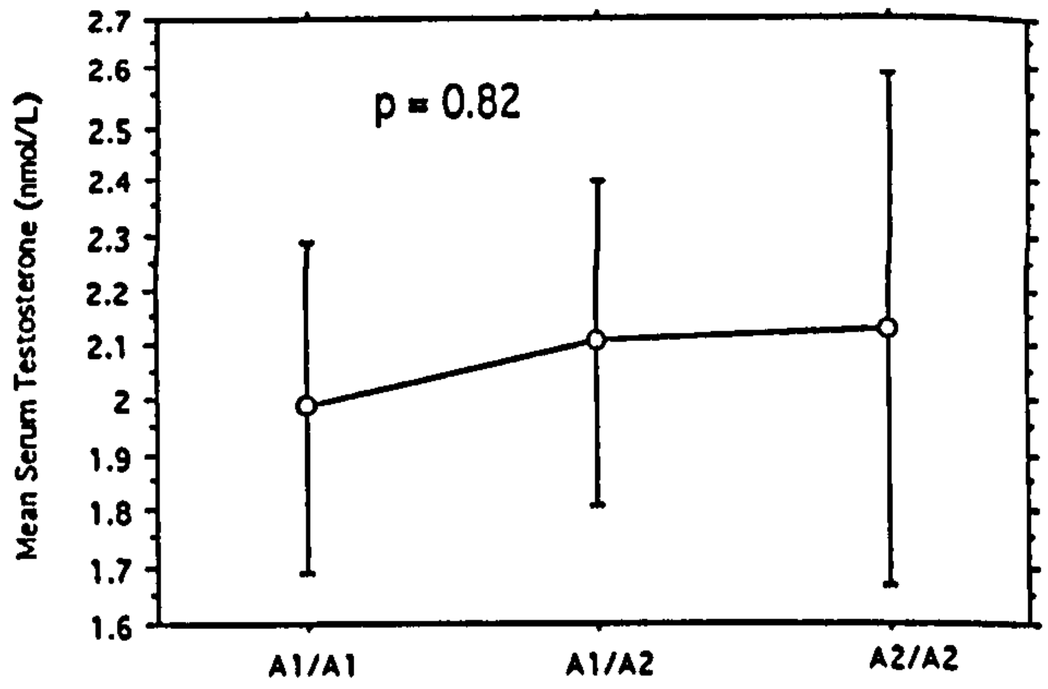


FIG. 1. Mean and 95% confidence intervals for serum testosterone levels in the three CYP17 polymorphism genotype groups. There are 26, 36, and 15 individuals in the A1/A1, A1/A2, and A2/A2 groups respectively. The mean serum testosterone levels for the different groups are A1/A1 (1.99 nmol/L), A1/A2 (2.10 nmol/L), and A2/A2 (2.13 nmol/L).

TABLE 1. Results of screening for the presence of the A2 allele

	A2/A2	A1/A2	A1/A1	% with A2 present
PCOS affected	11	22	16	67%
Normal control	5	24	18	62%

There was no significant difference between PCOS and controls in the allelic distribution.

Neda Gharani, Dawn M. Waterworth, Robert Williamson, Stephen Franks
Imperial College School of Medicine at St. Mary's
London W2 1PG, UK
Murdoch Institute, Royal Children's Hospital
Melbourne, Australia (R.W.)

References

1. Adams J, Polson DW, Franks, S. 1986 Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. Br Med J. 293:355–359.
2. Hull MG. 1987 Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. Gynecol Endocrinol. 1:235–245.
3. Fox R, Corrigan E, Thomas PA, Hull MGR. 1991 The diagnosis of polycystic ovaries in women with oligo-amenorrhoea: predictive power of endocrine tests. Clin Endocrinol (Oxf). 34:127–131.
4. Franks S. 1989 Polycystic ovary syndrome: a changing perspective. Clin Endocrinol (Oxf). 31:87–120.
5. Franks S. 1991 The ubiquitous polycystic ovary. J Endocrinol. 129:317–319.
6. Simpson JL. 1992 Elucidating the genetics of polycystic ovary syndrome. In: Dunaif A, Givens JR, Haseltine FP, Merriam GR, eds. Polycystic ovary syndrome. Oxford: Blackwell Scientific Publications; pp 59–77.
7. Carey AH, Chan KL, Short F, White D, Williamson R, Franks S. 1993 Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. Clin Endocrinol. 38:653–658.
8. Rosenfield RL, Barnes RB, Cara JF, Lucky AW. 1990 Dysregulation of cytochrome P450c 17α as the cause of polycystic ovarian syndrome. Fertil Steril. 53:785–791.
9. Carey AH, Waterworth DM, Patel K, et al. 1994 Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet. 3:1873–1876.
10. Pugeat M, Nicolas MH, Cousin P, et al. Polymorphism of the 5′ promoter of the human CYP17 gene encoding P450c17α and adrenal secretion in hirsute women. Proceedings of The 10th International Congress of Endocrinology, San Francisco, June 1996. Abstract P-626 (published by The Endocrine Society, 1996).

* Received July 25, 1996. Address correspondence to: Neda Gharani, Department of Molecular Genetics, Imperial College School of Medicine at St. Mary's, Norfolk Place, London, England W2 1PG.

© 1997 Oxford University Press

Human Molecular Genetics, 1997, Vol. 6, No. 3 397–402

Association of the steroid synthesis gene *CYP11a* with polycystic ovary syndrome and hyperandrogenism

Neda Gharani[1,*], Dawn M. Waterworth[1], Sari Batty[2], Davinia White[2], Carole Gilling-Smith[2], Gerard S. Conway[4], Mark McCarthy[1,3], Stephen Franks[2] and Robert Williamson[1,5]

[1]Department of Molecular Genetics, [2]Department of Obstetrics and Gynaecology and [3]Unit of Metabolic Medicine, Imperial College School of Medicine at St Mary's, London W2 1PG, UK, [4]Department of Endocrinology, The Middlesex Hospital, Mortimer Street, London W1N 8AA, UK and [5]Murdoch Institute, Royal Children's Hospital, Melbourne, Australia

Received September 16, 1996; Revised and Accepted January 6, 1997

Biochemical data implicate an underlying disorder of androgen biosynthesis and/or metabolism in the aetiology of polycystic ovary syndrome (PCOS). We have examined the segregation of the genes coding for two key enzymes in the synthesis and metabolism of androgens, cholesterol side chain cleavage (*CYP11a*) and aromatase (*CYP19*), with PCOS in 20 multiply-affected families. All analyses excluded *CYP19* co-segregation with PCOS, demonstrating that this locus is not a major determinant of risk for the syndrome. However, our results provide evidence for linkage to the *CYP11a* locus (NPL score = 3.03, $p = 0.003$). Parametric analysis using a dominant model suggests genetic heterogeneity, generating a maximum HLOD score of 2.7 ($\alpha = 0.63$). An association study of 97 consecutively identified Europids with PCOS and matched controls demonstrates significant allelic association of a *CYP11a* 5′ UTR pentanucleotide repeat polymorphism with hirsute PCOS subjects ($p = 0.03$). A strong association was also found between alleles of this polymorphism and total serum testosterone levels in both affected and unaffected individuals ($p = 0.002$). Our data demonstrate that variation in *CYP11a* may play an important role in the aetiology of hyperandrogenaemia which is a common characteristic of polycystic ovary syndrome.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a highly prevalent endocrine disorder which is characterised by hyperandrogenaemia and represents the most common cause of anovulatory infertility and hirsutism (1–3). PCOS has been estimated to have a population prevalence of between 5–10%. The characteristic polycystic ovarian morphology, however, may be found in up to 22% of the normal population, with >90% of these women having at least one mild symptom that may be considered a clinical marker of PCOS (4).

Although it is clear that PCOS is a familial disorder, with a risk to siblings of ~50%, the precise mode of inheritance has not been agreed (5–11). A review of the various family studies carried out (12) has highlighted problems such as a lack of consensus regarding clinical phenotypes for PCOS, as well as a lack of agreement on the male phenotype. A dominant mode of inheritance has been suggested and there is evidence that premature male pattern baldness (MPB) may be the male phenotype (6,9,11). Rather than representing a simple, single gene disorder, it is more likely that PCOS has a complex genetic basis, where the interaction of multiple genetic and environmental factors determine the development of the syndrome. This would explain the observation that sisters with polycystic ovaries may present with different clinical symptoms.

Hyperandrogenaemia is seen both in women with PCOS and men with premature male pattern baldness suggesting an underlying disorder of androgen biosynthesis or metabolism (13–15). Androgens are synthesised by the adrenals, the theca cell layer of the developing ovarian follicle and the testicular Leydig cells. Both scalp hair loss and hirsutism are known to be mediated by androgens (16–18). The sensitivity of the hair follicle to androgens is dependent on a number of factors, such as serum concentrations of bioavailable androgens and the presence and number of androgen receptors (19,20).

It has been shown that theca cells from polycystic ovaries show a significant increase in both androstenedione and progesterone production *in vitro* when compared to normal theca (21). This suggests that the putative defect in steroidogenesis occurs at or above the level of progesterone biosynthesis. The enzyme cholesterol side chain cleavage (cytochrome P450scc) catalyzes the conversion of cholesterol to pregnenolone, the first, rate-limiting step at the start of the steroid hormone biosynthesis pathway (22,23). We hypothesise that upregulation of this enzyme could lead to an increase in androgen production.

A second steroidogenic enzyme, aromatase (cytochrome P450arom), catalyzes the conversion of the C_{19} steroids (androgens) to the C_{18} oestrogens. Evidence for altered regulation

*To whom correspondence should be addressed

of this enzyme in PCOS comes from the observation that granulosa cells from anovulatory polycystic ovaries are hyperresponsive to FSH *in vitro*, displaying significantly greater oestradiol production than granulosa cells from normal ovaries (24). Conversely, two other studies have demonstrated cases in which aromatase deficiency was associated with hyperandrogenaemia (25,26). These studies suggest that an altered regulation of this enzyme may be involved in PCOS.

The aim of the present study was to investigate P450scc and P450arom, as potential candidates for involvement in PCOS/MPB. The genes encoding P450 scc (*CYP11a*) and P450 arom (*CYP19*) are both located on the long arm of chromosome 15 at positions 15q24 and 15q21.1 respectively. Given uncertainties regarding the segregation model for PCOS, both parametric and non-parametric linkage analyses were conducted in 20 PCOS/MPB pedigrees. Mutation screening of the promoter region of *CYP11a* was also carried out. Finally, using a *CYP11a* intragenic marker, we carried out an association study in a case-control data set of consecutively identified Europid women with PCOS and matched controls.

RESULTS

Both parametric and non-parametric linkage analyses were carried out for each of the candidate loci using the GENEHUNTER program (27). The position and chromosomal order of the microsatellite markers used in the analyses are shown in Figure 1. The non-parametric results for the *CYP11a* locus have provided evidence for excess allele sharing, generating a maximum NPL score of 3.03 ($p = 0.003$) at the *CYP11a* locus (Table 1). Under the parametric analysis (using a dominant model with 95% penetrance and 10% phenocopy rate), the overall multipoint LOD score obtained at the *CYP11a* $(ac)_n$/D15S169 locus was 1.74: a homogeneity test on these data suggested that ~60% of the pedigrees may be linked to this locus [maximum HLOD score of 2.7, $\alpha = 0.63$, computed at the *CYP11a* $(ac)_n$/D15S169 locus]. It has been suggested that a *p*-value of <0.001 (for a non-parametric analysis) and a LOD score of >1.9 (for a parametric analysis) may be taken as suggestive evidence for linkage for a dense, complete genome scan (28). These criteria take into account the multiple testing aspect of a genome scan.

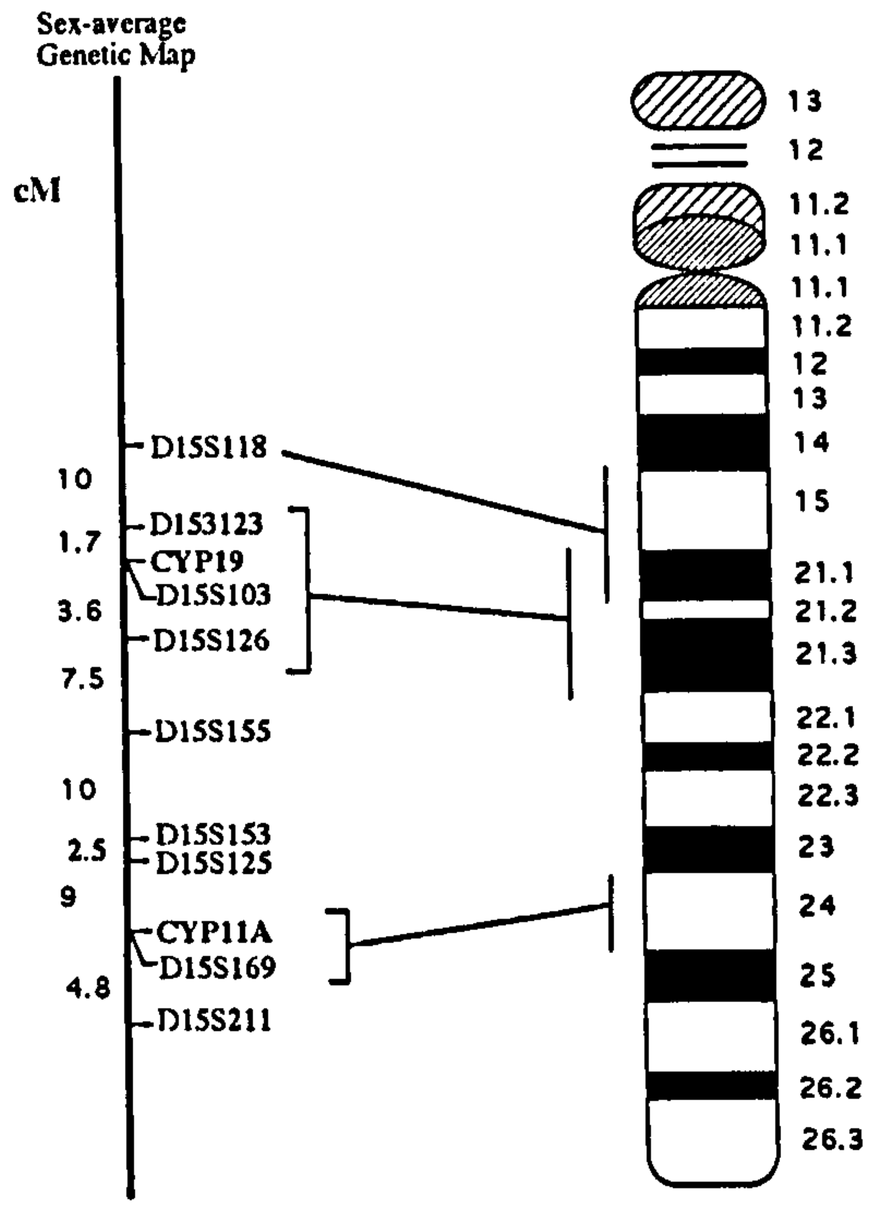


Figure 1. Physical and genetic map of chromosome 15. Chromosomal order and interval distances were obtained from the chromosome 15 map at: http://cedar.genetics.soton.ac.uk/pub/chr15. All D15S markers are Genethon microsatellite markers (37), except markers D15S169 and D15S103 (36). For each of the above two loci an intragenic microsatellite has also been used in the linkage analysis, *CYP19* $(ttta)_n$ (38) and *CYP11a* $(ac)_n$ (36).

Since we studied only a small number of relevant candidate genes, these criteria are likely to be overly stringent for our study.

The results for the *CYP19* locus are summarised in Table 2. Clear exclusion of the *CYP19* locus was obtained; there was no evidence for excess allele sharing across the whole region and the maximum negative LOD score generated at the *CYP19* locus is –7.45.

Table 1. Multipoint parametric and non-parametric analysis in the *CYP11a* region

Marker	(cM)	LOD score	(alpha, HLOD)	NPL score	*p*-value	Information
D15S155	0.00	–3.52	(0.23, 0.35)	0.625	0.250	0.820
D15S153	10.00	–2.98	(0.25, 0.44)	1.093	0.130	0.934
D15S125	12.50	–1.60	(0.28, 0.48)	1.243	0.102	0.943
CYP11a* $(ac)_n$/*D15S169	**21.50**	**1.74**	**(0.63, 2.67)**	**3.034**	**0.003**	0.958
D15S211	26.30	–0.82	(0.41, 0.96)	2.093	0.021	0.915

Interlocus scores of the multipoint data are available on request. Maximum parametric and non-parametric scores are shown in bold. There is zero genetic distance between the *D15S169* and *CYP11a* $(ac)_n$ markers. Non-parametric linkage score (NPL) is a Z score. The position of the markers along the chromosome are given in centimorgans (cM) moving from the most centromeric marker to the most telomeric. HLOD is the maximum LOD score allowing for heterogeneity, where alpha is the estimated proportion of families linked to this locus. An estimation of the information content at each point is also given (one being the maximum).

Table 2. Multipoint parametric and non-parametric analysis in the *CYP19* region

Marker	(cM)	LOD score	NPL score	*p*-value	Information
D15S118	0.00	–5.04	0.189	0.400	0.849
D15S123	10.00	–7.18	–0.421	0.664	0.954
CYP19(ttta)$_n$/*D15S103*	11.70	–7.45	–0.279	0.604	0.975
D15S126	15.30	–5.18	0.386	0.321	0.941

Interlocus scores of the multipoint data are available on request. Maximum negative parametric LOD score is shown in bold. Non-parametric analysis provides no significant evidence for excess allele sharing across the whole region. There is zero genetic distance between the *D15S103* and *CYP19*(ttta)$_n$ markers. For abbreviations refer to legend of Table 1.

Table 3. Comparison of genotype distributions between different sub-groups in the case-control data set

	Genotypes		*p*-value	
	216+	216–	(total group)	(excluding normal controls)
Analysis by quartiles of serum testosterone:				
Top quartile	19	15	0.015	
50th-75th percentile	24	7		
25th-50th percentile	33	7		
Bottom quartile	27	4		
Analysis by clinical diagnosis:				
PCOS	72	25	0.03	
aPCO/Normal controls	95	15		
Analysis by hirsutism:				
Hirsute	25	14	0.01	0.03
Non-hirsute	51	11		
Analysis by ovulatory status:				
Anovulatory	66	22	0.08	0.23
Ovulatory	101	18		

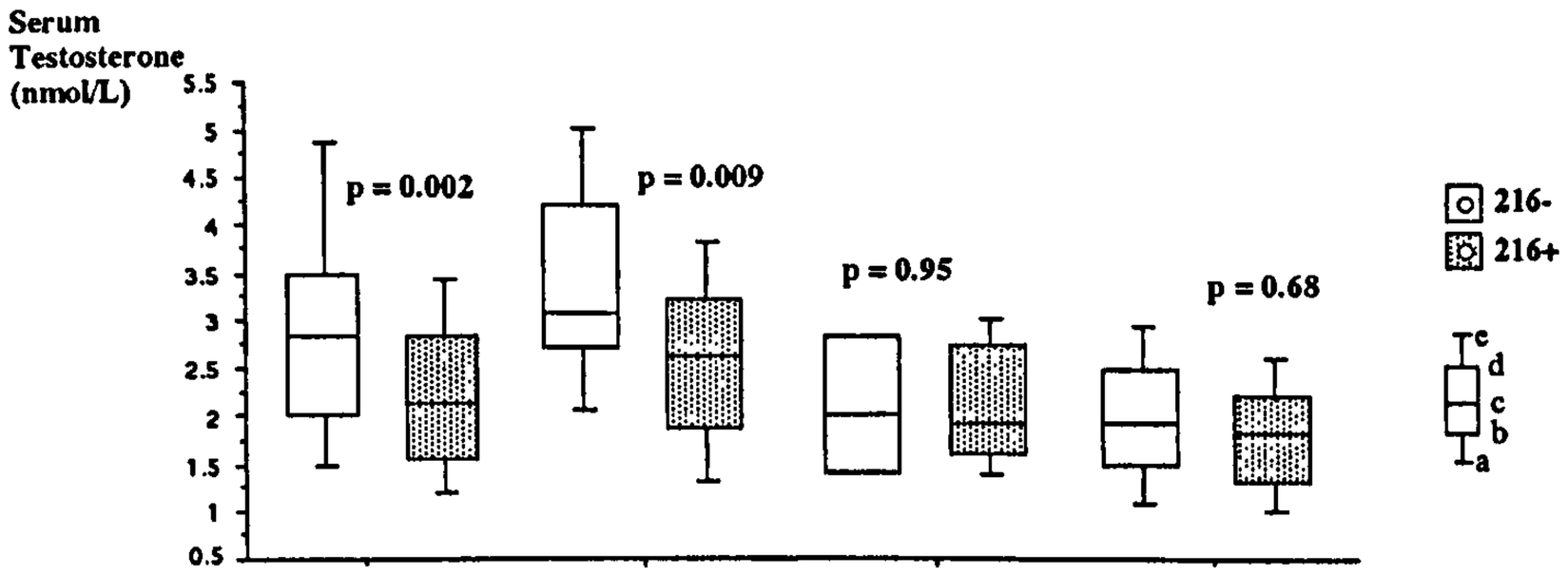
The 'Top quartile' includes women with serum testosterone levels of 2.90 nmol/l and greater, the '50th-75th percentile' includes levels between 2.2 to 2.9 nmol/l, the '25th-50th percentile' includes levels between 1.6 to 2.2 nmol/l and the 'Bottom quartile' includes those with levels of 1.60 nmol/l and lower, for the entire data-set (affecteds and controls). '216+' genotype includes individuals with at least one 216 allele and '216–' includes those with no 216 allele. χ^2 contingency tables were used to compare genotype distributions between the different groups. Genotype comparisons for the 'Analysis by hirsutism' and 'Analysis by ovulatory status' were carried out for the total group (genotype distributions given above), as well as after the exclusion of normal controls (*p*-values for both comparisons have been reported). For the 'Analysis by hirsutism', only individuals with a clear assignment as either hirsute or non-hirsute were included.

Any pathogenic mutations in *CYP11a* are expected to lead to upregulation; therefore we focused our mutation screening on the promoter region of *CYP11a*. This region is believed to contain multiple cAMP-regulated elements, responsible for increasing basal transcriptional activity, and elements involved in repression of expression (29). PCR amplification, followed by direct sequencing analysis, was used to screen 1.85 kb of sequence in ten affected and eight unaffected members of the pedigrees. No alterations to the published sequence were found in any individual apart from variation in the number of repeat units of the two microsatellite polymorphisms present in this region [a dinucleotide $(ac)_n$ repeat, and a pentanucleotide $(tttta)_n$ repeat, at positions –1376 and –528 respectively from the ATG start of translation site].

Individuals in the case-control study were typed for the *CYP11a* $(tttta)_n$ marker. Four alleles were observed in our data set, alleles 216, 226, 236 and 241 with observed frequencies of 0.59, 0.28, 0.04 and 0.09 respectively. In order to reduce multiple allele comparisons for infrequent alleles, individuals were allocated to two groups according to their genotype, those having at least one copy of the common, four-repeat units allele (designated 216+) and those with no 216 allele (designated 216–). An association between this polymorphism and serum testosterone levels was found in the total data set (Mann-Whitney, $p = 0.002$) (Fig. 2). We also compared the genotype distributions of each quartile of serum testosterone level within the total data set. The proportion of individuals with 216– genotypes in each quartile were (from highest to lowest) 0.44, 0.23, 0.18 and 0.13 ($p = 0.015$; Table 3).

On subgroup analysis, the association between *CYP11a* genotype and testosterone was only evident in the women with PCOS (Mann-Whitney, $p = 0.009$; Fig. 2), there being no significant difference between the mean testosterone levels of 216– and 216+ individuals for either the normal controls or the aPCO groups. However, the numbers in these groups were small, raising the possibility of a type 2 error.

We next compared the *CYP11a* genotype distributions between subject groups (Table 3). Since the association between serum testosterone levels and *CYP11a* genotype was only seen in the PCOS group, taking the entire data-set (including those without testosterone levels available), we compared PCOS women with a combined normal control/aPCO group. There was a significant



33	107	22	47	3	25	7	35	(n)
2.80	2.10	3.05	2.60	2.00	1.90	1.80	1.90	median testosterone (nmol/L)
2.90	2.22	3.36	2.55	2.10	2.20	1.94	1.77	mean testosterone (nmol/L)
1.19	0.95	1.13	0.93	0.95	1.06	0.68	0.68	± SD

Figure 2. Association of *CYP11a* with total serum testosterone levels in the case control data-set. The Mann-Whitney test was used to compare total serum testosterone levels between the two *CYP11a* genotype groups. *P*-values obtained for each test are given above. The 'Total group' included all the women in the data set. The 'PCOS' group are the women with polycystic ovaries and symptoms of anovulation and/or hirsutism. The 'aPCO' group are the asymptomatic women with polycystic ovaries and the 'normal control' group are the control subjects with normal ovarian morphology. '216+' sub-group are individuals with at least one 216 allele and '216–' are those with no 216 allele. 'a', 'b', 'c', 'd' and 'e' are the 10th, 25th, 50th, 75th and 90th percentiles of serum testosterone levels respectively. The median, mean and standard deviation (±SD) of serum testosterone levels for each group are also given above.

difference in the distribution of genotypes between these groups ($p = 0.03$).

When women were classified according to hirsutism (hirsute vs non-hirsute) we also found a marked difference in genotype distributions ($p = 0.01$ for the total group and $p = 0.03$ for women with polycystic ovaries). No such difference was seen when subjects were reclassified according to ovulatory status ($p = 0.08$ for the total group and $p = 0.23$ for women with polycystic ovaries).

DISCUSSION

Our data show that the gene coding for aromatase (*CYP19*) can be excluded as a major determinant of risk for PCOS/MPB. In contrast, the linkage results for *CYP11a* (encoding side chain cleavage) provide support for the involvement of this gene in the aetiology of PCOS/MPB. The association data demonstrate that allelic variants of *CYP11a* mediate the development of hyperandrogenaemia, which is in turn associated with PCOS and hirsutism. The association of this gene with hirsutism and the lack of association with ovulatory status indicates that *CYP11a* predominantly has a role in the development of hirsutism in PCOS.

We have shown a significant association between alleles of the *CYP11a* $(tttta)_n$ marker and serum testosterone levels. It is feasible that allelic variants of this rate-limiting enzyme could influence the level of androgen production. This association however, can only be demonstrated in the PCOS group. The failure to demonstrate association in the aPCO and normal groups may reflect the small sample sizes. Alternatively it may be that this locus has a modifying effect, interacting with other genetic and/or environmental factors controlling the development of the ovarian changes and thereby influencing the clinical presentation of PCOS.

CYP11a $(tttta)_n$ is present in the promoter region of *CYP11a* and sequencing analysis of the surrounding regions does not reveal any other mutations likely to have direct pathogenic effects. However only 1.85 kb of sequence was screened and the putative mutation could be located outside this region. Recently a number of gene-associated VNTRs (variable number tandem repeats) have been implicated in the regulation of the genes, such as the insulin gene VNTR (30). Also, direct association between alleles of a $(tttta)_n$ repeat polymorphism, present in the 5′ flanking region of the apolipoprotein(a) gene, and plasma levels of Lp(a) has been demonstrated (31). Expression studies have shown a 5-fold higher transcriptional activity in constructs containing eight copies of the apolipoprotein(a) (tttta) repeat compared to those with nine copies (32).

As yet no regulatory function has been assigned to the *CYP11a* $(tttta)_n$ polymorphism and further investigation of the promoter region of *CYP11a* is required to determine the relationship between the (tttta) repeat polymorphism and the regulation of this gene.

We propose that allelic variants of *CYP11a* have a role in the aetiology of hyperandrogenaemia. This in turn may be sufficient to cause the altered ovarian morphology or may act in conjunction with other genetic factors to modify the phenotypic expression of PCOS.

Table 4. Summary of information on the PCR fragments S1, S2 and S3

PCR fragment	Primer-set sequence	Size of product (bp)	Distance from ATG codon
S1	cttctgagggaggaatgtgg ggttgcttagagctaggggg	700 bp	−1174 to −1874
S2	tcctctaattcctctatccccc gagagtttcattcttgttgccc	660 bp	−551 to −1211
S3	ggtgaaactgtgccattgc ctgtacctgctccacttcagc	581 bp	−21 to −602

MATERIALS AND METHODS

Subjects

Twenty pedigrees (145 members) were selected from the Reproductive Medicine and Endocrinology clinics at the Samaritan and St. Mary's Hospital as described previously (11). Probands presented with either menstrual disturbances and/or hirsutism and all had bilateral polycystic ovaries on ultrasound scan. Other causes of anovulation and hirsutism such as Cushing's syndrome and late-onset congenital adrenal hyperplasia were excluded. All family members underwent a full screening procedure. Assignment of affection status in all women of reproductive age was based solely on the presence of bilateral polycystic ovaries on ultrasound scan. The only exception to this strict diagnostic scheme was the assignment of affection status in post menopausal women in whom ovarian imaging is not informative given reduced ovarian activity. In such women we assigned individuals as affected on the basis of a clear history of previous menstrual dysfunction and/or hirsutism: by this criteria, six (6) post menopausal women were assigned as affected. Pre-menarchal women and post-menopausal women, with a negative history, were assigned as unknown. Men were assigned as affected if they demonstrated significant premature male pattern baldness (MPB), defined as greater than a revised Hamilton IIa score (33), before the age of 40. If MPB was found to be present, age of onset was estimated using both historical photographic evidence and subjective assessment. Our previous definition of onset before the age of 30 years (11) has not been used because during the collection of further families we have observed that within certain families the age of onset of hair loss is between 30 and 40 years. This observation is generally conserved within a single pedigree. Thus to avoid the loss of power in those pedigrees we have used the Ferriman and Purdie definition of MPB (6). Sixteen of the pedigrees were Europid, two were Iranian and two Asian. Pedigree sizes range from 5 to 19 members. There are 59 affected women and 25 affected men. Over 80% of the affected women in the pedigrees are symptomatic.

Women in the case-control study were all Europid and were collected consecutively from two centres: the Reproductive Medicine and Endocrinology clinics at the Samaritan and St. Mary's Hospital, and from the Endocrinology clinics at the Middlesex Hospital, London. This data-set consisted of 97 probands presenting with anovulation and/or hirsutism and polycystic ovaries on ultrasound scan, designated 'PCOS'. In addition 51 non-hirsute women having regular menstrual cycles and polycystic ovaries (referred to as 'asymptomatic PCO' or 'aPCO') were collected together with 59 non-hirsute control subjects with regular menstrual cycles and normal ovarian morphology on ultrasound scan ('normal controls'). Serum testosterone levels of 69 PCOS, 28 aPCO and 42 normal subjects were measured by radioimmunoassay. Ovarian morphology was defined by pelvic ultrasonography (1). Women were assigned as hirsute if they obtained a score of at least 8 using the Ferriman–Gallwey index (34). Women with oligomenorrhea (intermenstrual interval of >6 weeks) and amenorrhea (intermenstrual interval >6 months) were designated anovulatory.

Sequencing analysis

Three overlapping PCR fragments were generated using primers designed from the published sequence of the 5′ region of *CYP11a* (35), spanning 1.85 kb of sequence upstream of the ATG start of translation site. The three fragments were designated S1, S2 and S3 (Table 4).

All polymerase chain reactions (PCR) of genomic DNA were carried out in a Hybaid OmniGene thermocycler. For each PCR reaction the 100 µl amplification mixture contained 50 ng of genomic DNA, 125 pmol of each primer, 1 × PCR reaction buffer [5 mM KCL, 1 mM Tris–HCl (pH 9), 0.01% Triton X-100 and 1 mM $MgCl_2$ (for the S2 fragment amplification) or 1.5 mM $MgCl_2$ (for the S1 and S3 fragments)], 100 µl each of dTTP/dCTP/dGTP/dATP and 1.5 U Amplitaq DNA polymerase (Cambio, UK). The reaction conditions were: 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 60°C (for S1 and S2), 63°C (for S3) and 1 min extension at 72°C. An initial denaturation step of 5 min at 94°C and a final extension of 10 min at 72°C were employed. In order to obtain single stranded DNA for sequencing in both the forward and reverse directions, two PCR reactions (one with only the forward strand biotinylated and the other with the reverse strand biotinylated) were carried out for each of the three fragments. Each PCR product was purified using the Promega Wizard PCR purification kit (Promega, UK). Biotinylated single stranded DNA was obtained using the Dynabeads M-280 Streptavidin system (Dynal Ltd, UK). Approximately 5 ng of each single stranded product was sequenced using the Sequanase version 2.0 DNA sequencing kit (USB Corporation, USA).

DNA samples from ten affected and eight unaffected members of the pedigrees, showing positive evidence for linkage, were screened.

Marker typing

All members of the pedigrees were typed for the markers *D15S118*, *D15S123*, *D15S103*,*CYP19* $(ttta)_n$, *D15S126*, *D15S155*, *D15S153*, *D15S125*, *CYP11a* $(ac)_n$, *D15S169* and *D15S211*, described elsewhere (36–38). Genotype analysis was carried out following PCR amplification.

All members of the case control study were typed for the *CYP11a* $(tttta)_n$ microsatellite marker, using the following primers designed from the published sequence: forward ggtgaaactgtgccattgc, reverse gtttgggggaaatgaggggc, using standard PCR conditions (39).

Data analysis

Multipoint parametric and non-parametric analyses were performed using the GENEHUNTER (27) programme. The non-parametric linkage score (NPL score) is a Z-score. For imperfect data (small number of pedigrees with different pedigree structures) the distribution of the Z is asymptotically normally distributed as the number of pedigrees is increased and as the information content approaches 100%. The significance level (p-value) assigned is conservative and becomes increasingly accurate as information content increases. This approach is referred to as the 'perfect data approximation'.

Model parameters for the parametric analysis were: disease frequency 20%, disease gene frequency 10%, penetrance 95%, phenocopy rate 10%, phenocopy risk 2.5%.

χ^2 contingency table analyses were used to compare the *CYP11a* $(tttta)_n$ genotype distributions in the case-control study. Mann-Whitney test was used to compare serum testosterone levels of the different groups in the case-control study.

ACKNOWLEDGEMENTS

We thank Dr Adam Carey for contribution to the family collection, Rosanne Joseph-Horne for assistance in collecting the female population and Julie Little and Patricia Novelli for help in marker typing. Work at ICSM at St. Mary's was funded by a ROPA award from the Medical Research Council and Unilever UK. Work at the Middlesex Hospital was funded by the British Heart Foundation, grant number PG/92138.

REFERENCES

1. Adams, J., Polson, D.W., Franks,S. (1986) Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *B. M. J.* **293**, 355–359.
2. Hull, M.G (1987) Epidemiology of infertilityand polycystic ovarian disease: endocrinological and demographic studies. *Gynecol. Endocrinol.* **1**, 235–245.
3. Fox, R., Corrigan, E., Thomas, P.A. and Hull, M.G.R. (1991) The diagnosis of polycystic ovaries in women with oligo-amenorrhoea: predictive power of endocrine tests. *Clin. Endocrinol. (Oxf)* **34**, 127–131.
4. Polson, D.W., Adams, J., Wadsworth, J. and Franks, S. (1988) Polycystic ovaries—a common finding in normal women. *Lancet* **1**, 870–872.
5. Cooper, H.E., Spellacy, W.N., Prem, K.A. and Cohen, W.D. (1968) Hereditary factors in the Stein-Leventhal syndrome. *Am. J. Obstet. Gynecol.* **100**, 371–387.
6. Ferriman, D.and Purdie, A.W. (1979) The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin. Endocrinol. (Oxf)* **11**, 291–300.
7. Givens, J.R. (1988) Familial polycystic ovarian disease. *Endocrinol. Metab. Clin. North Am.* **17**, 771–783.
8. Hague, W.M., Adams, J., Reeders, S.T., Peto, T.E.and Jacobs, H.S. (1988) Familial polycystic ovaries: a genetic disease? *Clin. Endocrinol. (Oxf)* **29**, 593–605.
9. Lunde, O., Magnus, P., Sandvik, L. and Hoglo, S. (1989) Familial clustering in the polycystic ovarian syndrome. *Gynecol. Obstet. Invest.* **28**, 23–30.
10. Simpson, J.L. (1992) Elucidating the genetics of polycystic ovary syndrome. In: Dunaif, A., Givens, J.R., Haseltine, F.P., Merriam, G.R., eds. *Polycystic Ovary Syndrome*. Oxford, England: Blackwell Scientific, 59–77.
11. Carey, A.H., Chan, K.L., Short, F., White,D., Williamson, R. and Franks, S. (1993) Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. *Clin. Endocrinol.* **38**, 653–658.
12. Legro, R.S. (1995) The genetics of polycystic ovary syndrome. *Am. J. Med.* **98**, 9S–16S.
13. Franks, S. (1989) Polycystic ovary syndrome: A changing perspective. *Clin. Endocrinol.* **31**, 87–120.
14. Franks, S. (1991) The ubiquitous polycystic ovary. *J. Endocrinol.* **129**, 317–319.
15. Stephens, J., Carey, A.H., Short, F., Williamson, R. and Franks, S. (1993) Premature balding, the male phenotype of polycystic ovaries (PCO). *J. Endocrinol.* **139**, 031.
16. Hamilton, J. B. (1942) Male hormone stimulation is a prerequisite and an incitant in common baldness. *Am. J. Anat.* **71**, 451–480.
17. Ebling, F.J.G. (1987) The biology of hair. *Dermatol. Clin.* **5**, 467–481.
18. Lobo, R.A. (1991) Hirsutism in polycystic ovary syndrome: current concepts. *Clin. Obstet. Gynecol.* **34**, 817–826.
19. Randall, V.A., Thorton, M.S., Hamada, K., Messenger, A.G (1992) Mechanisms of androgen action in cultured dermal papilla cells derived from human hair follicles with varying response to androgens in vivo. *J. Invest. Dermatol.* **98**, 865–915.
20. Barth, J.H. (1988) Alopecia and hirsuties: current concepts in pathogenesis and management. *Drugs* **35**, 83–91.
21. Gilling-Smith, C., Willis, D.S., Beard, R.W. and Franks, S. (1994) Hypersecretion of androstenedione by isolated thecal cells from polycystic ovaries. *J. Clin. Endocrinol. Metab.* **79**, 1158–1165.
22. Stone, D. and Hechter, O. (1955) Studies on ACTH action in perfused bovine adrenals: aspects of progesterone as an intermediary in corticosteroidogenesis. *Arch. Biochem. Biophys.* **54**, 121.
23. Halkerston, I.D.K., Eichhorn, J. and Hechter, O. (1961) A requirement for reduced triphosphopyridine nucleotide for cholesterol side-chain cleavage by mitochondrial fractions of bovine adrenal cortex. *J. Biol. Chem.* **236**, 374.
24. Mason, H.D., Willis, D.S., Beard, R.W., Winston, R.M.L., Margara, R. and Franks, S. (1994) Estradiol production by Granulosa cells of normal and polycystic ovaries: Relationship to menstrual cycle history and concentrations of gonadotropins and sex steroids in follicular fluid. *J. Clin. Endocrinol. Metab.* **79**, 1355–1360.
25. Harada, N., Ogawa, H., Shozu, M. and Yamada,K. (1992) Genetic studies to characterise the origin of the mutation in placental aromatase deficiency. *Am. J. Hum. Genet.* **51**, 666–667.
26. Ito, Y., Fisher, C.R., Conte, F.A., Grumbach, M.M. and Simpson, E.R. (1993) Molecular basis of aromatase deficiency in an adult female with sexual infantilism and polycystic ovaries. *Proc. Natl. Acad. Sci. USA* **90**, 11673–11677.
27. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996) Parametric and non-parametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363.
28. Lander, E.S. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247.
29. Moore, C.C.D., Brentano, S. T. and Miller, W. L. (1990) Human P450scc gene transcription is induced by cyclic AMP and repressed by 12-O-tetradecanoylphorbol-13-acetate and A23187 through independent cis elements. *Mol. Cell. Biol.* **10**, 6013–6023.
30. Kennedy GC German, M.S., Rutter, W.J. (1995) The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nature Genet.* **9**, 293–298.
31. Mooser, V., Francesco, P.M., Bopp, S., Petho-Schramm, A., Guerra, R., Boerwinkle, E., Muller, Hans-Joachin and Hobbs, H.H. (1995) Sequence polymorphisms in the Apo(a) gene associated with specific levels of Lp(a) in plasma. *Hum. Mol. Genet.* **4**, 173–181.
32. Wade, D.P., Clarke, J.G., Lindahl, G.E., Liu, A.C., Zysow, B.R., Meer, K., Shwartz, K.S. and Lawn, R.M. (1993) 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc. Natl. Acad. Sci. USA* **90**, 1369–1373.
33. Lesko, S.M., Rosenberg, L. and Sharpiro, S. (1993) A case control study of baldness in relation to myocardial infarction in men. *Proc. Natl. Acad. Scl. USA* **81**, 3443–3446.
34. Ferriman, D.and Gallwey, J. (1961) Clinical assessment of hair growth in women. *J. Clin. Metab.* **21**, 1440–1445.
35. Morohashi, K., Sogawa, K., Omura, T. and Fujii-kuriyama Y. (1987) Gene structure of human cytochrome P-450 (SCC), cholesterol desmolase. *J. Biochem.* **101**, 879–887.
36. Beckmann, J.S., Tomfohrde, J., Barnes, R.I., Williams, M., Broux, O., Richard, I., Weissenbach, J., Bowcock, A.M. (1993) A linkage map of human chromosome 15 with an average resolution of 2cM and containing 55 polymorphic microsatellites. *Hum. Mol. Genet.* **2**, 2019–2030.
37. Gyapay, G., Morrisette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., Weissenbach, J. (1994) The 1993–1994 Genethon human genetic linkage map. *Nature Genet.* **7**, 246–339.
38. Polymeropoulos, M.H., Xiao, H., Rath, D. S. and Merril, C. R. (1991) Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). *Nucleic Acids Res.* **19**, 195.
39. Weber, J.L.,and May, P. E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396.

Human Reproduction vol.12 no.12 pp.2641–2648, 1997

The genetic basis of polycystic ovary syndrome

Stephen Franks[1,4], Neda Gharani[2], Dawn Waterworth[2], Sari Batty[1], Davinia White[1], Robert Williamson[2] and Mark McCarthy[2,3]

Departments of [1]Obstetrics and Gynaecology, [2]Biochemistry and Molecular Genetics and [3]Unit of Metabolic Medicine, Imperial College School of Medicine, St Mary's Hospital, London W2 1PG, UK

[4]To whom correspondence should be addressed

Polycystic ovary syndrome (PCOS) is the most common endocrinopathy in women of reproductive age. Familial clustering of cases suggests that genetic factors play an important part in its aetiology. A number of studies of families with several cases of PCOS have produced results suggesting an autosomal dominant trait. Detailed analysis of a large number of affected families has, however, cast some doubt about the mode of inheritance. An autosomal dominant trait remains possible but a more complex aetiology seems more likely. The results of our recent studies support the concept of an oligogenic disorder in which genes affecting metabolic pathways in glucose homeostasis and steroid biosynthesis are both involved. We review evidence for an important role for the insulin gene minisatellite in the aetiology of anovulatory PCOS and for the gene coding for P450 cholesterol side chain cleavage (*CYP11a*) in the mechanism of excessive androgen secretion in women with polycystic ovaries. We propose that the heterogeneity of clinical and biochemical features in PCOS can be explained by the interaction of a small number of key genes with environmental, particularly nutritional, factors.

Key words: anovulation/*CYP11a*/folliculogenesis/insulin/polycystic ovary syndrome

Introduction

Polycystic ovary syndrome (PCOS) is a heterogeneous endocrine disorder which is considered to be the commonest cause of anovulatory infertility and hirsutism (Adams *et al.*, 1986; Hull, 1987). The most widely accepted definition of polycystic ovary syndrome is the association of anovulation (manifesting itself as irregular menses, oligomenorrhoea or amenorrhoea) with clinical or biochemical evidence of androgen excess (Zawadzki and Dunaif, 1992) but the identification of polycystic ovaries ultrasonographically has called into question the validity of this definition (Conway *et al.*, 1989; Franks 1989, 1995). It is now clear that the majority of hirsute women with regular menses have polycystic ovaries (Franks 1989; O'Driscoll *et al.*, 1994). Furthermore, the estimated prevalence of polycystic ovaries, as diagnosed by ultrasonography, in a normal (volunteer) population has been found to be over 20% (Polson *et al.*, 1988; Clayton *et al.*, 1992; Farquhar *et al.*, 1994) and even within this 'normal' group, many of these women will have symptoms which are considered to be typical of the syndrome.

Given the heterogeneous nature of its clinical and biochemical features, it has been suggested that PCOS represents a range of disorders rather than a single entity (Simpson, 1992). Although it seems likely that there is more than one cause of the syndrome, there are, nevertheless, certain biochemical features which are common to all groups of subjects with ultrasonographic evidence of polycystic ovaries irrespective of the clinical presentation. Serum levels of luteinizing hormone (LH) in hirsute women with polycystic ovaries and regular cycles, whilst lower than those in anovulatory subjects, are still significantly higher than normal (Adams *et al.*, 1986; Conway *et al.*, 1989; Franks, 1989). The most consistent endocrine feature in women with polycystic ovaries, however, appears to be hyperandrogenaemia, whether the mode of presentation is as the 'classic' syndrome or as an incidental finding on ultrasound examination (Franks, 1991). There has always been a vigorous debate about the source and aetiology of hyperandrogenaemia in PCOS. The weight of evidence suggests that the ovary is the major source of excess androgen (reviewed by Franks, 1995). Recent data from both clinical investigations and studies of isolated human theca cells implicate a primary ovarian abnormality rather than hypersecretion of androgens as a result of abnormal gonadotrophins (Gilling-Smith *et al.*, 1994, 1997; Ibañez *et al.*, 1996).

In addition to the well-described abnormalities of the pituitary ovarian axis, polycystic ovary syndrome is characterized by significant metabolic abnormalities. These include fasting and glucose-stimulated hyperinsulinaemia, peripheral insulin resistance (affecting predominantly muscle and adipose tissue), abnormalities of energy expenditure (reduced postprandial thermogenesis) and dyslipidaemia (reviewed by Dunaif, 1993; Franks, 1995; Holte, 1996). Furthermore, it has emerged that PCOS represents a major risk factor for non-insulin dependent diabetes mellitus (NIDDM). The prevalence of impaired glucose tolerance or frank diabetes in obese young women with PCOS lies (depending on the population studied) between 11% and 38% (Dunaif, 1993; Dunaif and Finegood, 1996; Holte, 1996). A long-term follow-up study of postmenopausal women with a previous history of PCOS found a 13% prevalence of NIDDM compared with <2% in the reference population – a seven-fold increase in risk (Dahlgren *et al.*, 1992a). Analysis of cardiovascular risk factors (such as hyperinsulinaemia and abnormal plasma lipids) suggests that

© European Society for Human Reproduction and Embryology

these patients are also at greater risk of developing cardiovascular disease in the future (Dahlgren *et al.*, 1992b). These findings emphasize that the significance of PCOS for women's health extends far beyond the implications for reproductive function, although these are important enough in themselves.

Polycystic ovary syndrome shows strong familial aggregation suggesting a major genetic component to its aetiology. In this paper, published and ongoing studies of the possible genetic basis of polycystic ovary syndrome, from this group, will be reviewed. This review will be set in the context of the clinical and biochemical background outlined above and in the light of previously published clinical and molecular genetic studies. We acknowledge that there is unlikely to be a single cause of the syndrome, but our hypothesis is that much of the clinical and biochemical variability within PCOS can be explained by the interaction of environmental (notably nutritional) factors with a small number of major causative genes which include those involved in androgen production and the secretion and/or action of insulin.

There are obvious problems which make genetic studies of polycystic ovary syndrome difficult to perform (Simpson, 1992; Legro, 1995). The heterogeneity and the lack of universally acceptable clinical or biochemical diagnostic criteria have been discussed. Another major handicap is that this is a disorder which primarily affects women of reproductive age and it is therefore very difficult for segregation studies to span more than one generation. In addition, as discussed below, there is no commonly accepted male phenotype. Lastly, the high prevalence of polycystic ovaries in the population means that large pedigrees, in particular, may include subjects with polycystic ovaries arising from a different genotype from that of the proband. Nevertheless, given modern methods of genetic modelling and molecular genotyoping these problems are not insurmountable, as we hope we will be able to illustrate in this review.

Family studies of polycystic ovary syndrome

A small number of clinical studies have been performed over the last 20 years which have drawn attention to the phenomenon of familial clustering of cases of polycystic ovary syndrome (Cooper *et al.*, 1968; Ferriman and Purdie, 1979; Givens *et al.*, 1988; Hague *et al.*, 1988; Lunde *et al.*, 1989; Carey *et al.*, 1993). Detailed analysis of these studies has been carried out in two excellent recent reviews (Simpson, 1992; Legro, 1995). Given that there is no unequivocal method of diagnosis, it is not surprising that the criteria used to identify probands and affected family members vary considerably between studies. A further confounding factor is that identification of affected family members was made by direct clinical observation in some studies, by questionnaire alone in some and by a combination of the two in others.

In one of the six largest studies (Hague *et al.*, 1988) no attempt was made to identify a male phenotype. In three others, premature balding was suggested as the likely manifestation of affected status in men but this was based, in two of the three, on evidence from questionnaires (Ferriman and Purdie, 1979; Lunde *et al.*, 1989) and, in the other, on a combination of data from direct observation, telephone interview and questionnaires (Carey *et al.*, 1993).

Similarly, there has been no general agreement about the mode of inheritance in PCOS. In four of six studies, segregation analysis gave results that were consistent with autosomal dominant inheritance (Cooper *et al.*, 1968; Ferriman and Purdie, 1979; Lunde *et al.*, 1989; Carey *et al.*, 1993) whilst one study suggested an X-linked mode (Givens *et al.*, 1988). In the other, the prevalence of polycystic ovaries among siblings was too high to be explained by a simple dominant model (Hague *et al.*, 1988). In the face of such inconsistencies, it is probably wise to make no assumptions about the mode of inheritance when performing linkage studies, as suggested below. One further difficulty relates to possible ethnic variations in the prevalence and presentation of the syndrome. None of the six studies has satisfactorily addressed this issue.

The St Mary's family studies

Our preliminary study at this centre, cited briefly in the previous section, focused on segregation analysis in 10 well characterized, multiply-affected families with polycystic ovaries (Carey *et al.*, 1993). It differed from those previously published in that it relied principally on direct interview and observation of relatives rather than on indirect evidence from questionnaires. Results from 50 women of reproductive age and 22 men were analysed. Assignment of affected status was made on the basis of ultrasound evidence of polycystic ovaries in the women and premature onset of fronto-parietal balding in men. In that study, to reduce the chance of false positive results, premature balding was defined as onset before the age of 30 although, conventionally, 40 years has been taken as the lower limit of normal (Ferriman and Purdie, 1979; Lunde *et al.*, 1989).

Although the diagnosis of polycystic ovaries was made ultrasonographically, 92% of affected female family members had at least one clinical (hirsutism, acne, menstrual disturbance) or biochemical feature (raised serum testosterone, LH) of polycystic ovary syndrome. The segregation ratio, expressed as the percentage of affected subjects in each generation (excluding the proband to avoid ascertainment bias), was calculated including data from the men and was found to be 51%, i.e. consistent with an autosomal dominant mode of inheritance. These initial results raised the clear prospect of a single gene effect and set us on a search for an appropriate candidate gene.

Subsequently, we have expanded some of the existing pedigrees and added new ones so that the number of families now includes 23 informative pedigrees. We have reviewed the data and have found that the picture is somewhat more complex than it appeared initially. This is well illustrated by the pedigree shown in Figure 1. In this family, the proband (14) was hirsute, anovulatory and had an elevated serum testosterone level. Both sisters (11, 16) had polycystic ovaries, were non-hirsute but had acne and raised serum testosterone; her brother (13) was prematurely bald. Her mother (5) was postmenopausal but was presumed to have been affected on the basis of a history of hirsutism, irregular menses and because her serum testosterone

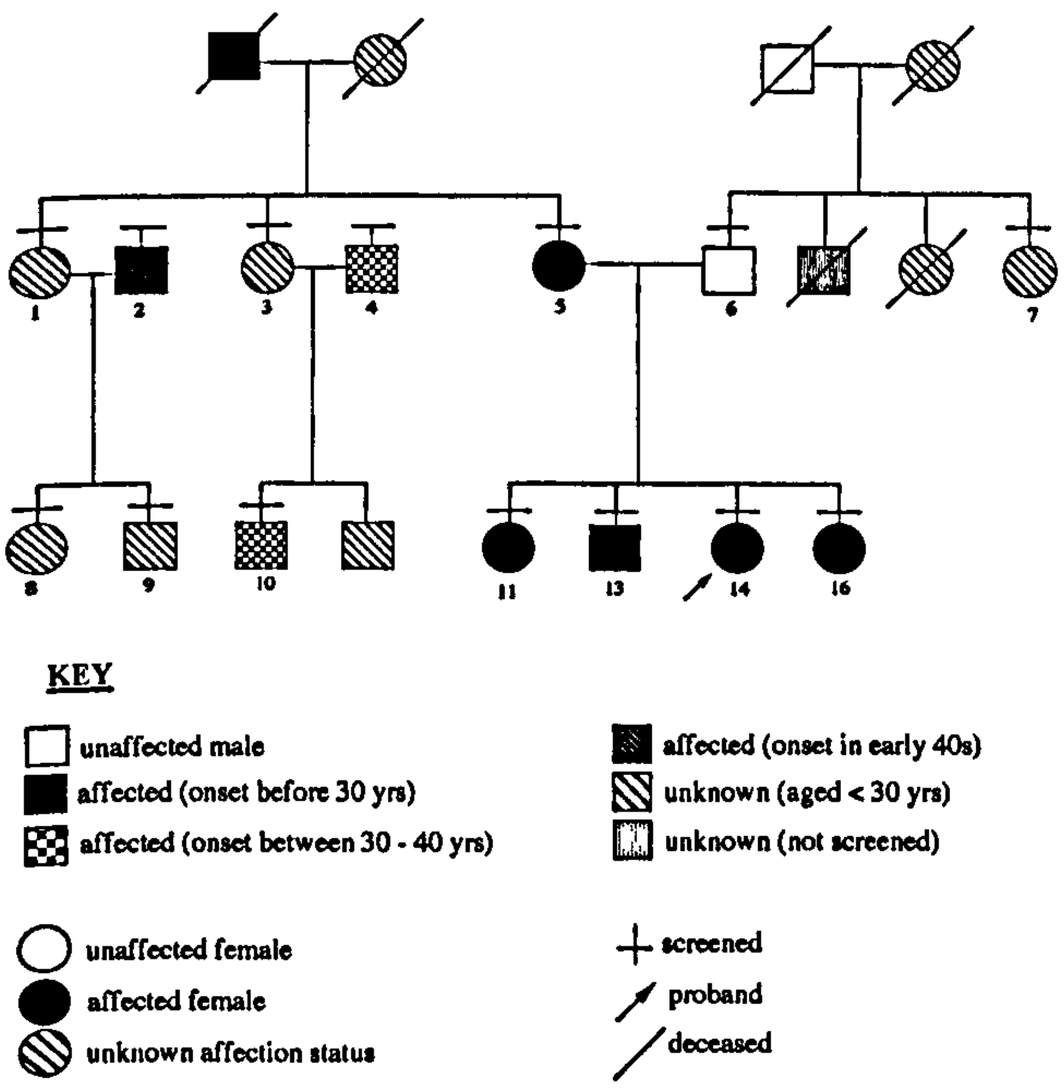


Figure 1. A large pedigree with familial polycystic ovary syndrome. Numbers below and horizontal bars above symbols indicate individuals who were fully screened by interview, examination, ovarian ultrasound (females) and biochemical testing. Affected status in other family members was assigned by history and/or photographic evidence.

was 4.7 nmol/l (normal range for premenopausal women 0.5–2.7 nmol/l). The two maternal aunts (1, 3) were also postmenopausal; neither had a history of hirsutism or menstrual disturbance but one had a serum testosterone of 5 nmol/l. One was married to a man (2) who became bald in his 40s and the other to a man (4) who developed significant hair loss between the ages of 30 and 40. Of their offspring (8, 9, 10), subject 8 was pregnant at the time of the study and the son of subjects 1 and 2 (9) was under 30. The son of parents 3 and 4 (10), like his father (4) became bald between 30 and 40 years of age. The proband's father (6) had no significant hair loss and of the two paternal aunts, one (7) was postmenopausal, but with no history suggestive of PCOS, and one was deceased.

This pedigree exemplifies the following points: (i) symptomatic heterogeneity between the proband and her sisters, all three of whom, nevertheless, had polycystic ovaries and hyperandrogenaemia; (ii) the problems associated with assigning definite affected status to more than one generation because of questions about the reliability of data from postmenopausal women; and (iii) difficulties in assigning affected status to men – especially those who were under 30 or who had noticed onset of balding around the age of 40 years.

Although the results are not incompatible with an autosomal dominant model it would be unwise to consider this mode of inheritance to the exclusion of all others. In this context, the suggestion by Simpson (1992) that PCOS should be treated as a quantitative trait disorder has considerable merit. This does not necessarily imply a truly polygenic aetiology because it would be possible to explain the variable phenotype on the basis of a small number of causative genes (a so-called oligogenic basis for disease). A candidate gene approach therefore remains valid but, rather than perform linkage studies using a single-gene autosomal dominant model, we have recently used a linkage analysis programme which makes no assumption about the mode of inheritance. In the following section the results of association and linkage studies applied to examination of the role of possible candidate genes will be discussed. Given the biochemical phenotype characteristic of women with polycystic ovaries we focused on genes coding for steroidogenic enzymes in the androgen biosynthetic pathway and those involved in the secretion and action of insulin.

Genes coding for steroidogenic enzymes

The 17-hydroxylase/17,20-lyase gene (CYP17)

On the basis of clinical studies which pointed to abnormal regulation of 17-hydroxylase/17,20-lyase (a known rate-limiting step in androgen biosynthesis) (Barnes *et al.*, 1989; Rosenfield *et al.*, 1990), our initial investigations focused on the possible role of *CYP17* (the gene encoding P450c17α). A 459bp fragment in the 5′ untranslated region of *CYP17* was amplified by polymerase chain reaction (PCR). A single base change (a T to C substitution at -34 base pairs from the starting point of translation) was found (Carey *et al.*, 1994). Conveniently, this variant allele includes a restriction site for the enzyme *Msp*-1, thus allowing a simple method of screening DNA by restriction fragment length polymorphism (RFLP) analysis.

Linkage studies were performed in PCOS families using polymorphic markers close to the gene and, on the basis of these, it was possible to exclude *CYP17* as a major causative gene. Nevertheless, using RFLP screening of the -34 allele, preliminary case-control data suggested an association between the variant allele of *CYP17* and PCOS (Carey *et al.*, 1994). These findings were, however, based on a relatively small population of subjects (71 patients and 33 controls) and subsequently we and others have been unable to confirm these results (Gharani *et al.*, 1996; Pugeat *et al.*, 1996; Techatraisak *et al.*, 1997; Franks, 1997). Critically, in none of these studies was any relationship found between the *CYP17* variant and serum androgen levels.

Cholesterol side chain cleavage gene, CYP11a

Our studies of ovarian theca cells in culture have demonstrated that PCO theca cells produce an excess of both androgens and progesterone (Gilling-Smith *et al.*, 1994; Franks *et al.*, 1996a). This prompted us to examine *CYP11a* [encoding P450 side chain cleavage (P450scc)] as a possible candidate gene for abnormal steroidogenesis (Gharani *et al.*, 1997). We therefore examined the segregation of *CYP11a* in 20 families and performed association studies in consecutively recruited, premenopausal, European women with polycystic ovaries on ultrasound and matched control women (with normal ovaries) from a similar ethnic background. We included 97 women with symptomatic PCOS, 51 subjects with polycystic ovaries and no symptoms, and 59 with normal ovaries.

Using an informative, microsatellite marker in the promoter

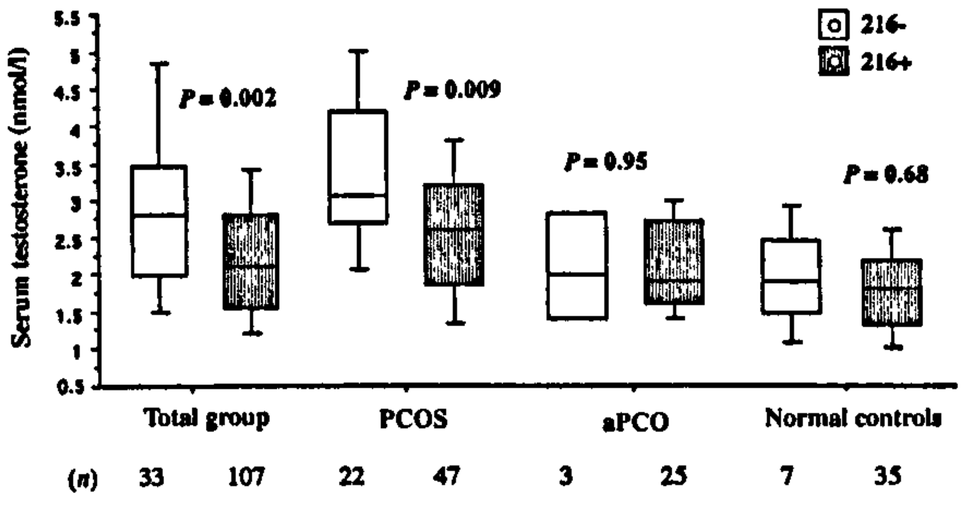


Figure 2. Relationship between serum testosterone and alleles of *CYP11a* in a case-control data-set. Serum testosterone concentrations in women with polycystic or normal ovaries are grouped according to genotype at the promoter region of *CYP11a*. 216+ individuals are those with at least one 216 alelle, 216- are subjects with no 216 allele. 'PCOS' are polycystic ovary syndrome subjects with anovulation and/or hirsutism; aPCO are those with polycystic ovaries but no symptoms. Values shown are 10th, 25th, 50th, 75th and 90th centiles for serum testosterone. Groups were compared by the Mann–Whitney test (from Gharani *et al.*, 1997, with permission).

region of *CYP11a*, genotype analysis was performed after PCR amplification. In the case-control study, subjects were allocated to one of two groups according to the presence or absence of the most common polymorphism, a pentanucleotide repeat $(tttta)_n$, -528bp from the AGT start of translation site. Individuals were designated as 216+ (at least one copy) or 216- (no 216 allele). Our results showed that variation at the *CYP11a* gene was associated with both PCOS and serum testosterone concentrations (Gharani *et al.*, 1997) (Figure 2). On further analysis, it was clear that differences in serum testosterone between 216+ and 216− subjects were maintained in the major subgroup of women with symptomatic PCO (i.e. with polycystic ovary syndrome). In a further analysis, the distribution in genotype was found to vary significantly if subjects were classified according to testosterone levels or by the presence of hirsutism.

Using a number of polymorphic markers in the region of *CYP11a*, we carried out non-parametric linkage analysis using the GENEHUNTER (multipoint linkage) programme (Kruglyak *et al.*, 1996). We found evidence for excess allele sharing (i.e. linkage) at the *CYP11a* locus, generating a maximum non-parametric linkage (NPL) score of 3.03 (P = 0.003). The data from both association and linkage studies suggest that *CYP11a* is a major genetic susceptibility locus for PCOS.

The aromatase gene

In the same population, the possible role of the gene encoding P450 aromatase (*CYP19*) was examined. There have been reports of hyperandrogenism occurring in rare patients with aromatase deficiency (Harada *et al.*, 1992; Ito *et al.*, 1993). In immunohistochemical studies of polycystic ovaries, Takayama *et al.* (1996) were unable to detect aromatase in antral follicles of various sizes. On the other hand, Mason *et al.* (1994) demonstrated enhanced oestradiol production by granulosa cells of antral follicles from polycystic ovaries, suggesting that, functionally, there was no evidence of an intrinsic deficiency of aromatase. Nevertheless, all these studies pointed to abnormal regulation of aromatase in women with hyperandrogenism. We therefore performed both a case control study and linkage analysis. The results revealed no association of alleles of *CYP19* with PCO and no evidence for excess allele sharing (Gharani *et al.*, 1997).

Genes involved in secretion and action of insulin

Numerous metabolic studies have revealed abnormalities of both insulin secretion and action in women with PCOS (reviewed by Dunaif, 1993; Holte, 1996). These studies have shown that there is an interaction between body weight and PCOS, so that individuals with PCOS are more insulin resistant than control subjects, even allowing for the effects of obesity. The results of such studies raise the possibility that genes implicated in the secretion and action of insulin may have a role in the aetiology of PCOS.

The insulin receptor gene

The demonstration of impaired sensitivity to insulin action *in vivo* and *in vitro* naturally led to the hypothesis that genetic abnormalities of the insulin receptor and/or post-receptor signalling were involved in the pathogenesis of familial PCOS. There have been sporadic reports of a PCOS-like phenotype occurring in patients with severe insulin resistance associated with defects of the insulin receptor gene (Moller and Flier, 1988) but Conway *et al.* (1994) were unable to detect any abnormalities of the tyrosine kinase domain of the insulin receptor gene in a population of 22 hyperinsulinaemic women with PCOS. These results are supported by those in a recently published paper by Talbot *et al.* (1996). In this study, molecular scanning of the entire coding region of the insulin receptor gene was carried out on DNA samples from 24 well-characterized women with PCOS. Common polymorphisms were detected, especially in the intron 5′ to exon 3, but no missense or nonsense mutations (i.e. those that would be expected to result in marked impairment of receptor function) were found. The authors concluded that mutations of the insulin receptor gene were rare in women with PCOS.

As far as post-receptor signalling is concerned, it remains to be determined whether there is a genetic basis for the putative abnormality of serine–threonine phosphorylation which characterizes a significant proportion of women with typical PCOS (Dunaif *et al.*, 1995). This observation is particularly intriguing, given that Miller's group have shown that serine phosphorylation is an important process in post-translational regulation of 17,20-lyase activity in steroidogenic tissue (Zhang *et al.*, 1995). These findings have led Miller and colleagues to put forward the hypothesis that a common, perhaps genetically-determined, biochemical abnormality could result in both insulin resistance and hyperandrogenism in patients with PCOS (Zhang *et al.*, 1995).

The insulin gene

Abnormalities of insulin secretion have been reported in recent studies of women with PCOS, with and without a family history of NIDDM (O'Meara *et al.*, 1993; Ehrmann *et al.*,

1995; Holte *et al.*, 1994, 1995; Dunaif and Finegood, 1996). Recent data from the Uppsala group have demonstrated that whereas insulin resistance was largely reversible by weight reduction (in obese PCOS subjects), an abnormality of first phase insulin secretion persisted, despite improved insulin sensitivity, thereby suggesting a fundamental disorder in pancreatic β-cell function (Holte *et al.*, 1995). We have therefore investigated the role of the insulin gene in the aetiology of PCOS. We evaluated the VNTR (variable number tandem repeats) minisatellite which lies 5′ to the insulin gene on chromosome 11p15.5, since variation at this element has been directly implicated in the regulation of insulin secretion, in susceptibility to NIDDM (Bennett *et al.*, 1995) and in hyperinsulinaemia related to central obesity (Weaver *et al.*, 1992). At this locus, there is a bimodal distribution of repeats, class I alleles being short (average 40 repeats) and class III alleles much longer (average 157).

We examined linkage of PCOS to the 11p15.5 locus in 17 families with several cases of PCOS and male pattern balding. We also looked for an association between the insulin gene VNTR (particularly class I and class III alleles) and polycystic ovaries in two additional populations of women (all European) presenting with symptoms of PCOS at two different endocrine centres (Waterworth *et al.*, 1997). We calculated the odds ratios for insulin VNTR genotypes either by using a conventional case-control approach (subjects from the St Mary's Hospital population) or by the use of affected family-based controls (AFBAC) (the Middlesex Hospital).

AFBAC and a related technique, the transmission disequilibrium test (TDT), are applicable if DNA is available from the proband and both parents (Spielman and Ewens, 1996). These methods compare alleles transmitted from parents to affected offspring with those not so transmitted. The latter generate 'control' genotypes or alleles which are matched for ethnicity to those in the sample from the affected case.

We found that class III alleles were associated with PCOS in each of the three populations (Figure 3). An important additional finding was that insulin VNTR class III alleles were most strongly associated with anovulatory PCOS. This is in keeping with the observation that hyperinsulinaemia is a more prominent feature in women with polycystic ovaries who have anovulatory menses (or amenorrhoea) than in equally hyperandrogenaemic subjects with regular menses (Dunaif *et al.*, 1987; Robinson *et al.*, 1993).

Another intriguing finding emerged from TDT analysis of the Middlesex Hospital population and of the 17 families with PCOS. Class III alleles were transmitted significantly more often from fathers than from mothers (Bennett *et al.*, 1997). This 'parent-of-origin effect' suggests genetic imprinting, as has previously been described for 11p15.5 in relation to type 1 (insulin dependent) diabetes (Bennett and Todd, 1996).

In the families, non-parametric linkage analysis was performed with the aid of five polymorphic markers in the region of 11p15.5, using the GENEHUNTER programme. We found evidence for excess allele sharing at the insulin gene VNTR locus, giving a maximum NPL score of 3.250 ($P = 0.002$). Using parametric analysis, we estimated that approximately 60% of families showed linkage to this locus. When we

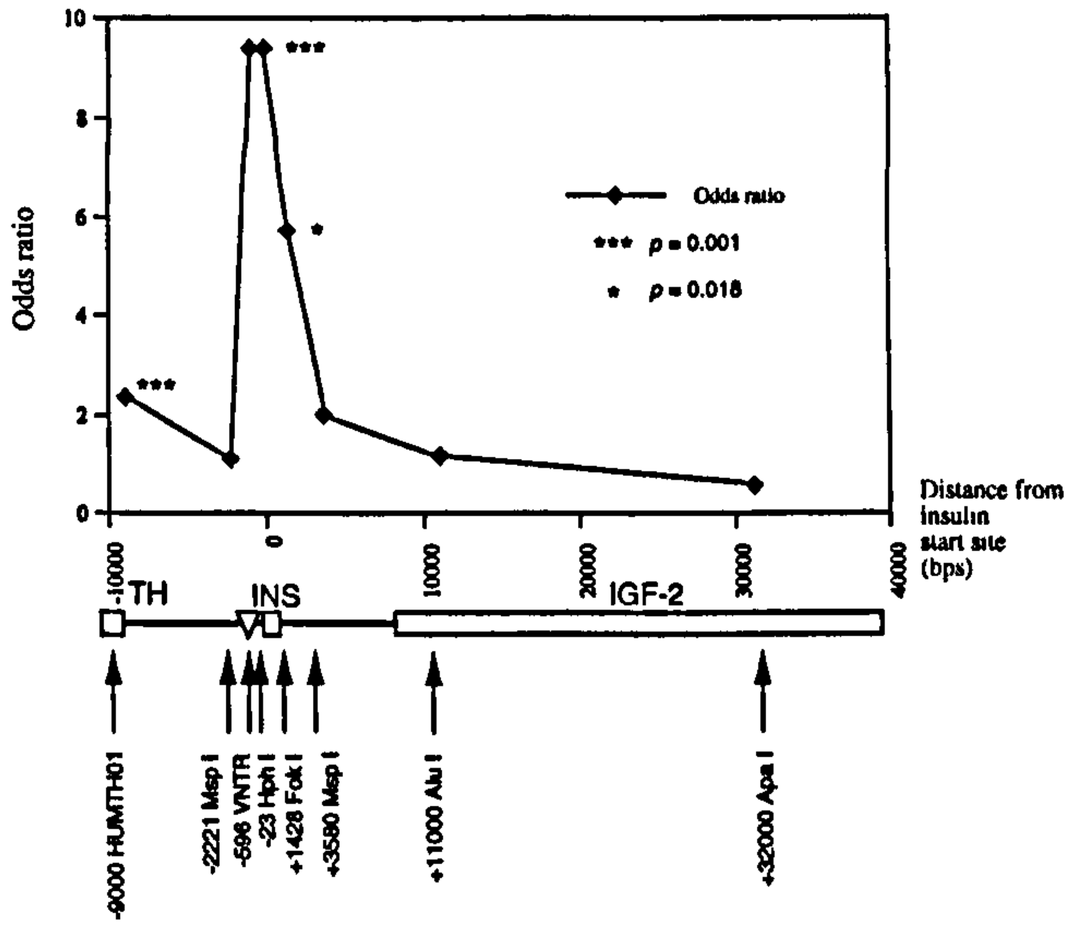


Figure 3. Localization of polycystic ovary syndrome (PCOS) susceptibility to the VNTR of the insulin gene on chromosome 11. Odds ratio (OR) values calculated for probands with anovulatory PCOS compared to non-affected controls. At the insulin gene VNTR locus, class III homozygotes were compared, similarly for alleles at neighbouring polymorphisms in linkage disequilibrium with class III VNTR alleles. 95% CI and P values for VNTR were: 2.01–44.2, $P = 0.001$ (from Waterworth *et al.*, 1997, with permission).

assigned data from families according to linkage score, we found that the geometric mean of fasting specific insulin levels was higher in those families with a positive LOD score than those with a negative score (Waterworth *et al.*, 1997).

In summary, in three different populations, we have uncovered strong evidence for both linkage and association between alleles at the VNTR 5′ to the insulin gene and PCOS. We conclude, from these data, that the VNTR of the insulin gene is a major susceptibility locus for PCOS, particularly anovulatory PCOS, and may contribute to the mechanism of hyperinsulinaemia and to the high risk of NIDDM in women with PCOS.

Future studies

To date, the approach we have taken of exploring candidate genes in both association and linkage studies has paid some dividends. Two loci, one related to control of androgen biosynthesis and one related to insulin secretion, have been identified as being of potential aetiological significance. It is important, however, for these findings to be supported in studies of other populations of women with PCOS (paying attention to any effect of ethnic origin) and to consider other methods for future studies. The potential pitfalls of case-control studies in relatively small populations have been illustrated by our experience with *CYP17*. Although we used a similar approach for the studies of *CYP11a*, these findings are likely to prove more robust for the following reasons: (i) the number of subjects examined in the case-control study was greater than in the initial *CYP17* study; (ii) in contrast to *CYP17*, we found a physiological correlation between alleles

of *CYP11a* and serum testosterone, supporting the concept that the variant allele has an effect on androgen production; (iii) non-parametric linkage analysis was undertaken, allowing for the fact that the mode of inheritance of PCOS remains uncertain.

As far as the insulin gene VNTR is concerned, the number of subjects studied in each population was not large but the consistency of results, using different methods in three separate groups of subjects, suggests that this is likely to be a sustainable finding. In studies of the insulin gene VNTR, the results of linkage analyses were similar even if data from the men in these families were omitted. This indicates that the results were not reliant on the, still controversial, assignment of premature balding as the male phenotype.

Nevertheless, consolidation of these findings and the search for other susceptibility genes demands an approach in which most of the disadvantages outlined above can be avoided. We believe that the strategy of assessing candidate genes remains viable. A more extensive 'anonymous' genome-wide scan to identify other susceptibility loci is also valid but requires many more subjects. Of course, the two approaches are not mutually exclusive. Linkage studies using families with several cases of PCOS/male balding are difficult, given the uncertainty, for example, about assignment of post-menopausal women and about the male phenotype. This has prompted us, and others, to consider using affected sibling pairs in which the minimum family unit would be two sisters with documented clinical, biochemical and ultrasonographic evidence of PCOS. Large numbers and resources are needed, especially for a genome-wide scan, but this approach conveniently side-steps the problem of the male phenotype and that of identifying, reliably, unaffected controls. Another strategy in association studies, which we have already found useful in the context of the insulin gene, is the use of AFBAC and TDT, as described above.

Summary

Using a candidate gene approach, we have found evidence for the involvement of two key genes in the aetiology of PCOS. From the results of both linkage and association studies, we suggest that the steroid synthesis gene *CYP11a* and the insulin VNTR regulatory polymorphism are important factors in the genetic basis of PCOS and may go some way to explaining the heterogeneity of the syndrome. Thus, differences in expression of *CYP11a* could account for variation in androgen production in women who have polycystic ovaries. We postulate that those subjects carrying class III alleles at the insulin gene VNTR locus are more likely to be hyperinsulinaemic and to suffer from menstrual disturbances.

These findings remain to be confirmed in larger studies and in other populations but, whatever the outcome of such studies, it is unlikely that these are the only genes to be involved in the aetiology of PCOS. Our earlier hypothesis, based on the initial family studies, that PCO/male balding could be explained by a single gene effect is no longer tenable. Our recent results lend weight to the idea that PCOS is an oligogenic disorder although it is quite possible that, within a given family, there is indeed one major gene which is dominantly inherited. Thus

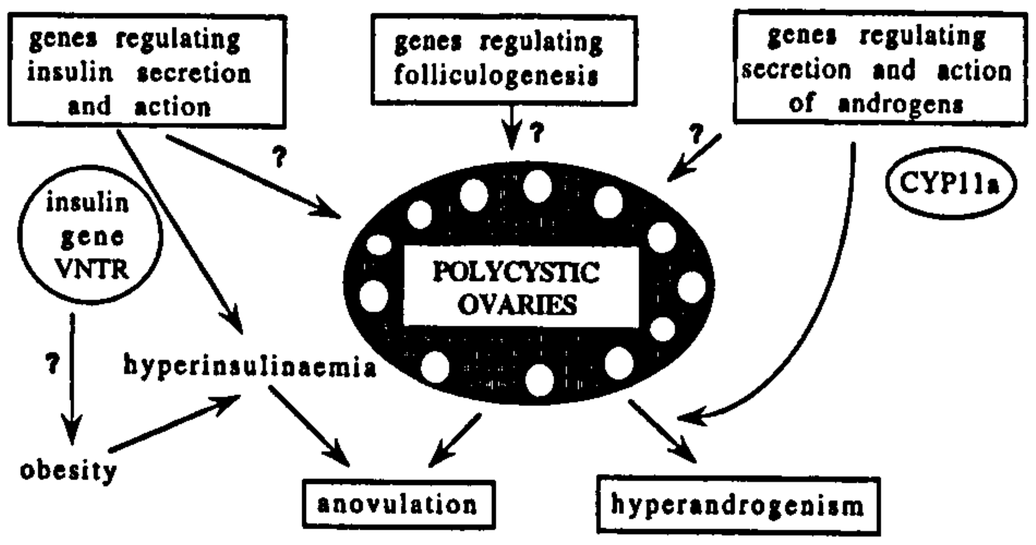


Figure 4. Suggested aetiological factors in polycystic ovary syndrome (PCOS).

PCOS appears to represent a quantitative trait in which a relatively small number of key genes contribute, in conjunction with environmental (particularly nutritional) factors, to the observed clinical and biochemical heterogeneity.

We propose that the underlying problem is the development of the polycystic ovarian morphology with an implicit disorder of folliculogenesis. This predisposes the subject to the development of polycystic ovary syndrome. The gene(s) determining the development of this distinct ovarian morphology remain unknown. *CYP11a* and insulin gene VNTR may act independently or in concert to determine abnormalities of ovarian function and (in the case of insulin) metabolism (Figure 4). It is possible that hyperinsulinaemia contributes to the morphological as well as the biochemical features of the polycystic ovary. Insulin has been shown to be even more effective than insulin-like growth factor-I in stimulating proliferation of ovarian stromal cells (Watson *et al.*, 1997).

Environmental factors can alter the clinical and biochemical presentation in those with a genetic predisposition to PCOS. This is illustrated by the effect of obesity (or, conversely, calorie restriction) on serum insulin levels, insulin sensitivity and menstrual function (Dunaif *et al.*, 1987; Holte *et al.*, 1995; Franks *et al.*, 1996b).

Identification of susceptibility genes in the aetiology of PCOS is not simply an intellectual exercise, as illustrated by the data regarding the insulin gene VNTR. If these findings are substantiated in further phenotype/genotype studies, they offer the prospect of a clinically important genetic marker, not only for PCOS but also for the future risk of NIDDM.

Acknowledgements

We thank Dr A.H. Carey, Dr C.M-T. Gilling-Smith and Ms R. Joseph-Horne (St Mary's Hospital, London), Dr G.S. Conway (The Middlesex Hospital, London), Mr S. Hague (Department of Biochemistry and Molecular Genetics, Imperial College School of Medicine at St Mary's, London), Dr S.T. Bennett and Professor J.A. Todd (Wellcome Trust Centre for Human Genetics, University of Oxford) for their invaluable contributions to these studies. We are very grateful for grant support from The Medical Research Council (Studentship for D.W. and ROPA award to S.F. and R.W.) and from the research department of Unilever, UK.

References

Adams, J., Polson, D.W. and Franks S. (1986) Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *Brit. Med. J.*, **293**, 355–359.

Barnes, R.B., Rosenfield, R.L., Burstein, S. and Ehrmann D.A. (1989). Pituitary-ovarian responses to nafarelin testing in the polycystic ovary syndrome. *N. Engl. J. Med.*, **320**, 559–565.

Bennett, S.T. and Todd, J.A. (1996) Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Ann. Rev. Genet.*, **30**, 343–370.

Bennett, S.T., Lucassen, A.M., Gough, S.C.L. *et al.* (1995) Susceptibility to human type 1 diabetes at *IDDM2* is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.*, **9**, 284–292.

Bennett, S.T., Todd, J.A., Waterworth, D.M. *et al.* (1997) Association of insulin gene VNTR polymorphism with polycystic ovary syndrome (letter). *Lancet*, **349**, 1771–1772.

Carey, A.H., Chan, K.L., Short, F. *et al.* (1993) Evidence for a single gene effect in polycystic ovaries and male pattern baldness. *Clin. Endocrinol.*, **38**, 653–658.

Carey, A.H., Waterworth, D., Patel, K. *et al.* (1994) Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum. Mol. Genet.*, **3**, 1873–1876.

Clayton, R.C., Ogden, V., Hodgkinson, J. *et al.* (1992) How common are polycystic ovaries in normal women and what is their significance for fertility in the general population? *Clin. Endocrinol.*, **37**, 127–134.

Conway, G.S., Avey, C. and Rumsby, G. (1994) The tyrosine kinase domain of the insulin receptor gene is normal in women with hyperinsulinaemia and polycystic ovary syndrome. *Hum. Reprod.*, **9**, 1681–1683.

Conway, G.S., Honour, J.W. and Jacobs, H.S. (1989) Heterogeneity of the polycystic ovary syndrome: clinical, endocrine and ultrasound features in 556 patients. *Clin. Endocrinol.*, **30**, 459–470.

Cooper, H., Spellacy, W., Prem, K. and Cohen, W. (1968) Hereditary factors in the Stein-Leventhal syndrome. *Am. J. Obstet. Gynecol.*, **100**, 371–387.

Dahlgren, E., Johannson, S., Lindstedt, G. *et al.* (1992a) Women with polycystic ovary syndrome wedge resected in 1956 to 1965: a long term follow up focusing on natural history and circulating hormones. *Fertil. Steril.*, **57**, 505–513.

Dahlgren, E., Janson, P.O., Johansson, S. *et al.* (1992b) Polycystic ovary syndrome and risk for myocardial infarction: evaluated from a risk factor model based on a prospective study of women. *Acta Obstet. Gynecol. Scand.*, **71**, 599–604.

Dunaif, A. (1993) Insulin resistance and ovarian dysfunction. In Moller, D. (ed.) *Insulin Resistance*. Wiley, New York, pp. 301–325.

Dunaif, A., Graf, M., Mandeli, J. *et al.* (1987) Characterization of groups of hyperandrogenic women with acanthosis nigricans, impaired glucose tolerance, and/or hyperinsulinemia. *J. Clin. Endocrinol. Metab.*, **65**, 499–507.

Dunaif, A., Xia, J., Book, C-B. *et al.* (1995) Excessive insulin receptor phosphorylation in cultured fibroblasts and in skeletal muscle. *J. Clin. Invest.*, **96**, 801–810.

Dunaif, A. and Finegood, D.T. (1996) β cell dysfunction independent of obesity in the polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **81**, 942–947.

Ehrmann, D., Sturis, J., Byrne, M. *et al.* (1995) Insulin secretory defects in polycystic ovary syndrome: relationship to insulin sensitivity and family history of non-insulin-dependent diabetes mellitus. *J. Clin. Invest.*, **96**, 520–527.

Farquhar, C.M., Birdsall, M., Manning, P. *et al.* (1994) The prevalence of polycystic ovaries on ultrasound scanning in a population of randomly selected women. *Aust. N. Z. J. Obstet. Gynaecol.*, **34**, 67–72.

Ferriman, D. and Purdie, A.W. (1979) The inheritance of polycystic ovarian disease and a possible relationship to premature balding. *Clin. Endocrinol.*, **11**, 291–300.

Franks, S. (1989) Polycystic ovary syndrome: a changing perspective. *Clin. Endocrinol.*, **31**, 87–120.

Franks, S. (1991) The ubiquitous polycystic ovary. *J. Endocrinol.*, **129**, 317–319.

Franks, S. (1995) Medical progress article: polycystic ovary syndrome. *N. Engl. J. Med.*, **333**, 853–861.

Franks, S. (1997) The 17α-hydroxylase-17,20-lyase gene (CYP17) and polycystic ovary syndrome (commentary). *Clin. Endocrinol.*, **46**, 135–136.

Franks, S., Willis, D., Mason, H. and Gilling-Smith, C. (1996a) Comparative androgen production from theca cells of normal women and women with polycystic ovaries. In Chang, R.J. (ed.) *Polycystic Ovary Syndrome*. Springer, New York, pp. 154–164.

Franks, S., Robinson, S. and Willis, D. (1996b) Nutrition, insulin and polycystic ovary syndrome. *Rev. Reprod.*, **1**, 47–53.

Gharani, N., Waterworth, D.M., Williamson, R. and Franks, S. (1996) 5′ polymorphism of the CYP17 gene is not associated with serum testosterone levels in women with polycystic ovaries (letter). *J. Clin. Endocrinol. Metab.*, **81**, 4174.

Gharani, N., Waterworth, D.M., Batty, S. *et al.* (1997) Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum. Mol. Genet.*, **6**, 397–402.

Gilling-Smith, C., Willis, D.S., Beard, R.W. and Franks, S. (1994) Hypersecretion of androstenedione by isolated theca cells from polycystic ovaries. *J. Clin. Endocrinol. Metab.*, **79**, 1158–1165.

Gilling-Smith, C., Story, E.H., Rogers, V. and Franks, S. (1997) Evidence for a primary abnormality of thecal cell steroidogenesis in the polycystic ovary syndrome. *Clin. Endocrinol.*, **47**, 93–99.

Givens, J.R. (1988) Familial polycystic ovarian disease. *Endocrinol. Metab. Clin. N. Am.*, **17**, 771–783.

Hague, W.M., Adams, J., Reeders, S.T. *et al.* (1988). Familial polycystic ovaries: a genetic disease? *Clin. Endocrinol.*, **29**, 593–605.

Harada, N., Ogawa, H., Shozu, M. and Yamada, K. (1992) Genetic studies to characterise the origin of the mutation in placental aromatase deficiency. *Am. J. Hum. Genet.*, **51**, 666–667.

Holte, J. (1996) Disturbances in insulin secretion and sensitivity in women with the polycystic ovary syndrome. *Clin. Endocrinol. Metab.*, **10**, 221–247.

Holte, J., Bergh, T., Berne, C. *et al.* (1994) Enhanced early insulin response to glucose in relation to insulin resistance in women with polycystic ovary syndrome and normal glucose tolerance. *J. Clin. Endocrinol. Metab.*, **78**, 1052–1058.

Holte, J., Bergh, T., Berne, C. *et al.* (1995) Restored insulin sensitivity but persistently increased early insulin secretion after weight loss in obese women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **80**, 2586–2593.

Hull, M.G. (1987) Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. *Gynecol. Endocrinol.*, **1**, 235–245.

Ibañez, L., Hall, J.E., Potau, N. *et al.* (1996) Ovarian 17-hydroxyprogesterone hyperresponsiveness to gonadotropin-releasing hormone (GnRH) agonist challenge in women with polycystic ovary syndrome is not mediated by luteinizing hormone hypersecretion: evidence from GnRH agonist and human chorionic gonadotropin stimulation testing. *J. Clin. Endocrinol. Metab.*, **81**, 4103–4107.

Ito, Y., Fisher, C.R., Conte, F.A. *et al.* (1993) Molecular basis of aromatase deficiency in an adult female with sexual infantilism and polycystic ovaries. *Proc. Natl. Acad. Sci. USA*, **90**, 11673–11677.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. *et al.* (1996) Parametric and non-parametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.

Legro, R.S. (1995) The genetics of polycystic ovary syndrome. *Am. J. Med.*, **98**, suppl 1A, 9S–16S.

Lunde, O., Magnus, P., Sandvik, L. and Hoglo, S. (1989) Familial clustering in the polycystic ovarian syndrome. *Gynecol. Obstet. Invest.*, **28**, 23–30.

Mason, H.D., Willis, D.S., Beard, R.W. *et al.* (1994) Estradiol production by granulosa cells of normal and polycystic ovaries: relationship to menstrual cycle history and to concentrations of sex steroids in follicular fluid. *J. Clin. Endocrinol. Metab.*, **79**, 1355–1360.

Moller, D.E. and Flier, J.S. (1988) Detection of an alteration in the insulin receptor gene in a patient with insulin resistance, acanthosis nigricans and the polycystic ovary syndrome. *N. Engl. J. Med.*, **319**, 1526–1529.

O'Driscoll, J.B., Mamtora, H., Higginson, J. *et al.* (1994) A prospective study of the incidence of clear-cut endocrine disorders and polycystic ovaries in 350 patients with hirsutism or androgenic alopecia. *Clin. Endocrinol.*, **41**, 231–236.

O'Meara, N., Blackman, J.D., Ehrmann, D.A. *et al.* (1993) Defects in B-cell function in functional ovarian hyperandrogenism. *J. Clin. Endocrinol. Metab.*, **76**, 1241–1247.

Polson, D.W., Adams, J., Wadsworth, J. and Franks, S. (1988) Polycystic ovaries - a common finding in normal women. *Lancet*, **1**(8590), 870–872.

Pugeat, M., Nicolas, M.H., Cousin, P. *et al.* (1996) Polymorphism in the 5′ promoter of the human gene encoding P450c17α and adrenal androgen secretion in hirsute women. *Programme of 10th International Congress of Endocrinology, San Francisco, June 1996.* Endocrine Society Press, Bethesda, p. 561.

Robinson, S., Kiddy, D., Gelding, S.V. *et al.* (1993) The relationship of insulin sensitivity to menstrual pattern in women with hyperandrogenism and polycystic ovaries. *Clin. Endocrinol.*, **39**, 351–355.

Rosenfield, R.L., Barnes, R.B., Cara, J.F. and Lucky, A.W. (1990) Dysregulation of cytochrome P450c 17 alpha as the cause of polycystic ovarian syndrome. *Fertil. Steril.*, **53**, 785–791.

Simpson, J.L. (1992) Elucidating the genetics of polycystic ovary syndrome. In Dunaif, A., Givens, J.R., Haseltine, F.P. and Merriam, G.R. (eds), *Polycystic Ovary Syndrome*. Blackwell Scientific, Oxford, pp. 59–77.

Spielman, R.S. and Ewens, W.J. (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.*, **59**, 983–989.

Takayama, K., Takao, T., Hironobu, S. *et al.* (1996) Immunohistochemical study of steroidogenesis and cell proliferation in polycystic ovarian syndrome. *Hum. Reprod.*, **11**, 1387–1392.

Talbot, J.A., Bicknell, E.J., Rajkhowa, M. *et al.* (1996). Molecular scanning of the insulin receptor gene in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **81**, 1979–1983.

Tetchatraisak, K., Conway, G.S. and Rumsby G. (1997) Frequency of a polymorphism in the regulatory region of the 17a-hydroxylase-17,20 lyase (CYP17) gene in hyperandrogenic states. *Clin. Endocrinol.*, **46**, 131–134.

Waterworth, D.M., Bennett, S.T., Gharani, N. *et al.* (1997) Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. *Lancet*, **349**, 986–989.

Watson, H., Willis, D., Mason, H. *et al.* (1997) The effects of ovarian steroids, epidermal growth factor, insulin and insulin-like growth factor-I on ovarian stromal cell growth. *Proceedings of the Endocrine Society, 79th Annual Meeting, June 11–14, 1997, Minneapolis, Minnesota.* Endocrine Society Press, Bethesda, p. 389.

Weaver, J.U., Kopelman, P.G. and Hitman, G.A. (1992) Central obesity and hyperinsulinaemia in women are associated with polymorphism in the 5′ flanking region of the human insulin gene. *Euro. J. Clin. Invest.*, **22**, 265–270.

Zawadski, J.K. and Dunaif, A. (1992) Diagnostic criteria for polycystic ovary syndrome towards a rational approach. In Dunaif, A., Givens, J.R., Haseltine, F.P. and Merriam, G.R. (eds), *Polycystic Ovary Syndrome*. Blackwell Scientific Publications, Oxford, pp. 377–384.

Zhang, L.H., Rodriguez, H., Ohno, S. and Miller, W.L. (1995) Serine phosphorylation of human P450c17 increases 17,20-lyase activity: implications for adrenarche and the polycystic ovary syndrome. *Proc. Natl. Acad. Sci. USA*, **92**, 10619–10623.

Received on June 10, 1997; accepted on August 28, 1997