POLYCYSTIC OVARY SYNDROME: FROM PHENOTYPE TO GENOTYPE

Yvonne Vera Louwers

Polycystic Ovary Syndrome: from phenotype to genotype
Thesis, Erasmus University Rotterdam, The Netherlands

The printing of this thesis has been financially supported by the Department of Obstetrics and Gynaecology, Erasmus MC Rotterdam, the Erasmus University Rotterdam and Nederlandse Vereniging voor Obstetrie en Gynaecologie.

Further support for this dissertation was kindly provided by:
Goodlife Healthcare

Cover: Sketch by Yvonne Louwers
Lay-out: Legatron Electronic Publishing, Rotterdam
Printing: Ipskamp Drukkers BV, Enschede

ISBN/EAN: 978-94-6259-027-4

POLYCYSTIC OVARY SYNDROME: FROM PHENOTYPE TO GENOTYPE

Het polycysteus ovarium syndroom: van fenotype naar genotype

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H. A. P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 26 februari 2014 om 13.30 uur

door

Yvonne Vera Louwers

geboren te Rotterdam

Promotiecommissie

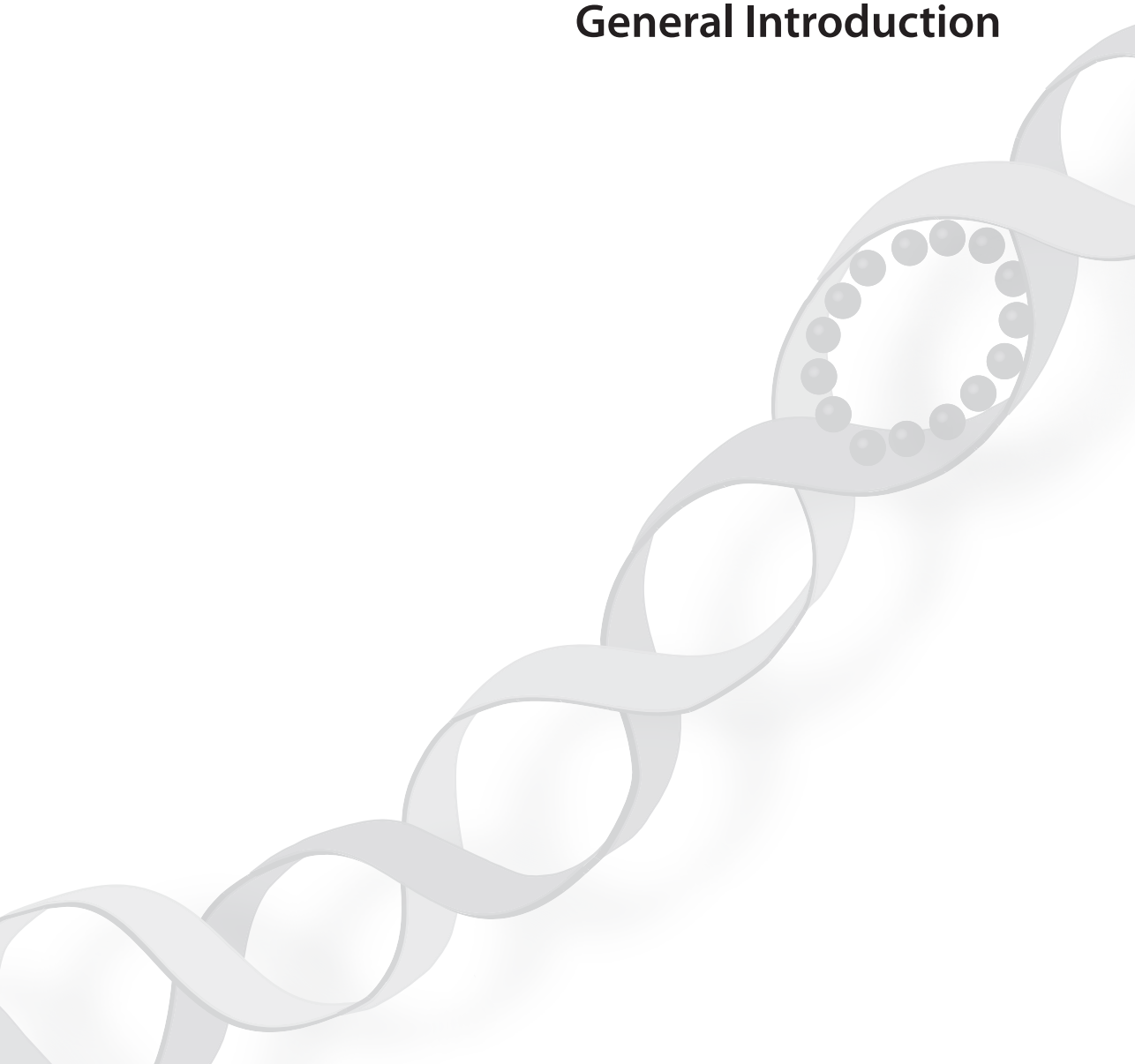
Promotoren:	Prof.dr. J.S.E. Laven
	Prof.dr. A.G. Uitterlinden
Overige leden:	Prof.dr. F.H. de Jong
	Prof.dr. E.J.G. Sijbrands
	Prof.dr. B.C.J.M. Fauser
Copromotor:	Dr. L. Stolk

Paranimfen:	K.M. Louwers
	W. van Dorp

Table of contents

Chapter 1

General Introduction



The phenotype of the polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder among women of reproductive age. It has an estimated prevalence of 5-15% worldwide [1-3]. PCOS represents the major cause of anovulatory infertility and is associated with distressing cutaneous manifestations of androgen excess, such as hirsutism and acne [4,5]. Moreover, obesity as well as insulin resistance co-occur with the syndrome [6,7]. Hence, it is a life-long condition with adverse reproductive as well as metabolic implications in affected women [8]. The pathogenesis of PCOS is multifactorial and far from being fully understood [9,10].

Historical (dis)agreement on the definition of PCOS

PCOS was first described in 1935 by Dr. Irving F. Stein and Dr. Michael L. Leventhal, who evaluated seven obese patients suffering from anovulation and having enlarged ovaries [11]. Since their observation, several diagnostic criteria have been proposed to define PCOS (Table 1). In 1990, the National Institutes of Health (NIH) provided the first set of diagnostic criteria and defined PCOS as the co-existence of ovulatory dysfunction and hyperandrogenism [12]. The Rotterdam consensus meeting in 2003 recognized PCOS as a syndrome that encompasses a broad spectrum of signs and symptoms, including at least two of the following: ovulatory dysfunction, hyperandrogenism and polycystic ovarian morphology (PCOM) [4,5]. Other causes that mimic these features, such as congenital adrenal hyperplasia, Cushing syndrome, androgen-producing tumours and hyperprolactinemia, should be excluded. In 2006, the Androgen Excess and PCOS Society (AES) proposed the AE-PCOS criteria emphasizing that hyperandrogenism should be considered as the key-feature of PCOS [13].

Obviously, this continuous debate regarding the definition has challenged clinical care and hampered research in this field because of lack of consensus. Therefore, in December 2012, the NIH organized the Evidence-based Methodology Workshop on PCOS [14]. During this meeting the PCOS definition was critically reappraised. An independent panel recommended maintaining the broad diagnostic criteria as formulated during the consensus meeting held in 2003 in Rotterdam [4,5].

Table 1 | Diagnostic criteria for PCOS and accompanying component phenotypes (1-2-3) and phenotype groups (A-B-C-D).

	Phenotype groups	NIH 1990	Rotterdam 2003	AE-PCOS 2006
A	1. Oligo-anovulation 2. Hyperandrogenism and/or hirsutism 3. Polycystic ovarian morphology	+	+	+
B	1. Oligo-anovulation 2. Hyperandrogenism and/or hirsutism	+	+	+
C	2. Hyperandrogenism and/or hirsutism 3. Polycystic ovarian morphology		+	+
D	1. Oligo-anovulation 3. Polycystic ovarian morphology		+	

NIH, National Institute of Health; *AE-PCOS,* Society of Androgen Excess and Polycystic Ovary Syndrome. Modified from Azziz et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report [15].

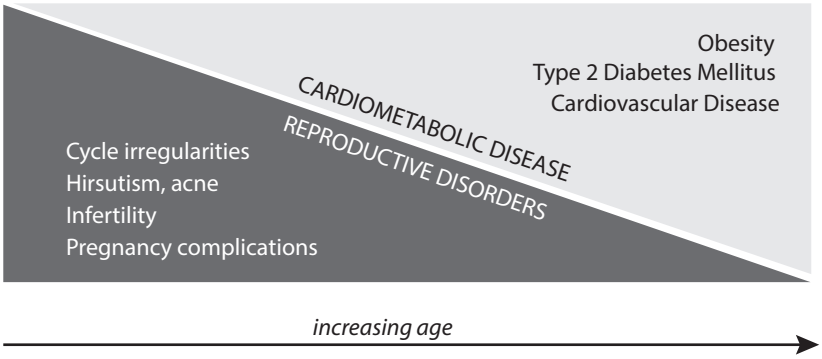
Variability within the PCOS phenotype

PCOS is usually diagnosed during the early reproductive years. It is a life-long lasting condition with substantial variation of symptomatology regarding reproductive, metabolic and cardiovascular health throughout a woman's life (Figure 1) [8]. During their reproductive years, patients with PCOS suffer from oligo-ovulation or anovulation resulting in decreased fertility [16,17]. Moreover, PCOS seems to be associated with an increased risk of complications during pregnancy, such as gestational diabetes, pregnancy-induced hypertension and preeclampsia [18]. Ovarian and/or adrenal hyperandrogenism is present in 50-80% of all patients, depending on which diagnostic criteria are used [19,20]. Hirsutism is the most common clinical sign of hyperandrogenemia. Acne and alopecia androgenetica are present to a lesser extent [21].

The metabolic disturbances become more pronounced with increasing age. Insulin resistance and hyperinsulinemia occur in 50-70% of all patients with PCOS [7,22]. Subsequently, patients are at increased risk of developing impaired glucose tolerance and type 2 diabetes mellitus [23]. Moreover, PCOS is associated with dyslipidemia, increased carotid intima-media thickness, increased coronary artery calcification and subclinical cardiovascular disease [24-27]. Overweight or obesity is present in up to 80% of all patients [6]. It aggravates the metabolic co-morbidities associated with PCOS [28]. Lifestyle interventions and weight-loss not only improve the metabolic disturbances, but also ameliorate the reproductive phenotype [29]. However, obesity does not fully account for the burden of these long-term health complications. Patients with PCOS also experience these complications independently of their BMI levels [30]. Adequately powered long-term follow-up studies of PCOS patients monitoring hard endpoints of cardiovascular risk, such as stroke, myocardial infarction, and mortality, have not been conducted yet. Hence, although suggested by surrogate makers,

it is still unclear whether or not patients with PCOS encounter a substantially increased risk of developing cardiovascular events. Moreover, these increased risks might be limited to patients with certain phenotypic characteristics. Therefore, after the diagnosis PCOS has been established, considerable differences amongst these patients might be present in terms of treatment-outcome, associated co-morbidities and the impact of the long-term health sequelae.

Figure 1 | The phenotype of PCOS varies throughout life.



Modified from Fauser and Balen ESHRE/ASRM PCOS Consensus meeting [8].

Pathogenesis of PCOS

The following paragraph provides a global overview of the dynamic processes occurring during an ovulatory menstrual cycle [31,32]. Moreover, it describes the potential underlying mechanisms of PCOS. However, these mechanisms are far from fully understood. Schematic representation of these processes is displayed in Figure 2.

Ovarian physiology

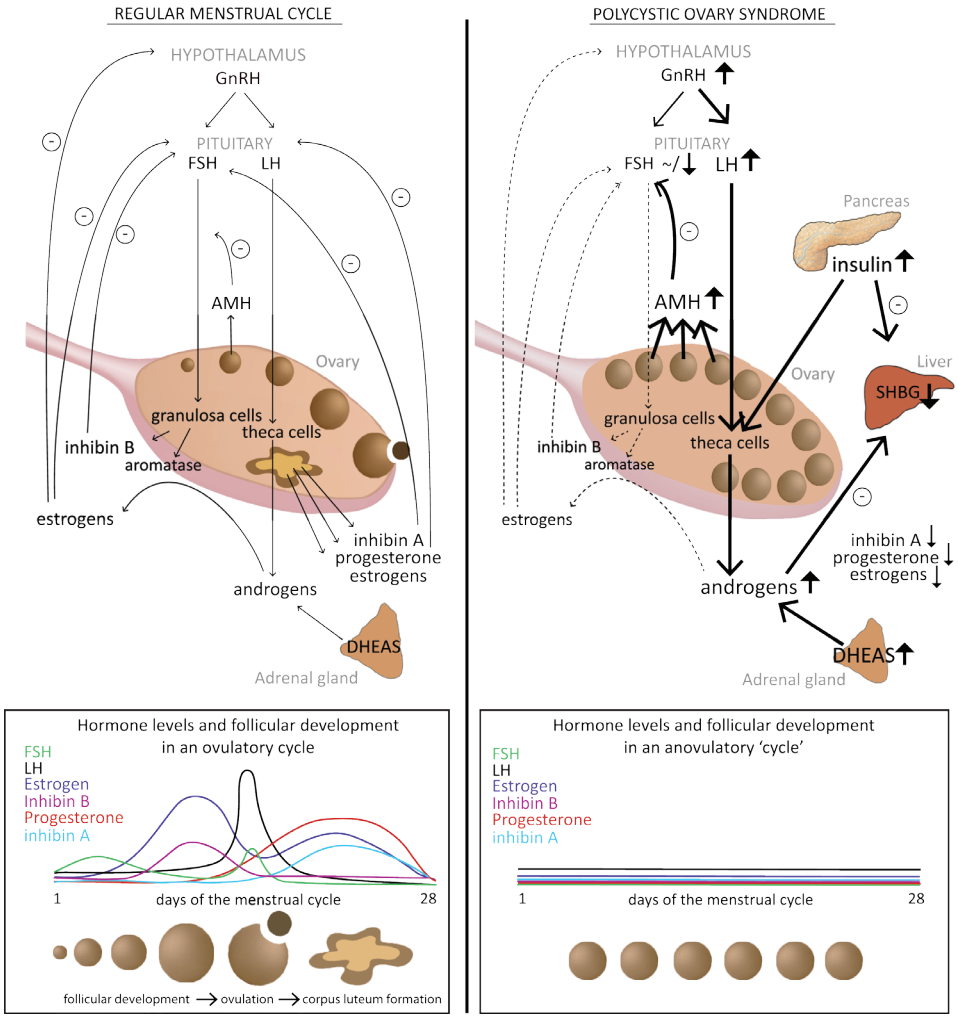
The menstrual cycle is an orderly sequence of events regulated by hypothalamic, pituitary and gonadal hormones. Normal reproductive function involves monthly follicular development, ovulation and preparation of the endometrium for implantation. The neurons of one of the basal nuclei in the hypothalamus control the pulsatile release of gonadotropin-releasing hormone (GnRH). In response to a slow GnRH pulse frequency, the anterior pituitary produces follicle-stimulating hormone (FSH). During the follicular phase, i.e. the first part of

the menstrual cycle, FSH gradually increases and is responsible for follicular development in the ovary. FSH stimulates the secretion of inhibin B from the granulosa cells. In response to luteinizing hormone (LH), the theca cells are stimulated to produce androgens that can then be converted, through FSH-induced aromatization, to estrogens in the granulosa cells. The secretion of inhibin B and estrogens rises with increasing diameter of follicles and suppresses further release of FSH by the pituitary via a negative feedback mechanism. The influence of estrogens on LH release varies with concentration and duration of exposure. At low levels, estrogens have a negative feedback on LH release. However, when estrogen levels reach a certain threshold, this negative feedback changes into a positive feedback triggering an increase in GnRH pulse frequency resulting in the LH surge. LH initiates luteinisation of the granulosa cells and synthesis of progesterone. Subsequently, the follicle ruptures, ovulation occurs and the corpus luteum is formed. The corpus luteum produces progesterone, inhibin A and estrogens. Elevated progesterone levels inhibit LH secretion through negative feedback at both the hypothalamic as well as the pituitary level. Inhibin A suppresses FSH production from the pituitary during the early and mid-luteal phase. In early pregnancy, human chorionic gonadotropin (hCG) rescues the corpus luteum. However, when a pregnancy does not occur, LH levels fall giving rise to the demise of the corpus luteum. This in turn results in a decrease in progesterone, inhibin A and estrogen levels. The decrease in inhibin A removes a suppressing influence on FSH secretion in the pituitary. The decrease in estrogen and progesterone results in an increase in GnRH pulse frequency. This increase in pulse frequency in combination with the removal of the suppressing influence on the pituitary leads to an increase in FSH levels allowing a new cohort follicles to be recruited.

Granulosa cells of primary and small antral follicles produce Anti Müllerian Hormone (AMH) [33]. AMH inhibits initial recruitment of primordial follicles and attenuates follicle sensitivity to FSH, thereby affecting folliculogenesis [33]. AMH production steadily declines as follicles grow. This decline appears to be an important requirement for dominant follicle selection and progression to ovulation [34].

The ovaries as well as the adrenal glands contribute to the androgen production in women. Dehydroepiandrosterone (DHEAS) is almost exclusively produced by the adrenal gland. DHEAS can be converted to dehydroepiandrosterone (DHEA), which is the precursor for biologically active androgens and estrogens [35]. One of these biologically active androgens is testosterone. The ovary is responsible for approximately half of the circulating testosterone, whereas the remaining part is produced by the adrenal gland. In the circulation testosterone is bound to albumin and sex-hormone binding globulin (SHBG), only around 1% of testosterone circulates freely and is biologically active in women.

Figure 2 | Schematic presentation of the processes and accompanying fluctuations of hormone levels during a regular ovulatory menstrual cycle (on the left) and a anovulatory cycle in PCOS (on the right).



The arrows represent a stimulus, unless otherwise indicated (-). The thicker arrows in the right figure represent an increase in production in PCOS compared to the production during an ovulatory menstrual cycle, whereas the dotted arrows indicate a decrease in hormone production. *GnRH,* Gonadotropin Releasing Hormone; *FSH,* Follicle Stimulating Hormone; *LH,* Luteinizing Hormone; *AMH,* Anti Müllerian Hormone; *SHBG,* Sex Hormone Binding Globulin; *DHEAS,* Dehydroepiandosterone-sulphate. The processes are described in more detail in the text.

Pathophysiology of PCOS

The pathophysiology of PCOS is far from fully understood. However, over the last decades potential underlying mechanisms have been proposed. An increased GnRH pulse frequency resulting in LH hypersecretion has been reported [36-38]. As stated, LH stimulates the ovarian theca cells to produce androgens, such as testosterone. Because of a relative FSH deficiency, testosterone is incompletely aromatized by the granulosa cells, resulting in hyperandrogenemia. There seems also to be an increase in the activity of steroidogenic enzymes in polycystic ovaries, which further adds to the androgen excess [39]. Moreover, although the ovaries are the main source of androgen excess in PCOS, also the adrenal glands contribute to the existing hyperandrogenism. Up to 50% of the patients with PCOS also have elevated DHEAS levels [40]. These elevated androgen levels can lead to symptoms of androgen excess such as hirsutism, acne and alopecia.

In PCOS, hyperandrogenism, hyperinsulinemia and altered intraovarian paracrine signalling can disrupt follicle growth [9,41]. The gonadotropin-independent development of preantral follicles seems disordered in PCOS [41]. Moreover, FSH does not seem to increase to the threshold levels which are required to stimulate normal follicular maturation resulting in follicular arrest [41,42]. The accumulation of small antral follicles result in elevated AMH levels [34,43,44]. These increased AMH levels seem to add to the anovulation in PCOS by reducing both primordial follicle growth and follicle sensitivity to FSH [34].

Hyperinsulinemia aggravates the androgen excess and ovarian follicular arrest. Insulin acts synergistically with LH on theca cells to stimulate ovarian androgen production [45,46]. Moreover insulin suppresses hepatic production of SHBG, resulting in higher levels of bioavailable testosterone [47]. In turn, androgens can produce insulin resistance by directly influencing insulin action in skeletal muscle and adipocytes, by altering adipokine secretion and by increasing visceral adiposity. However, these androgen effects on insulin action are modest [7].

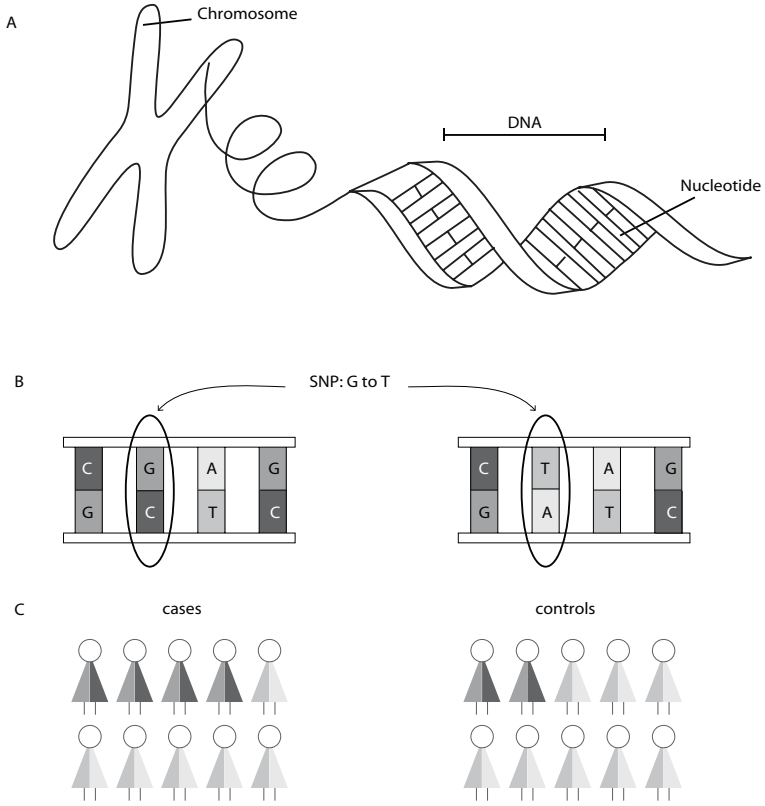
Genetics of PCOS

PCOS is considered to be a so-called complex genetic disorder. Complex genetic disorders are diseases of which the etiology is explained by the effects of genetic and environmental factors. PCOS as well as its associated co-morbidities cluster within families. Approximately 20-40% of the first-degree female family members of patients with PCOS are also diagnosed with the syndrome [48,49]. Moreover, associated co-morbidities, such as hyperandrogenemia, hyperinsulinemia and disturbed insulin secretion, also cluster within these families of patients with PCOS [50-53]. Finally, PCOS seems to be more common amongst sisters of monozygotic twin pairs compared to dizygotic twins indicating a high degree of heritability in patients with PCOS [54]. Heritability (h^2) is the proportion of the total phenotypic variation that can be attributed to additive genetic effects. For PCOS the heritability is estimated to be around 65% [54], indicating that genetic factors largely determine PCOS susceptibility.

Genetic variation in humans

The human genome is made up of Deoxyribo Nucleic Acid (DNA), which consists of a long sequence of four types of nucleotide bases, i.e. adenine (A), cytosine (C), guanine (G), and thymine (T) typically intertwined with each other [55] (Figure 3A). These nucleotides bind in pairs, i.e. A with T and C with G, resulting in so-called double-stranded helices. The complete DNA sequence of the human genome consists of about 3,3 billion base pairs. Approximately 3% of the genome consists of coding sequences [56]. It is estimated that approximately 30,000 to 40,000 protein-coding genes exist [57-59]. DNA is transcribed into messenger Ribo Nucleic Acid (mRNA), which encodes for amino acids that can built up a protein. This process is called translation. Changes in the DNA sequences can result in alterations in protein function and thereby affecting disease risk [56].

Figure 3 | Single Nucleotide Polymorphisms and genetic association studies.



Panel A shows DNA which is packed into chromosomes as two long twisted strands. ***Panel B*** shows an a G to T SNP. ***Panel C*** represent the principle used in genetic association studies: the SNP is considered to be associated with the disease, when it is more frequently present in individuals with the disease (cases) compared to individuals without the disease (controls). *DNA,* Deoxyribo Nucleic Acid; *A,* adenine; *C,* cytosine; *G,* guanine; *T,* thymine; *SNP,* Single Nucleotide Polymorphism.

The genome of any two unrelated individuals is for 99.9% identical. Consequently, the remaining 0.1% of the genome is responsible for differences between two unrelated individuals. If the DNA sequence at a given locus, i.e. an unique chromosomal location defining the position of a gene or DNA sequence, varies in the population, each different version is called an allele [60]. If there are two alleles at a given locus, the allele that is less common in the population is the minor allele. Many types of DNA sequence variations exist (56). The most common DNA variation is the Single Nucleotide Polymorphism (SNP), which constitutes a single base variation (Figure 3B). A SNP is defined as a locus at which two alleles are present, both at a frequency of at least 1% [61]. It is estimated that > 50 million SNPs are present across the human genome [62]. SNPs in protein-coding regions are either non-synonymous or synonymous, depending on whether they do or do not have a modifying effect on the amino acid sequence of the gene's protein product [57,60]. Intronic and intergenic SNPs lie in the non-coding regions of the genome. All types of SNPs can contribute to susceptibility of disease [60].

Genetic association studies

Candidate-gene studies versus genome-wide approach

Genetic association studies aim to detect the association of a disease with a SNP. If a certain SNP is more frequently present in individuals with the disease compared to individuals without the disease, the SNP is considered to be associated with the disease (Figure 3C). Two types of methods to study such associations exist: the candidate gene study and the genome-wide association study (GWAS). Using candidate gene analysis, SNPs in genes are selected based on prior knowledge of biological mechanisms underlying the disease. In case of a GWAS the association of a trait or disease with over 500,000 SNPs is tested using a hypothesis-free approach. The level of statistical significance required to establish association for a common variant in GWASs is $< 5 \times 10^{-8}$, applying Bonferroni correction for approximately 1,000,000 independent genomic markers (0.05/1 million; genome-wide significant, gws). GWAS has proven its value for implicating novel biological pathways in disease pathogenesis [63].

Correlation of genetic variants

The identified SNP may be the causal variant, but is more likely to be highly correlated to the actual causal variant due to lack of genetic recombination between the two. This concept of non-random association of alleles at 2 or more loci due to infrequent recombination events, is called linkage disequilibrium (LD). Using LD it is possible to select optimized subsets of SNPs, in such a way that all preferred variants are either directly genotyped or are strongly correlated with the genotyped SNP [57]. Indirect association mapping relies on LD in the sense that the functional variant needs not to be studied at all, as long as one measures a variant that is in LD with this functional variant [64]. The index r^2 is used as a measurement of LD and is the square of the conventional correlation coefficient of the relationship between the allele at the typed

locus and the allele at the causal locus. A high r^2, i.e. $r^2 = 1$, indicates that the alleles can predict each other completely [65]. Using the concept of r^2 it is possible to predict genotypes that are not directly measured. The term for this process is genotype imputation [66]. Imputing against a reference genotype dataset, such as HapMap or 1000Genomes, increases the number of SNPs that can be tested for association.

Meta-analysing genetic data

Meta-analysis combines the results of different studies investigating the same research question in order to identify patterns among the study results. The aim of the meta-analysis in genetic epidemiology is to evaluate the diversity, i.e. so-called heterogeneity, among the results of the different studies and estimating a pooled effect of the trait-associated SNP [67]. Moreover, by meta-analysing data a more robust indication of whether or not a certain genetic variant is really of importance in the disease under study might be achieved. Depending on the amount of heterogeneity, either a fixed or random effects meta-analysis is used. A fixed effects meta-analysis assumes that the genetic effects are similar across the combined studies and that all observed differences are due to chance [68]. Random effects calculations assume that due to true differences because of various reasons, the estimates of the genetic effects may vary across different studies [67]. The between-studies heterogeneity can be measured by the metric I^2, which is the percentage of total variation across studies due to heterogeneity beyond chance [67,69]. The I^2 ranges from 0 to 100%. An I^2 exceeding 25% is considered to indicate moderate to large heterogeneity [67]. Therefore, when an $I^2 > 25\%$ is present, a random effects meta-analysis is used in this thesis to combine data from different studies.

PCOS as a complex genetic disorder

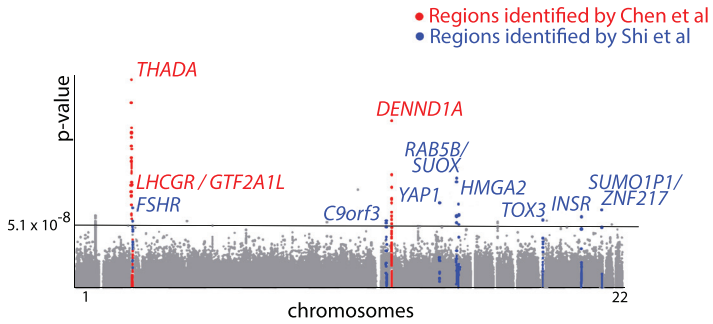
Candidate gene studies

Hitherto, candidate gene studies have been the predominant approach used in PCOS genetics (for review see [9,70-73]). In these studies over a hundred candidate genes have been studied, mainly focussing on candidate genes selected from logical pathways involved in or regulating gonadotropin secretion, steroid synthesis and insulin signalling [70]. Varying diagnostic criteria, small sample sizes in many studies, and more importantly, the lack of positive replication studies along with the broad phenotype of PCOS have hindered progress in this field [9,72]. Although, in general, this candidate gene approach has not been tremendously successful, some very promising candidate genes have been identified. The most thoroughly investigated functional candidates for PCOS susceptibility are the fibrillin (*FBN3*) gene, which is involved in TGF-β signalling; genes involved in insulin signalling and type 2 diabetes mellitus; and obesity genes [71].

Genome-wide association studies

Obviously, candidate gene selection is also limited by our incomplete knowledge of the pathophysiology of PCOS [70]. Three GWASs of PCOS have been published: two in Han-Chinese patients and one in Korean patients with PCOS (74-76). The results of these GWASs are presented in Figure 4 and Table 2.

Figure 4 | Manhattan plot of the results of the association analysis of the GWASs in the Han-Chinese PCOS patients.



Results of the studies by Chen et al. [74] and Shi et al. [76]. On the Y-axis the logarithm of the *P* value is plotted. The chromosomes are plotted on the X-axis. Every dot represents a SNP.

The *LHCGR* and *FSHR* gene are plausible PCOS candidate genes due to their role in follicular development and ovarian physiology. SNPs mapping to the *FSHR* gene has been previously reported to be associated with PCOS [77]. In addition, the GWASs also provided new directions to further unravel underlying biological mechanisms of PCOS. *GTF2A1L* is highly expressed in adult testis and may play a role in spermatogenesis [78]. The *THADA* (thyroid adenoma associated) region has previously been associated with type 2 diabetes mellitus in Europeans [79]. Genetic variants mapping to the *DENND1A* (DENN domain-containing protein 1A) locus might influence the pathogenesis of PCOS through misregulation of endoplasmic reticulum aminopeptidase 1 (ERAP1) [74]. The other identified PCOS association signals were mapping to genes related to insulin signalling, sexual hormone function, type 2 diabetes calcium signalling and endocytosis. The GWAS in Korean patients with PCOS did not reveal genome-wide significant hits [75].

Table 2 | Results of GWASs in PCOS.

	Discovery set (n)		Replication set (n)		chr	Nearby genes	Ethnic origin
	Cases	Controls	Cases	Controls			
Chen et al.[1]	744	895	3338	5792	2	*GTF2A1L*	Han-Chinese
						LHCGR	
						THADA	
					9	*DENND1A*	
Shi et al.[2]	2254	2911	8226	7578	2	*FSHR*	Han-Chinese
					9	*C9orf3*	
					11	*YAP1*	
					12	*RAB5B,*	
						SUOX	
						HMGA2	
					16	*TOX3*	
					19	*INSR*	
					20	*SUMO1P1*	
						ZNF217	
Hwang et al.[3]	1741	967	-	-	no SNP with p-value < 5×10^{-8}		Korean

GWAS, Genome Wide Association Study; *n,* number; *Chr,* chromosome; *SNP,* single Nucleotide Polymorphism.

[1] Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nature genetics*. 2011; Jan; 43: 55-9.

[2] Shi Y, Zhao H, Cao Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nature genetics*. 2012; Sep; 44: 1020-5.

[3] Hwang JY, Lee EJ, Jin Go M, et al. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet*. 2012; Oct; 57: 660-4.

Large phenotypic differences are present among patients with PCOS from various ethnicities [80,81]. Therefore, genetic variants might act differently in PCOS patients of various ethnic groups. The association of SNPs mapping to the *THADA*, *DENND1A* and *FSHR* susceptibility loci in Chinese patients with PCOS have been replicated in patients from Northern European descent (82-84). However, more studies are needed to further determine the role of genetic factors in explaining ethnic differences in PCOS [8].

Aims of the thesis

The variety of phenotypic characteristics accompanying PCOS might be partly reflected by the influence of genetic factors, environmental factors and ethnic origin. Subsequently, not all patients with PCOS encounter similar health risks.

The aims of this thesis were:

1. to determine which patients, based on specific characteristics within the broad PCOS phenotype, have the highest risk of developing long-term health complications,
2. to identify genetic variants associated with PCOS based on candidate-gene analysis as well as on hypothesis-free approach using GWAS

The first part in this thesis focusses on identification of patients with a high risk of developing adverse long-term health implications. The second part focusses on identification of additional PCOS-loci using the candidate gene and genome-wide association approach to obtain novel biological insights into the pathogenesis of PCOS and the mechanisms of its phenotypic heterogeneity. Moreover, the studies in this thesis address the cross-ethnic effect of genetic variants and the influence of genetic ancestry on phenotypic characteristics of PCOS.

Outline of the thesis

Chapter two defines the risk of cardiometabolic complications in patients with PCOS based on the presence of hyperandrogenism. **Chapter three** determines the mortality risk of parents of patients with PCOS, especially of parents with type 2 diabetes mellitus. **Chapter four** includes the results of three candidate-gene studies for PCOS. **Chapter five** describes the preliminary results of a genome-wide association study in patients with PCOS of European descent.

Chapter six assesses the effects of genetic variants identified in Chinese patients with PCOS in patients from Northern European descent. **Chapter seven** compares the self-reported ethnicity of patients with PCOS to their genetic ancestry based on genome-wide genetic data. Moreover, it determines which of these two is the best predictor for phenotypic characteristics of the syndrome.

Chapter eight summarizes the results of this thesis, places them in broader context and discusses the implications for clinical practice as well as the directions for future research.

References

1. Asuncion M, Calvo RM, San Millan JL, et al. A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. *J Clin Endocrinol Metab*. 2000; Jul; 85: 2434-8.
2. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
3. Knochenhauer ES, Key TJ, Kahsar-Miller M, et al. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab*. 1998; Sep; 83: 3078-82.
4. ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004; Jan; 81: 19-25.
5. ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Hum Reprod*. 2004; Jan; 19: 41-7.
6. Barber TM, McCarthy MI, Wass JA, et al. Obesity and polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2006; Aug; 65: 137-45.
7. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr Rev*. 1997; Dec; 18: 774-800.
8. Fauser BC, Tarlatzis BC, Rebar RW, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril*. 2012; Jan; 97: 28-38 e25.
9. Goodarzi MO, Dumesic DA, Chazenbalk G, et al. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol*. 2011; Apr; 7: 219-31.
10. Ehrmann DA. Polycystic ovary syndrome. *N Engl J Med*. 2005; Mar 24; 352: 1223-36.
11. Stein IF, Leventhal ML. Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol*. 1935; 29: 181-91.
12. Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. In: Dunaif A, Givens JR and Haseltine (eds), Polycystic Ovary Syndrome. Boston: Blackwell Scientific; 1992. p. 377-84.
13. Azziz R, Carmina E, Dewailly D, et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4237-45.
14. National Institutes of Health. Final Statement Evidence-based Methodology Workshop on Polycystic Ovary Syndrome. 2012.
15. Azziz R, Carmina E, Dewailly D, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril*. 2009; Feb; 91: 456-88.
16. Consensus on infertility treatment related to polycystic ovary syndrome. *Hum Reprod*. 2008; Mar; 23: 462-77.
17. Consensus on infertility treatment related to polycystic ovary syndrome. *Fertil Steril*. 2008; Mar; 89: 505-22.
18. Boomsma CM, Fauser BC, Macklon NS. Pregnancy complications in women with polycystic ovary syndrome. *Semin Reprod Med*. 2008; Jan; 26: 72-84.
19. Broekmans FJ, Knauff EA, Valkenburg O, et al. PCOS according to the Rotterdam consensus criteria: Change in prevalence among WHO-II anovulation and association with metabolic factors. *BJOG*. 2006; Oct; 113: 1210-7.
20. Huang A, Brennan K, Azziz R. Prevalence of hyperandrogenemia in the polycystic ovary syndrome diagnosed by the National Institutes of Health 1990 criteria. *Fertil Steril*. 2010; Apr; 93: 1938-41.

21. Escobar-Morreale HF, Carmina E, Dewailly D, et al. Epidemiology, diagnosis and management of hirsutism: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update*. 2012; Mar-Apr; 18: 146-70.
22. DeUgarte CM, Bartolucci AA, Azziz R. Prevalence of insulin resistance in the polycystic ovary syndrome using the homeostasis model assessment. *Fertil Steril*. 2005; May; 83: 1454-60.
23. Legro RS, Kunselman AR, Dodson WC, et al. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999; Jan; 84: 165-9.
24. Valkenburg O, Steegers-Theunissen RP, Smedts HP, et al. A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. *J Clin Endocrinol Metab*. 2008; Feb; 93: 470-6.
25. Carmina E, Orio F, Palomba S, et al. Endothelial dysfunction in PCOS: role of obesity and adipose hormones. *Am J Med*. 2006; Apr; 119: 356 e1-6.
26. Talbott EO, Zborowski JV, Rager JR, et al. Evidence for an association between metabolic cardiovascular syndrome and coronary and aortic calcification among women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2004; Nov; 89: 5454-61.
27. Talbott EO, Zborowski JV, Sutton-Tyrrell K, et al. Cardiovascular risk in women with polycystic ovary syndrome. *Obstet Gynecol Clin North Am*. 2001; Mar; 28: 111-33, vii.
28. Welt CK, Gudmundsson JA, Arason G, et al. Characterizing discrete subsets of polycystic ovary syndrome as defined by the Rotterdam criteria: the impact of weight on phenotype and metabolic features. *J Clin Endocrinol Metab*. 2006; Dec; 91: 4842-8.
29. Moran LJ, Hutchison SK, Norman RJ, et al. Lifestyle changes in women with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2011: CD007506.
30. Moran LJ, Misso ML, Wild RA, et al. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update*. 2010; Jul-Aug; 16: 347-63.
31. Fritz AF, Speroff L. Clinical Gynecologic Endocrinology and Infertility2010.
32. Strauss JF, Barbieri RL. Reproductive Endocrinology.
33. Durlinger AL, Visser JA, Themmen AP. Regulation of ovarian function: the role of anti-Mullerian hormone. *Reproduction*. 2002; Nov; 124: 601-9.
34. Pellatt L, Rice S, Mason HD. Anti-Mullerian hormone and polycystic ovary syndrome: a mountain too high? *Reproduction*. 2010; May; 139: 825-33.
35. Labrie F. Intracrinology. *Mol Cell Endocrinol*. 1991; Jul; 78: C113-8.
36. Taylor AE, McCourt B, Martin KA, et al. Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 1997; Jul; 82: 2248-56.
37. Waldstreicher J, Santoro NF, Hall JE, et al. Hyperfunction of the hypothalamic-pituitary axis in women with polycystic ovarian disease: indirect evidence for partial gonadotroph desensitization. *J Clin Endocrinol Metab*. 1988; Jan; 66: 165-72.
38. Laven JS, Imani B, Eijkemans MJ, et al. New approach to polycystic ovary syndrome and other forms of anovulatory infertility. *Obstet Gynecol Surv*. 2002; Nov; 57: 755-67.
39. Nelson VL, Legro RS, Strauss JF, 3rd, et al. Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. *Mol Endocrinol*. 1999; Jun; 13: 946-57.
40. Kumar A, Woods KS, Bartolucci AA, et al. Prevalence of adrenal androgen excess in patients with the polycystic ovary syndrome (PCOS). *Clin Endocrinol (Oxf)*. 2005; Jun; 62: 644-9.
41. Franks S, Stark J, Hardy K. Follicle dynamics and anovulation in polycystic ovary syndrome. *Hum Reprod Update*. 2008; Jul-Aug; 14: 367-78.
42. Hillier SG. Current concepts of the roles of follicle stimulating hormone and luteinizing hormone in folliculogenesis. *Hum Reprod*. 1994; Feb; 9: 188-91.
43. Pigny P, Merlen E, Robert Y, et al. Elevated serum level of anti-mullerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. *J Clin Endocrinol Metab*. 2003; Dec; 88: 5957-62.

44. Laven JS, Mulders AG, Visser JA, et al. Anti-Mullerian hormone serum concentrations in normoovulatory and anovulatory women of reproductive age. *J Clin Endocrinol Metab*. 2004; Jan; 89: 318-23.
45. Barbieri RL, Makris A, Randall RW, et al. Insulin stimulates androgen accumulation in incubations of ovarian stroma obtained from women with hyperandrogenism. *J Clin Endocrinol Metab*. 1986; May; 62: 904-10.
46. Nestler JE, Jakubowicz DJ, de Vargas AF, et al. Insulin stimulates testosterone biosynthesis by human thecal cells from women with polycystic ovary syndrome by activating its own receptor and using inositolglycan mediators as the signal transduction system. *J Clin Endocrinol Metab*. 1998; Jun; 83: 2001-5.
47. Nestler JE. Insulin regulation of human ovarian androgens. *Hum Reprod*. 1997; Oct; 12 Suppl 1: 53-62.
48. Kahsar-Miller MD, Nixon C, Boots LR, et al. Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. *Fertil Steril*. 2001; Jan; 75: 53-8.
49. Lunde O, Magnus P, Sandvik L, et al. Familial clustering in the polycystic ovarian syndrome. *Gynecol Obstet Invest*. 1989; 28: 23-30.
50. Colilla S, Cox NJ, Ehrmann DA. Heritability of insulin secretion and insulin action in women with polycystic ovary syndrome and their first degree relatives. *J Clin Endocrinol Metab*. 2001; May; 86: 2027-31.
51. Legro RS, Bentley-Lewis R, Driscoll D, et al. Insulin resistance in the sisters of women with polycystic ovary syndrome: association with hyperandrogenemia rather than menstrual irregularity. *J Clin Endocrinol Metab*. 2002; May; 87: 2128-33.
52. Legro RS, Driscoll D, Strauss JF, 3rd, et al. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *Proc Natl Acad Sci U S A*. 1998; Dec 8; 95: 14956-60.
53. Yildiz BO, Goodarzi MO, Guo X, et al. Heritability of dehydroepiandrosterone sulfate in women with polycystic ovary syndrome and their sisters. *Fertil Steril*. 2006; Dec; 86: 1688-93.
54. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2100-4.
55. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953; Apr 25; 171: 737-8.
56. Strachan T, Read AP. Human Molecular Genetics 3. Oxford: Garland Science Publishers; 2003.
57. Carlson CS, Eberle MA, Kruglyak L, et al. Mapping complex disease loci in whole-genome association studies. *Nature*. 2004; May 27; 429: 446-52.
58. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; Feb 15; 409: 860-921.
59. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001; Feb 16; 291: 1304-51.
60. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet*. 2005; Sep 10-16; 366: 941-51.
61. The International HapMap Project. *Nature*. 2003; Dec 18; 426: 789-96.
62. Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; Nov 1; 491: 56-65.
63. Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. *N Engl J Med*. 2009; Apr 23; 360: 1699-701.
64. Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet*. 2005; Oct 1; 366: 1223-34.
65. Cordell HJ, Clayton DG. Genetic association studies. *Lancet*. 2005; Sep 24-30; 366: 1121-31.
66. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010; Jul; 11: 499-511.
67. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One*. 2007; 2: e841.

68. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med*. 1997; Nov 1; 127: 820-6.

69. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; Jun 15; 21: 1539-58.

70. Goodarzi MO. Looking for polycystic ovary syndrome genes: rational and best strategy. *Semin Reprod Med*. 2008; Jan; 26: 5-13.

71. Kosova G, Urbanek M. Genetics of the polycystic ovary syndrome. *Mol Cell Endocrinol*. 2013; Jul 5; 373: 29-38.

72. Urbanek M. The genetics of the polycystic ovary syndrome. *Nat Clin Pract Endocrinol Metab*. 2007; Feb; 3: 103-11.

73. Urbanek M, Legro RS, Driscoll D, et al. Searching for the polycystic ovary syndrome genes. *J Pediatr Endocrinol Metab*. 2000; 13 Suppl 5: 1311-3.

74. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011; Jan; 43: 55-9.

75. Hwang JY, Lee EJ, Jin Go M, et al. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet*. 2012; Oct; 57: 660-4.

76. Shi Y, Zhao H, Cao Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012; Sep; 44: 1020-5.

77. Laven JS, Mulders AG, Suryandari DA, et al. Follicle-stimulating hormone receptor polymorphisms in women with normogonadotropic anovulatory infertility. *Fertil Steril*. 2003; Oct; 80: 986-92.

78. Huang M, Wang H, Li J, et al. Involvement of ALF in human spermatogenesis and male infertility. *Int J Mol Med*. 2006; Apr; 17: 599-604.

79. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008; May; 40: 638-45.

80. Valkenburg O, Lao O, Schipper I, et al. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *J Clin Endocrinol Metab*. 2011; Jul; 96: E1181-7.

81. Welt CK, Arason G, Gudmundsson JA, et al. Defining constant versus variable phenotypic features of women with polycystic ovary syndrome using different ethnic groups and populations. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4361-8.

82. Goodarzi MO, Jones MR, Li X, et al. Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *J Med Genet*. 2012; Feb; 49: 90-5.

83. Mutharasan P, Galdones E, Penalver Bernabe B, et al. Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. *J Clin Endocrinol Metab*. 2013; Jan; 98: E185-90.

84. Welt CK, Styrkarsdottir U, Ehrmann DA, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. *J Clin Endocrinol Metab*. 2012; Jul; 97: E1342-7.

Chapter 2

The Cardiometabolic Profile of Women with Hyperandrogenic PCOS Compared to Women with Non-hyperandrogenic PCOS

N.M.P. Daan
Y.V. Louwers
M.P.H. Koster
M.J.C. Eijkemans
Y.B. de Rijke
E.W.G. Lentjes
B.C.J.M. Fauser
J.S.E. Laven

Manuscript submitted

Abstract

Context: Different sub-phenotypes within Polycystic Ovary Syndrome (PCOS) are associated with differences in cardiometabolic presentation.

Objective: To compare the known markers for cardiometabolic disease and cardiovascular risk factor (CVRF) distribution of patients with hyperandrogenic PCOS, with patients having ovulatory dysfunction and polycystic ovarian morphology (PCOM).

Setting: Specialized reproductive outpatient clinics of the Erasmus Medical Centre Rotterdam and the University Medical Centre Utrecht, the Netherlands.

Patients or Other Participants: 2288 distinctly phenotyped reproductive-aged patients with PCOS were included.

Intervention(s): Women with a suspicion of oligo- or anovulation underwent a standardized screening consisting of a systematic general medical, reproductive and family history taking, anthropometric measurements, transvaginal ultrasonography, extensive endocrine and metabolic laboratory assessment.

Main Outcome Measure(s): Differences in cardiometabolic profile and CVRF prevalence between different PCOS sub-phenotypes, i.e. obesity and overweight, hypertension, hyperglycemia, insulin resistance, dyslipidemia and the metabolic syndrome (MetS).

Results: Patients with hyperandrogenic PCOS presented with a poor cardiometabolic profile upon screening, resulting in a higher prevalence of CVRFs such as obesity and overweight (P <0.001), insulin resistance (P <0.001) and MetS (P <0.001), compared with patients with non-hyperandrogenic PCOS. No significant differences in cardiometabolic risk factors, except for overweight or obesity, were observed amongst the different hyperandrogenic PCOS characteristics.

Conclusions: Patients with hyperandrogenic PCOS have an unfavourable cardiometabolic profile and higher prevalence of cardiovascular risk factors compared to patients with ovulatory dysfunction and PCOM. Whether these risk factors result in cardiovascular disease and actual cardiovascular events should be determined in prospective follow-up studies.

Introduction

The Polycystic Ovary Syndrome (PCOS) is the most common endocrinopathy in women of reproductive age. The prevalence of PCOS has been estimated to be up to 12% depending on which diagnostic criteria are used [1]. Ever since Stein and Leventhal first described a series of women with oligmenorrhea or amenorrhea and polycystic ovaries in 1935 [2], there has been an ongoing international debate about the legitimacy of various criteria which are used to diagnose PCOS. In the absence of an universally accepted definition the 1990 National Institute of Health (NIH) criteria were introduced, delineating PCOS as a syndrome of both chronic anovulation and clinical/biochemical signs of hyperandrogenism (with exclusion of other etiologies) [3].

During the years thereafter the broad clinical expression of PCOS gave rise to the perception that no single criterion was mandatory in diagnosing PCOS resulting in the 2003 Rotterdam consensus criteria [4,5]. According to these criteria, PCOS is diagnosed when at least two of the following criteria are present: ovulatory dysfunction (oligo- or amenorrhea), hyperandrogenism (either biochemical or clinical being hirsutism) and/or polycystic ovarian morphology (PCOM). With the introduction of these broadened criteria, the prevalence of PCOS amongst women with normogonadotropic normo-oestrogenic anovulation as well as amongst women in the general population, increased considerably [6].

Over the years evidence has accumulated that the presence of hyperandrogenism is associated with an increased prevalence of metabolic disturbances in patients with PCOS [7-11]. Therefore, in response to the introduction of the Rotterdam criteria, it has been postulated that PCOS should indeed be primarily considered a disorder of hyperandrogenemia [12]. It may even be justified to distinguish the hyperandrogenic PCOS phenotype as a separate entity given its profound metabolic implications, and assign this phenotype with a new name since the current denomination is unsatisfying for both clinicians and patients [13].

It remains to be elucidated to what extent women with ovarian dysfunction and polycystic ovarian morphology, face the same long term health implications as "classic" hyperandrogenic patients with PCOS. At the NIH Evidence-based Methodology Workshop on PCOS in December 2012 an independent expert panel recommended maintaining the broad diagnostic Rotterdam criteria, along with the usage of distinct phenotyping in PCOS [14]. Furthermore, the panel underlined several future research priorities, one of which is the determination of the prevalence of cardiometabolic implications of the individual PCOS phenotypes, as was also stressed at the previous ESHRE-ASRM endorsed, third consensus meeting [15].

The aim of the current large cohort study was to compare the known markers for cardiometabolic disease as well as the cardiovascular risk factor (CVRF) distribution, amongst women with the hyperandrogenic PCOS sub-phenotypes and those with the ovulatory dysfunction and polycystic ovarian morphology (PCOM) phenotype.

Materials and Methods

Study population

Women diagnosed with PCOS according to the Rotterdam criteria, who were screened at the outpatient clinics of either the Erasmus Medical Centre Rotterdam or the University Medical Centre Utrecht between January 1st 2004 and May 1st 2013, were eligible for inclusion in this study. This screening was performed by trained professionals according to a standardized protocol that has been previously described in great detail [6,9]. The screening procedure started with a thorough general medical, reproductive and family history taking, which was followed by anthropometric measurements including height, weight, body mass index (BMI), systolic and diastolic blood pressure (BP), waist circumference (WC), hip circumference and the measurement of hirsutism using the Ferriman-Gallwey (FG) score. Furthermore, a systematic transvaginal ultrasonography was performed assessing double endometrial thickness, ovarian volume and the total number of antral follicles, measuring 2-10 mm. Finally, an extensive fasting endocrine and metabolic profile was assessed.

PCOS was diagnosed according to the Rotterdam criteria; i.e. requiring the presence of at least two out of the following three criteria: ovulatory dysfunction, androgen excess and/or polycystic ovarian morphology [4,5]. Ovulatory dysfunction was defined as oligomenorrhea, i.e. mean bleeding interval between 35 days – 182 days, or amenorrhea, i.e. absence of menstrual bleeding for more than 182 days. We defined androgen excess as a FG score ≥ 9, and/or a free androgen index > 4.5 [FAI: (Testosterone / SHBG) x 100)] [16]. An ovary was designated as polycystic if the volume exceeded 10 cm^3 and/or the follicle count (2-10 mm) was ≥ 12 [17].

Normogonadotropic normoestrogenic patients who underwent the aforementioned standardized screening and were diagnosed with PCOS, were classified into one of the four potential PCOS phenotypes:

1. Androgen Excess + Ovulatory Dysfunction (AE + OD)
2. Androgen Excess + Polycystic Ovarian Morphology (AE + PCOM)
3. Ovulatory Dysfunction + Polycystic Ovarian Morphology (OD + PCOM)
4. Androgen Excess + Ovulatory Dysfunction + Polycystic Ovarian Morphology (AE + OD + PCOM).

Patients had to be aged between 18-45 years upon the moment of screening. Patients were excluded if they were not fasting at the time of the blood withdrawal.

Endocrine and metabolic assessment

With respect to cardiometabolic presentation we collected information on BMI, WC, systolic and diastolic BP. Furthermore, we selected the following endocrine and metabolic laboratory parameters for comparative analysis between PCOS sub-phenotype groups: FSH, LH, estradiol, testosterone, dehydroepiandrosterone sulphate (DHEAS), SHBG, insulin, glucose, total

cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C). Low-density lipoprotein cholesterol (LDL-C) was calculated using the Friedewald formula [LDL-C = TC – HDL-C – (TG/5)] [18]. Insulin resistance was assessed using the homeostasis model assessment (HOMA-IR): fasting glucose(mmol/L) x fasting insulin(mU/L) / 22.5 [19]. During the study period both centres used various assays to measure these parameters. Details about the used assays and their mutual comparability are available in Supplementary Table 1. Uniformity between different assays was assured by using annual calibration data from the Dutch Foundation for quality assessment in clinical laboratories (SKML) , which allowed combined interpretation of laboratory results from both centres. All data were recalculated accordingly.

Assessment of cardiometabolic profile

We compared the presence of the following CVRFs between different PCOS sub-phenotype groups: obesity (BMI ≥ 30 kg/m^2), overweight (BMI 25-29.9 kg/m^2), hypertension (BP ≥ 140/90 mmHg), increased WC (WC ≥ 88 cm), hyperglycemia (fasting glucose > 6.0 mmol/L), insulin resistance (1/HOMA-IR < 0.47), dyslipidemia (TC ≥ 5.0 mmol/L, TG > 1.7 mmol/L, LDL-C ≥ 3.0 mmol/L, HDL-C < 1.2 mmol/L) and presence of the metabolic syndrome (MetS) (≥ 3 of any of these clinical features; WC ≥ 88 cm, fasting glucose ≥ 5.6 mmol/L, BP ≥ 130/85 mmHg, HDL < 1.3 mmol/L, TG ≥ 1.7 mmol/L) [20-23].

Statistical analyses

Lipid profile assessment was only performed in a subset of the patients screened at the Erasmus Medical Centre. To avoid any potential bias that may occur in complete-case analysis, multiple imputations (10×) were applied using observed patient characteristics [24]. Missing data were imputed using a logistic regression model. Baseline variables were expressed as means with standard deviation or numbers with percentage. Multiple imputation and other statistical analyses were performed using SPSS statistics version 20.0.

For the first analyses, we compared the hyperandrogenic patients with PCOS (phenotypes: AE + OD, AE + PCOM and AE + OD + PCOM) to the non-hyperandrogenic patients with PCOS (phenotype OD + PCOM) for potential variations in ethnic distribution and differences in CVRF prevalence. Univariate logistic regression was used to compare the endocrine and cardiometabolic characteristics of hyperandrogenic to those of non-hyperandrogenic patients with PCOS. Subsequently, multivariate logistic regression analyses were performed to correct for potential confounders including age, BMI, smoking and ethnicity. A *P* value of < 0.05 was considered to be statistically significant. Moreover, in a subgroup analysis we compared the three individual hyperandrogenic sub-phenotype groups to detect differences in CVRF prevalence. In pairwise comparisons between any two hyperandrogenic sub-phenotype groups, a Bonferroni corrected *P* value for multiple testing < 0.0167 (0.05/3) was considered statistically significant.

Ethical approval

This study was reviewed by the local Institutional Review Board (IRB) Committee of both the Erasmus Medical Centre Rotterdam and the University Medical Centre Utrecht. Since the study neither implied that patients would receive a particular treatment, nor imposed on their behaviour as described in the Medical Research Involving Human Subjects Act (WMO), the IRB officially stated that the WMO did not apply.

Results

In total, 2510 patients diagnosed with PCOS were eligible for inclusion in this study. Of these, 222 (8.9%) patients could not be assigned to a PCOS sub-phenotype group because of missing data: 13 women were excluded because of uncertainty about their (current) menstrual cycle pattern, 99 women had incomplete or inconclusive ultrasonography data, of one patient data on androgen levels was missing, 71 women were excluded since they were not fasting at the time of blood withdrawal and 38 women were excluded because they were under 18 years of age at time of screening. Hence, 2288 patients with PCOS were included in our final analyses, of which 1437 patients were diagnosed with PCOS at the Erasmus Medical Centre Rotterdam and 851 patients at the University Medical Centre Utrecht.

In our study population 1219 (53.3%) women exhibited a hyperandrogenic form of PCOS (AE +OD, AE + PCOM, AE + OD + PCOM). The 1069 (46.7%) remaining patients with PCOS, solely had signs of ovulatory dysfunction and polycystic ovarian morphology (OD + PCOM) as is depicted in Figure 1.

The majority (69.7%) of included patients were from Northern European descent which is shown in Table 1. The distribution of the various ethnicities differed significantly between patients with hyperandrogenic PCOS and patients with non-hyperandrogenic PCOS. The proportion of women from Northern European descent was significantly lower in hyperandrogenic patients with PCOS than in non-hyperandrogenic patients with PCOS (58.0% vs. 82.8% respectively; P <0.001). Furthermore each one of the other ethnicities was more prevalent amongst patients with hyperandrogenic PCOS compared to patients with non-hyperandrogenic PCOS.

Patients with hyperandrogenic PCOS differed significantly from patients with non-hyperandrogenic PCOS as far as the majority of endocrine and cardiometabolic parameters were concerned (Table 1). Patients with hyperandrogenic PCOS had a more unfavourable cardiometabolic profile, including higher BMI, WC, systolic and diastolic BP, and fasting insulin, LDL-C, and lower HDL-C levels. Moreover, a higher prevalence of CVRFs in patients with hyperandrogenic PCOS, compared to patients with non-hyperandrogenic PCOS was observed (Table 2). In patients with hyperandrogenic PCOS the most prevalent CVRFs were overweight or obesity (67.9%), raised LDL-C (61.8%) and elevated waist circumference (53.5%). This CVRF distribution was dissimilar to the distribution seen in non-hyperandrogenic PCOS

patients. In the latter raised LDL-C was the most prevalent risk factor (52.2%) followed by raised Total-C (47.4%) and overweight or obesity (28.5%). When the prevalence of individual CVRFs were compared between the hyperandrogenic and non-hyperandrogenic PCOS group, we observed that hyperandrogenic patients were significantly more often overweight/obese, hyperglycemic, insulin resistant and hypertensive. Furthermore, lipid profiles of these patients were significantly more disturbed, and the MetS was more prevalent in the hyperandrogenic PCOS group (25.8% vs 6.5%, $P < 0.001$) compared to the non-hyperandrogenic PCOS sub-phenotype.

Multiple logistic regression analysis, correcting for age, BMI, smoking and ethnicity, revealed that the differences in the incidence of CVRF between hyperandrogenic and non-hyperandrogenic patients attenuated, but did not disappear. However, the difference in the incidence of hypertension between patients with hyperandrogenic and patients with non-hyperandrogenic PCOS was no longer significant. Also, differences in lipid profile were less pronounced after this correction, except for HDL-C levels ($P = 0.006$). Differences in prevalence of MetS and insulin resistance remained significant ($P = 0.025$ and $P < 0.001$, respectively).

When comparing the individual hyperandrogenic sub-phenotypes (AE+OD, AE + PCOM, AE+OD+PCOM) for CVRF incidence (Table 3), we observed no significant differences between these sub-phenotypes after correction for multiple testing ($P < 0.0167$), except for overweight/obesity between the OD + AE phenotype and the OD + AE + PCOM phenotype ($P = 0.010$).

Figure 1 | Distribution of PCOS phenotypes.

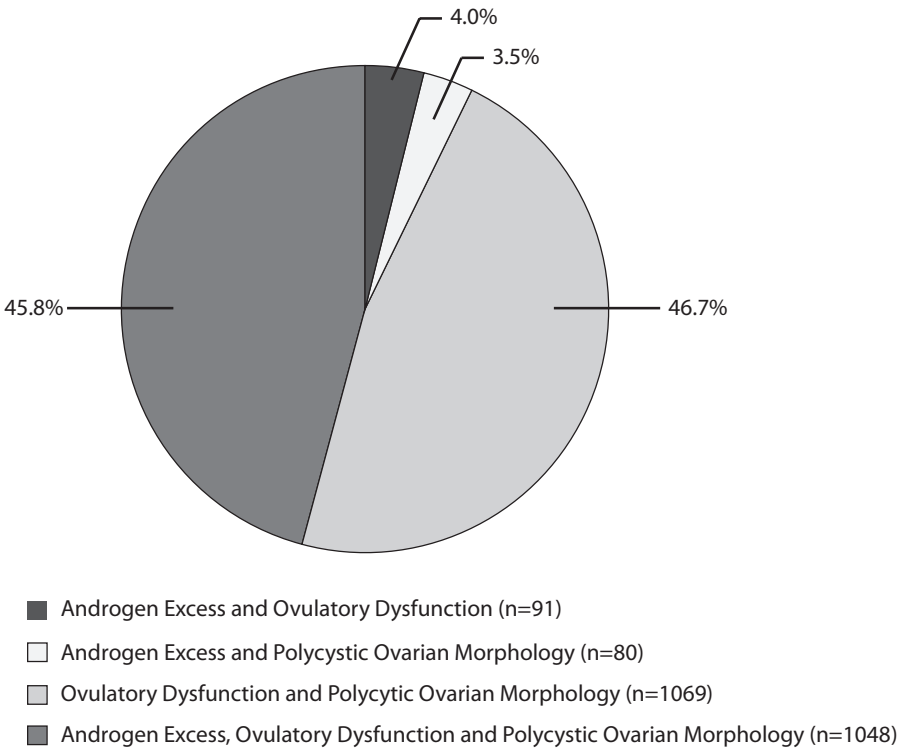


Table 1 | Comparison of the baseline and endocrine characteristics of the hyperandrogenic and the non-hyperandrogenic patients with PCOS.

Parameters	**HA – PCOS (n=1219)**	**non HA – PCOS (n=1069)**	***P values***	***Adjusted P values***
Age (years)	27.5 (5.1)	29.1 (4.5)	< 0.001	-
Ethnicity			< 0.001	-
Northern European	707 (58.0)	886 (82.8)		
Turkish / Moroccan	123 (10.1)	69 (6.5)		
Negroid	206 (16.9)	41 (3.8)		
Hindustani	80 (6.6)	22 (2.1)		
Asian	57 (4.7)	29 (2.7)		
Other	46 (3.7)	22 (2.1)		
BMI (kg/m^2)	28.8 (6.5)	23.6 (4.7)	< 0.001	< 0.001*
WC (cm)	90.5 (16.3)	77.6 (14.0)	< 0.001	< 0.001
BP systolic	120.6 (14.1)	117.2 (12.7)	< 0.001	0.053
BP diastolic	78.3 (10.4)	74.7 (9.7)	< 0.001	0.010
LH (U/L)	11.4 (9.7)	9.0 (8.6)	< 0.001	< 0.001
FSH (U/L)	6.2 (2.2)	6.2 (2.6)	0.604	< 0.001
Estradiol (pmol/L)	223.6 (198.0)	226.1 (205.3)	0.768	0.414
Testosterone (nmol/L)	2.2 (0.9)	1.5 (0.6)	< 0.001	< 0.001
DHEAS (µmol/L)	5.8 (9.7)	4.3 (1.8)	< 0.001	< 0.001
Androstenedione (nmol/L)	6.9 (2.7)	4.9 (1.8)	< 0.001	< 0.001
SHBG (nmol/L)	32.2 (17.2)	62.9 (29.4)	< 0.001	< 0.001
FAI	8.6 (6.5)	2.7 (1.1)	< 0.001	< 0.001
Glucose (mmol/L)	4.9 (0.8)	4.9 (2.0)	0.174	0.117
Insulin (mIU/L)	11.6 (10.6)	6.0 (4.9)	< 0.001	< 0.001
TG (mmol/L)	1.2 (0.9)	1.0 (4.1)	0.171	0.954
Total-C (nmol/L)	5.2 (1.5)	5.1 (1.6)	0.236	0.719
LDL-C (nmol/L)	3.4 (1.3)	3.2 (1.3)	< 0.001	0.356
HDL-C (nmol/L)	1.4 (0.5)	1.7 (0.5)	< 0.001	0.002

Values are means (standard deviations) or numbers (percentages). Adjusted *P* values are adjusted for age, smoking, BMI and ethnicity. *Corrected for age, smoking and ethnicity only. *HA,* Hyperandrogenic; *BMI,* body mass index; *BP,* blood pressure; *WC,* waist circumference, *FAI,* free androgen index; *TG,* triglycerides; *Total-C,* total cholesterol; *LDL-C,* LDL-cholesterol; *HDL-C*, HDL-cholesterol.

Table 2 | Prevalence of cardiovascular risk factors in hyperandrogenic and non-hyperandrogenic patients with PCOS.

Parameters	HA – PCOS (n=1219)	non HA – PCOS (n=1069)	*unadjusted P value*	*adjusted P value*
Obesity (BMI ≥ 30 kg/m^2)	498 (40.9)	113 (10.6)	< 0.001	< 0.001*
Overweight or Obesity (BMI ≥ 25 kg/m^2)	828 (67.9)	305 (28.5)	< 0.001	< 0.001*
Hyperglycemia (fasting glucose > 6.0 mmol/L)	49 (4.0)	22 (2.1)	0.010	0.831
1/HOMA – IR < 0.47	528 (43.3)	150 (14.0)	< 0.001	< 0.001
BP ≥ 140 systolic and/or 90 diastolic mmHg	214 (17.6)	126 (11.8)	< 0.001	0.912
WC ≥ 88 cm	652 (53.5)	188 (18.8)	< 0.001	0.009
TG > 1.7 mmol/L	191 (15.6)	78 (7.3)	< 0.001	0.485
Total-C ≥ 5.0 mmol/L	588 (48.2)	507 (47.4)	0.696	0.957
LDL-C ≥ 3.0 mmol/L	753 (61.8)	558 (52.2)	< 0.001	0.233
HDL-C <1.2 mmol/L	498 (40.9)	193 (18.1)	< 0.001	0.006
MetS	314 (25.8)	70 (6.5)	< 0.001	0.025

Values are numbers of patients (percentages). Adjusted *P* values; adjusted for smoking, ethnicity, age and BMI. *Corrected for smoking and ethnicity only. *BMI,* body mass index; *BP,* blood pressure; *WC,* waist circumference; *TG,* triglycerides; *Total-C,* total cholesterol; *LDL-C,* LDL-cholesterol; *HDL-C,* HDL-cholesterol; *MetS,* metabolic syndrome.

Table 3 | Prevalence of cardiovascular risk factors in different hyperandrogenic PCOS phenotypes.

Parameters	Phenotype 1 AE + OD (n=91)	Phenotype 2 AE + PCOM (n=80)	Phenotype 3 AE+ OD + PCOM (n=1048)
Obesity (BMI ≥ 30 kg/m^2)	48 (52.7)	29 (36.3)	422 (40.3)
Overweight or Obesity (BMI ≥ 25 kg/m^2)	73 (80.2) [a]	57 (71.3)	698 (66.6) [a]
Hyperglycemia (fasting glucose > 6.0 mmol/L)	7 (7.7)	2 (2.5)	39 (3.7)
1/HOMA – IR < 0.47	46 (50.5)	32 (40.0)	450 (42.9)
BP ≥ 140 systolic and/or 90 diastolic mmHg	20 (22.0)	16 (20.0	177 (16.9)
WC ≥ 88 cm	59 (64.8)	37 (46.3)	557 (53.1)
TG > 1.7 mmol/L	18 (19.8)	11 (13.8)	162 (15.5)
Total-C ≥ 5.0 mmol/L	42 (46.2)	33 (41.3)	513 (49.0)
LDL-C ≥ 3.0 mmol/L	55 (60.4)	48 (60.0)	651 (62.1)
HDL-C < 1.2 mmol/L	49 (53.8)	33 (41.3)	417 (39.8)
MetS	32 (35.2)	22 (27.5)	260 (24.8)

Differences were considered significant when *P* < 0.0167 (Bonferroni correction). [a] Significant difference (*P* 0.010) between phenotype 1 and 3, all other differences were not significant. *AE,* androgen excess; *OD,* ovulatory dysfunction; *PCOM*, polycystic ovarian morphology; *BMI* body mass index; *BP,* blood pressure; *WC,* waist circumference; *TG* triglycerides, *Total-C,* total cholesterol; *LDL-C,* LDL-cholesterol; *HDL-C* HDL-cholesterol; *MetS,* metabolic syndrome.

Discussion

In this study we have observed that the hyperandrogenic sub-phenotype in patients with PCOS is clearly associated with an unfavourable cardiometabolic profile compared to patients who exhibit the non-hyperandrogenic PCOS sub-phenotype. In our study population, i.e. women with PCOS within their fertile years, about half of all patients were categorised into the hyperandrogenic PCOS sub-phenotype whereas the other half suffered from ovulatory dysfunction in combination with PCOM, i.e. the non-hyperandrogenic PCOS sub-phenotype. Hyperandrogenic patients with PCOS had a higher prevalence of nearly all studied CVRFs compared to the women with the non-hyperandrogenic PCOS sub-phenotype. This unfavourable profile remained even after correction for BMI, age, smoking and ethnic origin. Thereby, we have confirmed results from previously published studies including considerably smaller sample sizes, that the so-called hyperandrogenic PCOS phenotype encompasses the patients with the most unfavourable cardiometabolic phenotype [7,25]. Overall, the presence of CVRFs within the different sub-phenotypes of hyperandrogenic patients with PCOS were similar. However, patients with androgen excess and ovulatory dysfunction were more often overweight or obese compared to patients with hyperandrogenism, ovulatory dysfunction and PCOM.

This study provides further insight in the prevalence of MetS in different PCOS sub-phenotypes. Our results yielded a 25.8% prevalence of MetS in women with hyperandrogenic PCOS which is in line with previous reported prevalences of MetS (ATP III criteria) varying between 20-46% in women with PCOS defined according to the former (1992) NIH criteria [8]. The significantly lower prevalence of MetS (6.5%) in non-hyperandrogenic PCOS women within our population, seems to resemble the prevalence of MetS as it is seen in the normal (non – PCOS) population of women of similar age, BMI and ethnic background , although studies reporting on this matter are limited. This might be explained by the fact that the non-hyperandrogenic PCOS sub-phenotype consists of women who merely exhibit symptoms of ovulatory dysfunction and PCOM on ultrasound. Previous studies have indicated that the combination of current advanced ultrasound techniques and maintenance of a follicle threshold at 12 follicles to determine PCOM, results in a high prevalence of PCOM amongst asymptomatic women of reproductive age [26,27]. This has led some authors to conclude that PCOM itself is not associated with metabolic abnormalities [26]. At the same time others have advocated to raise the antral follicle count (AFC) threshold in the ultrasonographic diagnosis of PCOM, or even replace the ultrasound as a diagnostic tool with determination of AMH levels being a more precise marker in AFC assessment [27,28].

According to these views some of the young women with non-hyperandrogenic PCOS in our study population, should not have been diagnosed with PCOM and consequently therefore not would be assigned as having PCOS. The PCOM trait in ovulatory women does not necessarily seem to result in the development of PCOS later in life [29]. Nonetheless, PCOM in ovulatory women has been associated with significantly higher circulating serum testosterone

levels compared to ovulatory women without PCOM [30].This suggests that PCOM might be a predisposing factor for the evolvement of PCOS in some women, and therefore should not be entirely disregarded as such.

Furthermore, we have demonstrated that the distribution of ethnic origin between non-hyperandrogenic and hyperandrogenic PCOS women was different stressing the importance of taking ethnicity into consideration when addressing PCOS and its long-term health implications. Patients with PCOS of non-Northern European descent were more often attributed to the hyperandrogenic phenotype group, consistent with the higher prevalence of hyperandrogenism in patients of Middle Eastern and Mediterranean origin [31,32]. Moreover, it has been demonstrated previously in community based mixed population studies in the USA that there are not only variations in prevalence of PCOS amongst women of different ethnic descent, but also significant in differences in associated CVRF prevalence such as dyslipidemia, hypertension, insulin resistance and diabetes between various ethnic groups [33,34]. These differences observed in patients with PCOS might also reflect variation in androgen levels in the general population of different ethnic descent [35]. Therefore, the use of ethnic-specific appropriate thresholds for determining presence of hyperandrogenism as well as identifying adverse cardiometabolic implications has been proposed [15,36]. To overcome the potential bias associated with differences in ethnicity in the current study, we have adjusted the analyses for ethnic origin of the patients.

We are well aware that the lack of (inter)national assay standardization, as well as the limitations of current clinical methods used to assess hyperandrogenism, make diagnosing hyperandrogenism in PCOS and non-PCOS women an utterly precarious matter [37,38]. A potential limitation of the study is that due to the extensive period of screening we were faced with various changes in endocrine and metabolic laboratory assays over time. Fortunately we were able to apply well established conversion factors when necessary to correct for potential variations due to these assay alterations, also facilitating between-centre assay comparisons. By doing so we feel that we have corrected for potential variations due to these assay alterations. Moreover, it has been shown that RIA's are comparable to liquid chromatography tandem mass spectrometry (LC-MS/MS) in accuracy and precision of testosterone measurements in patients with PCOS [39]. At low testosterone levels, poor precision with all assays and significant variability between different LC-MS/MS assays remain problematic [40]. Hence, it is implausible that overestimation of testosterone levels has interfered with our results.

Although, the current study underlines disturbances in the cardiometabolic profile of hyperandrogenic patients with PCOS, it has to be noted that patients with the non-hyperandrogenic PCOS sub-phenotype also encountered derangements in their cardiometabolic profile at this relatively young age. According to the treatment recommendations of both the American Heart Association and European Society of Cardiology, intervention for cardiovascular risk factor reduction is required for some of these patients [23,41].

The high prevalence of CVRFs suggests that patients with PCOS, especially in patients with hyperandrogenic PCOS, are at a higher risk of developing cardiovascular disease and events later in life. These complications might also affect their life expectancy. Available studies reporting on this matter either lack numbers of well phenotyped PCOS patients, or have limited time to follow up [42-44]. Recognizing this necessity, we are currently embarking on a prospective long term follow-up study in PCOS women using a standardized screening for cardiovascular and metabolic risk factors. Furthermore, in older PCOS patients the incidence of cardiovascular events will be assessed. Hopefully this will be of help to improve the ability of both patient and physician to improve the health of PCOS patients and their offspring throughout life.

References

1. March WA, Moore VM, Willson KJ, et al. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Hum Reprod*. 2010; Feb; 25: 544-51.
2. Stein IF, Leventhal ML. Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol*. 1935; 29: 181-91.
3. Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. In: Dunaif A, Givens JR and Haseltine (eds), Polycystic Ovary Syndrome. Boston: Blackwell Scientific; 1992. p. 377-84.
4. ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004; Jan; 81: 19-25.
5. ESHRE ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.
6. Broekmans FJ, Knauff EA, Valkenburg O, et al. PCOS according to the Rotterdam consensus criteria: Change in prevalence among WHO-II anovulation and association with metabolic factors. *BJOG*. 2006; Oct; 113: 1210-7.
7. Goverde AJ, van Koert AJ, Eijkemans MJ, et al. Indicators for metabolic disturbances in anovulatory women with polycystic ovary syndrome diagnosed according to the Rotterdam consensus criteria. *Hum Reprod*. 2009; Mar; 24: 710-7.
8. Moran LJ, Misso ML, Wild RA, et al. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update*. 2010; Jul-Aug; 16: 347-63.
9. Valkenburg O, Steegers-Theunissen RP, Smedts HP, et al. A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. *J Clin Endocrinol Metab*. 2008; Feb; 93: 470-6.
10. Wild RA, Rizzo M, Clifton S, et al. Lipid levels in polycystic ovary syndrome: systematic review and meta-analysis. *Fertil Steril*. 2011; Mar 1; 95: 1073-9 e1-11.
11. Yildiz BO, Bozdag G, Yapici Z, et al. Prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome under different diagnostic criteria. *Hum Reprod*. 2012; Oct; 27: 3067-73.
12. Azziz R, Carmina E, Dewailly D, et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4237-45.
13. Dunaif A, Fauser B. Renaming PCOS - A Two State Solution. *J Clin Endocrinol Metab*. 2013; Sep 5.
14. National Institutes of Health. Final Statement Evidence-based Methodology Workshop on Polycystic Ovary Syndrome. 2012.
15. Fauser BC, Tarlatzis BC, Rebar RW, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril*. 2012; Jan; 97: 28-38 e25.
16. van Santbrink EJ, Hop WC, Fauser BC. Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril*. 1997; Mar; 67: 452-8.
17. Balen AH, Laven JS, Tan SL, et al. Ultrasound assessment of the polycystic ovary: international consensus definitions. *Hum Reprod Update*. 2003; Nov-Dec; 9: 505-14.
18. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972; Jun; 18: 499-502.

19. Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985; Jul; 28: 412-9.
20. Carmina E, Lobo RA. Use of fasting blood to assess the prevalence of insulin resistance in women with polycystic ovary syndrome. *Fertil Steril*. 2004; Sep; 82: 661-5.
21. de Paula Martins W, Santana LF, Nastri CO, et al. Agreement among insulin sensitivity indexes on the diagnosis of insulin resistance in polycystic ovary syndrome and ovulatory women. *Eur J Obstet Gynecol Reprod Biol*. 2007; Aug; 133: 203-7.
22. Grundy SM. Metabolic syndrome scientific statement by the American Heart Association and the National Heart, Lung, and Blood Institute. *Arterioscler Thromb Vasc Biol*. 2005; Nov; 25: 2243-4.
23. Perk J, De Backer G, Gohlke H, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012) : the fifth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). *Int J Behav Med*. 2012; Dec; 19: 403-88.
24. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009; 338: b2393.
25. Moran L, Teede H. Metabolic features of the reproductive phenotypes of polycystic ovary syndrome. *Hum Reprod Update*. 2009; Jul-Aug; 15: 477-88.
26. Johnstone EB, Davis G, Zane LT, et al. Age-related differences in the reproductive and metabolic implications of polycystic ovarian syndrome: findings in an obese, United States population. *Gynecol Endocrinol*. 2012; Oct; 28: 819-22.
27. Kristensen SL, Ramlau-Hansen CH, Ernst E, et al. A very large proportion of young Danish women have polycystic ovaries: is a revision of the Rotterdam criteria needed? *Hum Reprod*. 2010; Dec; 25: 3117-22.
28. Dewailly D, Gronier H, Poncelet E, et al. Diagnosis of polycystic ovary syndrome (PCOS): revisiting the threshold values of follicle count on ultrasound and of the serum AMH level for the definition of polycystic ovaries. *Hum Reprod*. 2011; Nov; 26: 3123-9.
29. Murphy MK, Hall JE, Adams JM, et al. Polycystic ovarian morphology in normal women does not predict the development of polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006; Oct; 91: 3878-84.
30. Robin G, Gallo C, Catteau-Jonard S, et al. Polycystic Ovary-Like Abnormalities (PCO-L) in women with functional hypothalamic amenorrhea. *J Clin Endocrinol Metab*. 2012; Nov; 97: 4236-43.
31. Glintborg D, Mumm H, Hougaard D, et al. Ethnic differences in Rotterdam criteria and metabolic risk factors in a multiethnic group of women with PCOS studied in Denmark. *Clin Endocrinol (Oxf)*. 2010; Dec; 73: 732-8.
32. Valkenburg O, Lao O, Schipper I, et al. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *J Clin Endocrinol Metab*. 2011; Jul; 96: E1181-7.
33. Kauffman RP, Baker VM, DiMarino P, et al. Hyperinsulinemia and circulating dehydroepiandrosterone sulfate in white and Mexican American women with polycystic ovary syndrome. *Fertil Steril*. 2006; Apr; 85: 1010-6.
34. Lo JC, Feigenbaum SL, Yang J, et al. Epidemiology and adverse cardiovascular risk profile of diagnosed polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006; Apr; 91: 1357-63.
35. Kim C, Golden SH, Mather KJ, et al. Racial/ethnic differences in sex hormone levels among postmenopausal women in the diabetes prevention program. *J Clin Endocrinol Metab*. 2012; Nov; 97: 4051-60.
36. Zhao Y, Qiao J. Ethnic differences in the phenotypic expression of polycystic ovary syndrome. *Steroids*. 2013; Aug; 78: 755-60.
37. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? *Clin Chem*. 2003; Aug; 49: 1250-1.
38. Yildiz BO. Diagnosis of hyperandrogenism: clinical criteria. *Best Pract Res Clin Endocrinol Metab*. 2006; Jun; 20: 167-76.

39. Janse F, Eijkemans MJ, Goverde AJ, et al. Assessment of androgen concentration in women: liquid chromatography-tandem mass spectrometry and extraction RIA show comparable results. *Eur J Endocrinol*. 2011; Dec; 165: 925-33.

40. Legro RS, Schlaff WD, Diamond MP, et al. Total testosterone assays in women with polycystic ovary syndrome: precision and correlation with hirsutism. *J Clin Endocrinol Metab*. 2010; Dec; 95: 5305-13.

41. Mosca L, Appel LJ, Benjamin EJ, et al. Evidence-based guidelines for cardiovascular disease prevention in women. American Heart Association scientific statement. *Arterioscler Thromb Vasc Biol*. 2004; Mar; 24: e29-50.

42. Pierpoint T, McKeigue PM, Isaacs AJ, et al. Mortality of women with polycystic ovary syndrome at long-term follow-up. *J Clin Epidemiol*. 1998; Jul; 51: 581-6.

43. Schmidt J, Landin-Wilhelmsen K, Brannstrom M, et al. Cardiovascular disease and risk factors in PCOS women of postmenopausal age: a 21-year controlled follow-up study. *J Clin Endocrinol Metab*. 2011; Dec; 96: 3794-803.

44. Wild S, Pierpoint T, McKeigue P, et al. Cardiovascular disease in women with polycystic ovary syndrome at long-term follow-up: a retrospective cohort study. *Clin Endocrinol (Oxf)*. 2000; May; 52: 595-600.

Supplementary Table 1 | Assays per centre with conversion factors.

Parameter	Erasmus Medical Centre Rotterdam (Y)			University Medical Centre Utrecht (X)			Between Centre Conversion
	Date	Assay	Conversion	Date	Assay	Conversion	
LH (U/l)	*	Immulite platform	-	start – 11/29/2006	ADVIA Centaur (x)	y=0.88x + 0.87	Y=1.06X
				11/30/2006 – end	Beckman Dxi (y)		
FSH (U/l)	*	Immulite platform	-	start – 11/29/2006	ADVIA Centaur (x)	y=1.16x + 0.46	Y=1.02X
				11/30/2006 – end	Beckman Dxi (y)		
Estradiol (pmol/l)	*	RIA DPC	-	start – 1/31/2010	internal RIA (x)	y=0.98x	Y=0.44X – 0.06
				2/1/2010 – end	Roche Modular (y)		
Testosterone (nmol/l)	start – 08/19/2012	RIA DPC (x)	y=1.05x + 0.06	*	Internal RIA	-	Y=0.85X
	08/20/2012 – end	LC-MS/MS (y)					
DHEAS (µmol/L)	start – 07/04/2013	Immulite platform (x)	y=0.84x + 0.27	start – 4/30/2012	RIA DPC (x)	y=1.39x + 0.033	Y=0.7X+ 0.38
	08/04/2013 – end	LC-MS/MS (y)		5/1/2012 - end	Roche Modular (y)		
Androstenedione (nmol/L)	start – 08/21/2012	Immulite 2000 (x)	y=0.46x + 0.44	*	Internal RIA		Y=0.91X – 0.51
	08/22/2012 – end	LC-MS/MS (y)					
SHBG (nmol/L)	*	Immulite platform	-	start- 4/25/2007	Immulite 1000 (x)	y=1.10x - 0.7	Y=0.91X
				4/26/2007 - end	Roche Modular (y)		
Insulin (mIU/L)	*	Immulite platform	-	start – 12/15/2010	Immulite 1000 (x)	y= 1.3x	Y=0,70X
				12/16/2010 – end	Roche Modular (y)		

Supplementary Table 1 | *Continued*

Parameter	Erasmus Medical Centre Rotterdam (Y)			University Medical Centre Utrecht (X)			Between Centre Conversion
	Date	Assay	Conversion	Date	Assay	Conversion	
Glucose (mmol/L)	start – 04/../2008	Hitachi 917	-	start – 11/30/2006	VITROS	-	-
	04/../2008 – 03/20/2013	Roche Modular		12/1/2006 – end	Unicell DxC 800		
	03/21/2013 – end	Cobas 8000					
Lipids (mmol/L) (HDL-C, TG, Total-C); LDL-C calculated		CA Wako		start – 11/30/2006	VITROS		
				12/1/2006 – 5/6/2012	Unicell DxC 800		
				5/7/2012 – end	AU 5811		

If applicable, results of assays were first internally adjusted per centre according to the standards of the last used assay. Subsequently only one between centre conversion factor was applied to ensure between centre homogeneity. * Assay used during entire study period, – conversion factor not applicable.

ADVIA Centaur: *Automated System, Bayer Corporation, Tarrytown, USA.*
Beckman Dxi system, Unicell DxC 800, AU 5811: *Beckman Coulter, Woerden, Netherlands*
Cobas 8000, Hitachi 917 analyzer, Roche Modular E170, Roche Modular P: *Roche Diagnostics Almere, Netherlands*..
Immulite platform, Immulite 1000, Immulite 2000, RIA DPC: *Siemens DPC Los Angeles, CA, USA.*
VITROS Chemistry System*: Ortho-Clinical Diagnostics, Strasbourg, France*
Internal RIA, In house developed extraction RIA.
LC-MS/MS, Liquid Chromatography-Tandem Mass Spectrometry

Chapter 3

Severe Excess Mortality in Mothers of Patients with Polycystic Ovary Syndrome

Y.V. Louwers
M.E. Roest-Schalken
N. Kleefstra
J. Roeters van Lennep
M. van den Berg
B.C.J.M. Fauser
H.J.G. Bilo
E.J.G. Sijbrands
J.S.E. Laven

Manuscript submitted

Abstract

Context: Because of the heritable nature of Polycystic Ovary Syndrome (PCOS), parents of these patients are also prone to develop type 2 diabetes mellitus, which might influence their life-expectancy.

Objective: All-cause mortality of parents of patients with PCOS was determined. Parents with diabetes were compared to controls, whose diabetes was not related to PCOS.

Design: Reverse parent-offspring study.

Setting: University hospital and general community.

Participants: The medical history of mothers (n=946) and fathers (n=902) was primarily obtained during the initial screening of the patient and updated via questionnaires. The control-population consisted of 1353 men and women diagnosed with type 2 diabetes mellitus.

Main outcome measures: Standardized mortality ratio (SMR) was calculated as the ratio of the observed mortality of the parents to the expected mortality in the Dutch general population. The mortality of parents with type 2 diabetes mellitus relative to controls with diabetes not related to PCOS was standardized for age, gender and calendar period using Poisson regression.

Results: 302 parents deceased in 62,693 person-years. Mothers above age 60 had significant excess mortality of 1.50 (95% CI 1.15 to 1.92). Moreover, mothers with diabetes had two-times higher mortality risk compared to control women with diabetes (RR 2.0, 95% CI 1.19 to 3.41). No excess mortality among fathers was observed.

Conclusions: Type 2 diabetes among mothers of PCOS patients resulted in severe excess mortality compared to women with diabetes from the general population. We strongly advise screening of these mothers to ensure that timely preventive and therapeutic measures are taken.

Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in women of reproductive age with a prevalence of 6-10% in an unselected population [1]. PCOS is a life-long condition with considerable variation of symptomatology between and within patients over time [2]. The syndrome is characterized by ovulatory dysfunction, hyperandrogenism and polycystic appearance of the ovaries [3]. Insulin resistance is a prominent feature of PCOS [4]. Also, a higher rate of impaired glucose tolerance, gestational diabetes and type 2 diabetes mellitus was found in patients with PCOS compared with controls [5,6]. Obesity is present in up to 60% of all patients with PCOS and is an exacerbating factor for the development of insulin resistance, impaired glucose tolerance and type 2 diabetes mellitus [7]. Type 2 diabetes mellitus increases the risk of cardiovascular and reduces lifetime expectancy [8]. Studies assessing mortality rates of patients with PCOS and their family members are scarce due to small sample sizes and limited follow-up time of the patients [9-12].

Genetic factors play an important role in the etiology of PCOS as suggested by the estimated heritability of 65% [13]. The syndrome itself and notably the associated co-morbidities cluster both in families. Parents of patients with PCOS have more often insulin resistance compared with age and BMI matched controls with a healthy daughter [14,15]. Consequently, they seem to be at increased familial risk of developing cardiometabolic problems [16]. A previous study indicated that fathers of patients with PCOS had a high prevalence of heart attacks and strokes compared to a reference population [17]. Moreover, an increase has been observed of premature cardiovascular events among mothers of patients with PCOS [18]. Whether presence of diabetes plays a role in this increased cardiovascular risk is insufficiently known.

We hypothesized that the susceptibility to type 2 diabetes mellitus explains an important part of the complications among the patients with PCOS and their parents. Long-term follow-up studies with adequate sample size and well-phenotyped patients with PCOS assessing mortality, the most indisputable outcome, are ongoing but still require approximately three decades of follow-up. Because of the high heritability and to overcome the lack of appropriate follow-up data, we determined all-cause mortality in parents of patients with PCOS compared with the mortality in the general Dutch population. Moreover, we determined all-cause mortality of parents with type 2 diabetes mellitus compared to the mortality of men and women with type 2 diabetes, who were recruited without selection for PCOS.

Methods

PCOS diagnosis and assessment

Women with oligomenorrhea (interval of menstrual periods of at least 35 days) or amenorrhea (absence of vaginal bleeding for over 6 months) were examined at the outpatient clinic for cycle disturbances of the ErasmusMC University Medical Center Rotterdam. Medical history

was obtained. Extensive clinical and endocrine examination was performed as described previously [19]. Diagnosis of PCOS was based on the Rotterdam criteria [3]. According to these criteria, two of the following symptoms need to be present: oligo- or anovulation, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology. Clinical hyperandrogenism was defined as a modified Ferriman-Gallwey score of at least 8 [20]. Biochemical hyperandrogenism was determined by the calculation of the Free Androgen Index (FAI): 100 x T in nmol/L / SHBG in nmol/L. A cut-off level of 4.5 was used to define biochemical hyperandrogenism. Polycystic ovarian morphology was assessed by transvaginal ultrasound and a cut-off of 12 or more follicles in at least one ovary and/or an ovarian volume of at least 10 ml was used to define polycystic ovarian morphology [21]. Informed consent from every patient who visited the outpatient clinic was obtained, according to international review board standards of the ErasmusMC University Medical Center Rotterdam. This protocol has been approved by the medical ethical review board of the ErasmusMC University Medical Center Rotterdam according to the Declaration of Helsinki.

Parental life years and parental medical history

Birthdates and mortality dates of the parents were obtained from a nationwide web-based municipal record database (Dutch name: Gemeentelijke Basisadministratie Persoons-gegevens). This nationwide municipal database contains official records of births, marriages, and deaths of all government registered people living in the Netherlands and the Netherlands Antilles. Parental person-years were calculated by using dates of birth, death, and censoring (end of follow-up: January 1, 2007). In case this information was untraceable for both parents, the patient was excluded from further analyses. The parental medical history was primarily obtained during the initial screening of the patient. To complete and update the information about the prevalence of type 2 diabetes in their parents, a questionnaire was sent to all screened patients with PCOS.

Population-based cohort of patients diagnosed with type 2 diabetes mellitus

Mortality rates of men and women diagnosed with type 2 diabetes mellitus were collected from the Zwolle Outpatient Diabetes Project Integrating Available Care (ZODIAC) study. This large diabetes project was initiated in 1998 in the Zwolle region of the Netherlands. In its first phase, general practitioners were assisted by hospital-based diabetes specialist nurses in providing care for patients with type 2 diabetes. Patients with a very short life expectancy and patients with insufficient cognitive abilities were excluded. A total of 1,353 (90%) patients agreed to participate in the study. The details of this study have been published previously [22]. The ZODIAC study was approved by the local medical ethics committee and all included subjects provided informed consent.

Statistical analysis

Baseline characteristics of the patients with PCOS were evaluated using a commercially available statistical package (IBM Statistical Package of the Social Science/Predictive Analytic Software version 20). Man-Whitney-U tests and Chi-square tests, if necessary with continuity correction, were performed to compare baseline characteristics of the patients.

The all-cause mortality of the parents of the patients with PCOS was compared to the all-cause mortality of men and women from the general Dutch population standardized for age, gender, and calendar period, as previously described [23,24]. The ratio of the observed to the expected numbers of deaths is the standardized mortality ratio (SMR). The expected mortality was calculated by multiplying the total number of years lived by the study population until January 1, 2007 in each calendar period, with the age and gender specific mortality rates in the Dutch general population for each calendar period. These data are available at Statistics Netherlands using software of the World Health Organization [23].

Statistics Netherlands is responsible for collecting and processing data of a multitude of societal aspects including mortality rates. The parental years before birth of the patient with PCOS were excluded from the analyses. Including these years would lead to selecting years without deaths resulting in underestimation of the mortality risk. Calendar periods were divided into five-year intervals, and to each of these intervals we applied the population mortality rates. We also calculated the SMR in different age-groups. The 95% Confidence Interval (CI) of the SMR was calculated assuming a Poisson distribution of the observed number of deaths and by using exact limits [24]. The SMR of the diabetic patients, who were recruited from the general population, was calculated after excluding the years of life before entering the ZODIAC study.

Finally, we used Poisson regression to compare the mortality in the parents of the patients with PCOS with the mortality in the patients with diabetes who were recruited from the general population by calculating the relative risk (RR) standardized for age, gender and calendar period. A p-value < 0.05 was considered to be statistically significant.

Results

In total 1088 patients diagnosed with PCOS were eligible. Of these, the address of 130 patients could not be traced. They were therefore considered as lost to follow-up. Dates of birth, death, and censoring of at least one parent were available for the remaining 958 (88.1%), patients with PCOS who were included for further analyses. Baseline characteristics of included and excluded patients with PCOS are shown in Table 1. PCOS features and associated characteristics were very similar in both groups.

Table 1 | Baseline characteristics of the patients with PCOS.

	Included patients (n=958)	Excluded patients (n= 130)	*P* value
Age (years)	28.7 (6.4)	27.5 (6.1)	0.06
BMI (kg/m^2)	25.6 (9.0)	27.0 (9.0)	0.24
Waist circumference (cm)	86 (23)	88 (22)	0.52
Polycystic Ovaries (%)	90.1 %	84.7%	0.09
Hyperandrogenism (%)	61.4 %	66.9 %	0.26
Amenorrhea (%)	27.6 %	22.3 %	0.25
FSH (IU/L)	4.6 (2.3)	4.30 (2.4)	0.07
LH (IU/L)	7.1 (5.9)	6.5 (6.3)	0.49
SHBG (nmol/L)	40.1 (36.3)	36.9 (30.5)	0.21
Testosterone (nmol/L)	2.0 (1.3)	2.0 (1.3)	0.47
FAI	4.9 (5.4)	5.3 (5.4)	0.49
Androstenedione (nmol/L)	12.4 (6.8)	11.8 (7.6)	0.86
Glucose (mmol/L)	4.1 (0.8)	4.20 (0.7)	0.03
Insulin (pmol/L)	71 (66)	72 (67)	0.46

Baseline characteristics are shown as median and interquartile ranges (IQR), unless indicated otherwise. *BMI,* Body mass index; *FSH,* Follicle Stimulating Hormone; *LH,* Luteinizing Hormone, *SHBG,* Sex Hormone Binding Globulin, *FAI,* Free Androgen Index.

First, we evaluated whether having a daughter with PCOS influenced all-cause mortality. Information on birth and mortality dates was available of 1848 (96.5%) of the 1916 parents, i.e. 946 (98.7%) of the mothers and 902 (94.1%) of the fathers. A total of 302 deaths were observed in 62,693 person years (Table 2). The overall life expectancy of the parents independent of their medical history was not influenced by the fact that their daughters were diagnosed with PCOS (SMR 1.04; 95% CI 0.98 to 1.23), nor was the life expectancy of mothers (SMR 1.13; 95% CI 0.93 to 1.37) and fathers (SMR 0.99; 95% CI 0.86 to 1.14) separately affected. As predefined in the SMR method, data were also analyzed in different age categories. Figure 1 summarizes the SMR of mothers and fathers in different age ranges standardized for gender and calendar period. The SMR of the mothers above 60 years of age was 1.50 (95% CI 1.15 to 1.92).

Table 2 | Standardized Mortality Ratio of parents of patients with PCOS compared to the general Dutch population and of patients with diabetes compared to the general Dutch population.

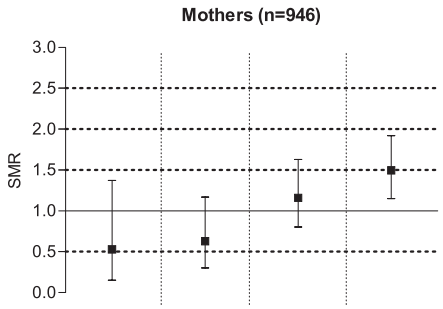
	Person years	Observed deaths	Expected deaths	SMR	95% CI	*P* value
All parents	62693.54	302	291.44	1.04	0.98-1.23	0.276
All mothers	32785.34	106	93.69	1.13	0.93-1.37	0.113
All fathers	29908.19	196	197.75	0.99	0.86-1.14	0.559
Mothers with type 2 diabetes*	3169.83	22	9.91	2.22	1.39-3.36	0.001
Fathers with type 2 diabetes*	3204.14	25	21.40	1.17	0.76-1.72	0.245
Female diabetic patients†	6063.67	326	219.94	1.48	1.33-1.65	< 0.001
Male diabetic patients†	4369.25	244	174.01	1.40	1.23-1.59	< 0.001

* Mothers and fathers diagnosed with type 2 diabetes mellitus and having a daughter diagnosed with PCOS. † Females and males diagnosed with type 2 diabetes mellitus and recruited from a population-based cohort study. *DM2,* type 2 diabetes mellitus; *SMR,* standardized mortality ratio; *CI,* confidence interval.

In total, 586 patients with PCOS replied to the questionnaires, resulting in a response rate of 61.2% (586 out of 958). The overall SMRs were similar between parents, whose information on medical health was available, and parents, whose information was missing. Type 2 diabetes was present in 15.6% (95% CI 13.5-17.6%) of the parents (mothers n=90; fathers n=98), while it was 8.1% in an age-matched samples of the Dutch population (difference 7.5%; 95% CI 5.4-9.5%; *P* value = < 0.001) [25]. Mothers with type 2 diabetes mellitus had excess mortality (SMR 2.22; 95% CI 1.39 to 3.36) compared to the general Dutch population (Table 2). Moreover, the mothers with diabetes had a two-fold excess mortality compared to the diabetic female patients who were recruited from the general population (RR 2.0: 95% CI 1.19 to 3.41; *P* value 0.009) standardized for age and calendar period. The fathers with type 2 diabetes mellitus (SMR 1.17; 95% CI 0.76-1.72) exhibited a shorter lifespan; however this difference was not significant. No differences in mortality were observed between fathers with diabetes and male patients with diabetes, who were recruited from the general population (RR 0.96; 95% CI 0.59 to 1.57; *P* value = 0.87).

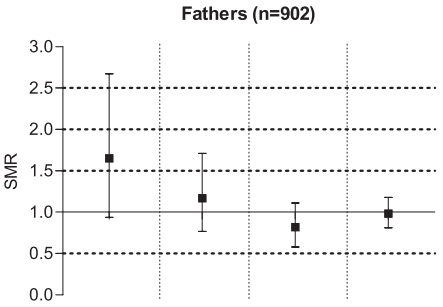
Figure 1 | Standardized mortality ratio, number of deaths and person years of mothers (**panel A**) and fathers (**panel B**) compared to the general population divided by age categories.

A



age (years)	< 40	40-49	50-59	> 60
deaths	4	10	33	59
PYRS	12558.71	9153.54	7335.36	3737.72

B



age (years)	< 40	40-49	50-59	> 60
deaths	16	27	41	112
PYRS	9176.04	8437.01	7440.19	4854.94

SMR, Standardized Mortality Ratio; *PYRS,* Person Years.

Discussion

Our study has shown that mothers above age 60 years with a daughter diagnosed with PCOS do have an increased mortality risk. Moreover, severe excess mortality was observed in mothers with type 2 diabetes mellitus, being twice as high as in female controls with diabetes.

The familial and heritable nature of PCOS has been well established [13]. Not only PCOS itself, but also associated co-morbidities of the syndrome evidently affect family members of

patients with PCOS. Our findings support the notion that parents of patients diagnosed with PCOS develop type 2 diabetes mellitus more often than age-matched subjects in the general population [25]. In the Dutch population 73% (95% CI 72.1-74.7%) of the patients had onset of type 2 diabetes mellitus above age 60 years [25]. We observed excess mortality in mothers above age 60 years relative to the general Dutch population, fully in line with the expected age of onset of chronic diseases. The presence of type 2 diabetes mellitus is associated with increased risk of early cardiovascular as well as all-cause mortality: it has been postulated that patients with type 2 diabetes mellitus should be immediately treated as if they had prior coronary disease [8].

Over the last few decades the prevalence of type 2 diabetes mellitus started to increase in the Netherlands [26]. Implementation of strict treatment guidelines and improvement of quality of diabetes care in shared care settings have introduced a trend towards mortality more similar to the general population [27]. Nonetheless, we even observed a two-fold increased mortality of diabetic mothers of patients with PCOS relative to female patients with diabetes recruited from the general population. Notably, a greater mortality risk has been observed in females than males with diabetes [28-30]. Although mechanisms underlying these gender differences are not fully understood, it has been shown that surrogate markers for cardiovascular disease, such as elevated blood pressure, dyslipidemia and increased carotid intima-media thickness, were more prominently present in females than in males with diabetes [31]. These risk markers are also well-known co-morbidities of PCOS [32-34]. Although we did not assess presence of PCOS in the post-menopausal mothers, previous studies have shown that mothers of PCOS patients do share clinical characteristics in at least half of all cases. Hence, they suffer more often from irregular menstrual cycles and hirsutism compared to mothers with daughters without PCOS [16]. Moreover, menstrual cycle irregularity, elevated androgen levels as well as presence of diabetes predispose to cardiovascular disease and events [35-37]. Therefore it can be hypothesized that the higher mortality in mothers with type 2 diabetes mellitus might be related to reproductive disorders, such as PCOS. Current efforts to obtain endpoint data based on large prospective follow-up studies of patients with PCOS and their family members will provide more profound insight in the coming decades.

The presented reverse parent-offspring analysis overcomes the problem of lacking adequate follow-up data and, because of high heritability of the syndrome, may serve as a warning for the future risk of the PCOS daughters. More importantly, the excess of mortality observed in diabetic mothers of patients with PCOS relative to the patients with diabetes recruited from the general population suggests that this specific high-risk group is either not sufficiently recognized limiting the use of preventive measures. Therefore, we feel that early and active screening as well as aggressive treatment for type 2 diabetes mellitus amongst these mothers of daughters suffering from PCOS is justified. We realize that we observed a limited number of deaths. However, the SMR method uses person-years and the corresponding population-based expected deaths. Therefore, power problems hardly occur and small numbers of deaths can be used to estimate differences with high confidence as long

as the number of person-years is large, which is the case in our study. The increased mortality of the mothers with type 2 diabetes mellitus is within the range of mortality caused by severe inherited cardiovascular disorders like familial hypercholesterolemia [38]. Early identification and treatment of cardiovascular risk factors may be of major importance in this specific high risk group, also in the light of the fact that the incidence of cardiovascular disease in women with diabetes begins to increase at least 15 years before diabetes is clinically diagnosed [39]. Moreover, it is also known that type 2 diabetes is much more common in patients with PCOS and about 20% of these develop diabetes as early as during their forth decade of life [40]. In case of such a double burden, early diagnosis of diabetes will hopefully lead to early and aggressive treatment of risk factors, which in turn will have its effects on primary outcome.

This is the first study assessing this excess risk in parents of patients diagnosed having PCOS, the most common endocrine disease in women of reproductive age. One could argue that we excluded severe cases from the population-based diabetic cohort. Excluding subjects with a very short life expectancy during the first months of a population-based cohort study is performed to avoid effects of severe, prevalent diseases on the follow-up of the cohort limiting the observations to truly incident cases. If this approach had introduced a bias, we would have expected excess mortality in the mothers below 60 years of age. However, this was not the case, implying that type 2 diabetes mellitus is indeed a disease of the elderly in our population. Questionnaires were used to get insight in the co-morbidities of the parents, and unfortunately we did not get a maximum response. Although this may have influenced our study, we did not observe any differences in mortality rates between the parents of the patients, who responded to the questionnaire, compared to those, who did not respond. The questionnaires may also have introduced recall bias for family history. However, this was previously demonstrated to be minimal for long-term chronic diseases [41]. Moreover, the high prevalence of type 2 diabetes mellitus in the parents of our patients is in agreement with earlier studies [12, 14]. We were not able to assess potential risk factors, including obesity hypertension or dyslipidemia, in the studied parents, which might shed some light on the potential mechanisms underlying the high mortality of the diabetic mothers. Future studies are needed to identify these potential additional risk factors.

In conclusion, we observed severe excess mortality above the age of 60 years in mothers of patients with PCOS, most likely as a result of type 2 diabetes mellitus. Mothers with type 2 diabetes mellitus had a two-fold excess mortality compared to females with diabetes recruited from the general population without selection on PCOS. Our findings justify screening for type 2 diabetes mellitus among the mothers with a daughter suffering from PCOS to ensure that good preventive and therapeutic measures according to the appropriate guidelines can be taken timely.

References

1. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
2. ESHRE ASRM Sponsored PCOS Consensus Workshop Group. Consensus on women's health aspects of polycystic ovary syndrome (PCOS). *Hum Reprod*. 2012; Jan; 27: 14-24.
3. ESHRE ASRM Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.
4. Dunaif A, Segal KR, Futterweit W, et al. Profound peripheral insulin resistance, independent of obesity, in polycystic ovary syndrome. *Diabetes*. 1989; Sep; 38: 1165-74.
5. Legro RS, Kunselman AR, Dodson WC, et al. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999; Jan; 84: 165-9.
6. Boomsma CM, Fauser BC, Macklon NS. Pregnancy complications in women with polycystic ovary syndrome. *Semin Reprod Med*. 2008; Jan; 26: 72-84.
7. Barber TM, Wass JA, McCarthy MI, et al. Metabolic characteristics of women with polycystic ovaries and oligo-amenorrhoea but normal androgen levels: implications for the management of polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2007; Apr; 66: 513-7.
8. Haffner SM, Lehto S, Ronnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med*. 1998; Jul 23; 339: 229-34.
9. Pierpoint T, McKeigue PM, Isaacs AJ, et al. Mortality of women with polycystic ovary syndrome at long-term follow-up. *J Clin Epidemiol*. 1998; Jul; 51: 581-6.
10. Schmidt J, Landin-Wilhelmsen K, Brannstrom M, et al. Cardiovascular disease and risk factors in PCOS women of postmenopausal age: a 21-year controlled follow-up study. *J Clin Endocrinol Metab*. 2011; Dec; 96: 3794-803.
11. Morgan CL, Jenkins-Jones S, Currie CJ, et al. Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a retrospective, observational study. *J Clin Endocrinol Metab*. 2012; Sep; 97: 3251-60.
12. Davies MJ, Marino JL, Willson KJ, et al. Intergenerational associations of chronic disease and polycystic ovary syndrome. *PLoS One*. 2011; 6: e25947.
13. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2100-4.
14. Sir-Petermann T, Angel B, Maliqueo M, et al. Prevalence of Type II diabetes mellitus and insulin resistance in parents of women with polycystic ovary syndrome. *Diabetologia*. 2002; Jul; 45: 959-64.
15. Yilmaz M, Bukan N, Ersoy R, et al. Glucose intolerance, insulin resistance and cardiovascular risk factors in first degree relatives of women with polycystic ovary syndrome. *Hum Reprod*. 2005; Sep; 20: 2414-20.
16. Sam S, Legro RS, Essah PA, et al. Evidence for metabolic and reproductive phenotypes in mothers of women with polycystic ovary syndrome. *Proc Natl Acad Sci U S A*. 2006; May 2; 103: 7030-5.
17. Taylor MC, Reema Kar A, Kunselman AR, et al. Evidence for increased cardiovascular events in the fathers but not mothers of women with polycystic ovary syndrome. *Hum Reprod*. 2011; Aug; 26: 2226-31.
18. Cheang KI, Nestler JE, Futterweit W. Risk of cardiovascular events in mothers of women with polycystic ovary syndrome. *Endocr Pract*. 2008; Dec; 14: 1084-94.
19. Imani B, Eijkemans MJ, te Velde ER, et al. Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab*. 1998; Jul; 83: 2361-5.
20. Ferriman D, Gallwey JD. Clinical assessment of body hair growth in women. *J Clin Endocrinol Metab*. 1961; Nov; 21: 1440-7.

21. Balen AH, Laven JS, Tan SL, et al. Ultrasound assessment of the polycystic ovary: international consensus definitions. *Hum Reprod Update*. 2003; Nov-Dec; 9: 505-14.
22. Ubink-Veltmaat LJ, Bilo HJ, Groenier KH, et al. Shared care with task delegation to nurses for type 2 diabetes: prospective observational study. *Neth J Med*. 2005; Mar; 63: 103-10.
23. Coleman MP, Herman C, Douglas A. Person-years (PYRS): a further program for cohort study analysis. *Lyon: WHO*. 1989.
24. VandenBroucke JP. A shortcut method for calculating the 95 percent CI of the SMR. *Am J Epidemiol*. 1982; 115: 303-4.
25. Ubink-Veltmaat LJ, Bilo HJ, Groenier KH, et al. Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: a prospective population-based study in The Netherlands (ZODIAC-1). *Eur J Epidemiol*. 2003; 18: 793-800.
26. Baan CA, van Baal PH, Jacobs-van der Bruggen MA, et al. [Diabetes mellitus in the Netherlands: estimate of the current disease burden and prognosis for 2025]. *Ned Tijdschr Geneeskd*. 2009; May 30; 153: 1052-8.
27. van Hateren KJ, Drion I, Kleefstra N, et al. A prospective observational study of quality of diabetes care in a shared care setting: trends and age differences (ZODIAC-19). *BMJ Open*. 2012; 2.
28. Booth GL, Kapral MK, Fung K, et al. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *Lancet*. 2006; Jul 1; 368: 29-36.
29. Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ*. 2006; Jan 14; 332: 73-8.
30. Becker A, Bos G, de Vegt F, et al. Cardiovascular events in type 2 diabetes: comparison with nondiabetic individuals without and with prior cardiovascular disease. 10-year follow-up of the Hoorn Study. *Eur Heart J*. 2003; Aug; 24: 1406-13.
31. Juutilainen A, Kortelainen S, Lehto S, et al. Gender difference in the impact of type 2 diabetes on coronary heart disease risk. *Diabetes Care*. 2004; Dec; 27: 2898-904.
32. Ehrmann DA, Liljenquist DR, Kasza K, et al. Prevalence and predictors of the metabolic syndrome in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006; Jan; 91: 48-53.
33. Guzick DS, Talbott EO, Sutton-Tyrrell K, et al. Carotid atherosclerosis in women with polycystic ovary syndrome: initial results from a case-control study. *Am J Obstet Gynecol*. 1996; Apr; 174: 1224-9; discussion 29-32.
34. Valkenburg O, Steegers-Theunissen RP, Smedts HP, et al. A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. *J Clin Endocrinol Metab*. 2008; Feb; 93: 470-6.
35. Krentz AJ, von Muhlen D, Barrett-Connor E. Searching for polycystic ovary syndrome in postmenopausal women: evidence of a dose-effect association with prevalent cardiovascular disease. *Menopause*. 2007; Mar-Apr; 14: 284-92.
36. Shaw LJ, Bairey Merz CN, Azziz R, et al. Postmenopausal women with a history of irregular menses and elevated androgen measurements at high risk for worsening cardiovascular event-free survival: results from the National Institutes of Health--National Heart, Lung, and Blood Institute sponsored Women's Ischemia Syndrome Evaluation. *J Clin Endocrinol Metab*. 2008; Apr; 93: 1276-84.
37. Solomon CG, Hu FB, Dunaif A, et al. Menstrual cycle irregularity and risk for future cardiovascular disease. *J Clin Endocrinol Metab*. 2002; May; 87: 2013-7.
38. Sijbrands EJ, Westendorp RG, Defesche JC, et al. Mortality over two centuries in large pedigree with familial hypercholesterolaemia: family tree mortality study. *BMJ*. 2001; Apr 28; 322: 1019-23.
39. Hu FB, Stampfer MJ, Haffner SM, et al. Elevated risk of cardiovascular disease prior to clinical diagnosis of type 2 diabetes. *Diabetes Care*. 2002; Jul; 25: 1129-34.
40. Legro RS, Gnatuk CL, Kunselman AR, et al. Changes in glucose tolerance over time in women with polycystic ovary syndrome: a controlled study. *J Clin Endocrinol Metab*. 2005; Jun; 90: 3236-42.
41. Bensen JT, Liese AD, Rushing JT, et al. Accuracy of proband reported family history: the NHLBI Family Heart Study (FHS). *Genet Epidemiol*. 1999; 17: 141-50.

Chapter 4

Candidate Gene Studies

Chapter 4.1

Replication of Association of a Novel Insulin Receptor Gene Polymorphism with Polycystic Ovary Syndrome

M.O. Goodarzi
Y.V. Louwers
K.D. Taylor
M.R. Jones
J. Cui
S. Kwon
Y.I. Chen
X. Guo
L. Stolk
A.G. Uitterlinden
J.S.E. Laven
R. Azziz

Fertility & Sterility 2011 Apr;95(5):1736-41

Abstract

Objective: To evaluate association with polycystic ovary syndrome (PCOS) of 295 variants in 39 genes central to metabolic insulin signaling and glycogen synthase kinase 3b (GSK-3b) regulation, followed by replication efforts.

Design: Case-control association study, with discovery and replication cohorts.

Setting: Subjects were recruited from reproductive endocrinology clinics, and controls were recruited from communities surrounding the University of Alabama at Birmingham and Erasmus Medical Center, Rotterdam.

Patient(s): A total of 273 cases with PCOS and 173 control subjects in the discovery cohort; and 526 cases and 3,585 control subjects in the replication cohort. All subjects were Caucasian.

Intervention(s): Phenotypic and genotypic assessment.

Main Outcome Measure(s): Detection of 295 single-nucleotide polymorphisms (SNPs), PCOS status.

Result(s): Several SNPs were associated with PCOS in the discovery cohort. Four insulin receptor (*INSR*) SNPs and three insulin receptor substrate 2 (*IRS2*) SNPs associated with PCOS were genotyped in the replication cohort. One *INSR* SNP (rs2252673) replicated association with PCOS. The minor allele conferred increased odds of PCOS in both cohorts, independent of body mass index.

Conclusion(s): A pathway-based tagging SNP approach allowed us to identify novel *INSR* SNPs associated with PCOS, one of which confirmed association in a large replication cohort.

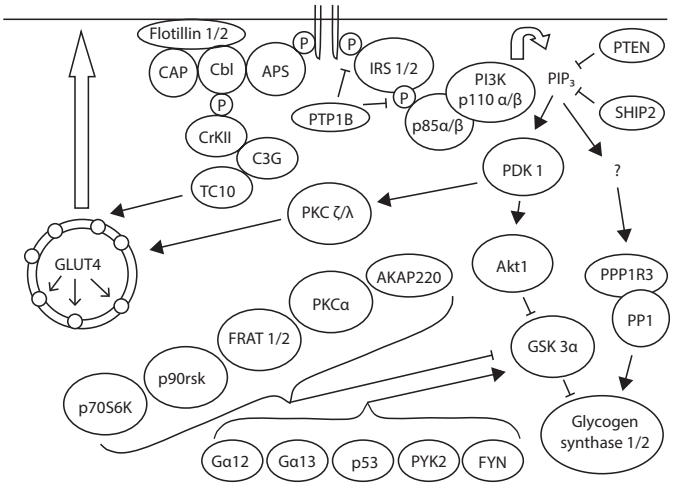
Introduction

Insulin resistance and compensatory hyperinsulinemia are frequent findings in polycystic ovary syndrome (PCOS), affecting 70% of cases [1]. The fact that insulin-sensitizing agents ameliorate features of PCOS points to insulin resistance as a key pathophysiologic factor [2].

Not only is PCOS itself heritable, but within PCOS, insulin resistance is under significant genetic control [3]. First-degree relatives of women with PCOS exhibit an increased prevalence of insulin resistance whether or not they have PCOS [4]. For these reasons, some of the earliest candidate genes examined for PCOS were components of the insulin-signaling pathway [5]. However, insulin signaling, which starts with insulin binding to the insulin receptor (INSR) on the cell surface, is a complex system that interacts with other signaling pathways [6]. INSR initiates metabolic effects (e.g., stimulation of glucose uptake and glycogen synthesis) via two main pathways, the phosphatidylinositol-3 (PI3) kinase cascade and the CAP/Cbl pathway [7]; it also promotes cell proliferation via the mitogen-activated protein kinase cascade. These are the best- understood immediate systems activated by the INSR. Given the considerable number of components of these systems, insulin-signaling pathways have been incompletely examined to date. Examples of studied genes include *AKT2*, *INSR*, *IRS1*, *IRS2*, *GSK3B*, *PTP1B*, *PPP1R3*, and *SORBS1* [8].

What has been needed is an analysis that considers all of the key components of the insulin signaling pathway by using a tagging SNP approach to encompass the majority of genetic variation across the target regions. Advances in high-throughput genotyping have made this possible. Therefore, to fully interrogate the insulin signaling pathway for susceptibility loci for PCOS, we genotyped the *INSR* gene plus 27 genes coding for pathway components related to metabolic effects (PI3 kinase and CAP/Cbl pathways; Figure 1). We also genotyped 11 genes that have been described as regulators of glycogen synthase kinase 3b [9], given our earlier genetic and physiologic results implicating this factor and its upstream regulator *AKT2* in PCOS [10-12]. The genes with the most significant SNPs associated with PCOS in the discovery phase were *INSR* and *IRS2*. Therefore, four SNPs in *INSR* and three SNPs in *IRS2* were genotyped and analyzed in a larger replication cohort. One of the *INSR* SNPs replicated association with PCOS.

Figure 1 | Metabolic insulin-signaling pathway genes genotyped in the discovery cohort.



The protein products of the 39 genes that were genotyped in the discovery cohort are indicated. Names containing a slash indicate two isoforms encoded by two different genes. *p,* phosphate group; *PIP3,* phosphatidylinositol (3,4,5)-triphosphate. Full gene names are listed in Supplementary Table 2.

Materials and Methods

Subjects and phenotyping

Discovery cohort

We studied 275 unrelated White PCOS patients and 173 White control women recruited at the University of Alabama at Birmingham (UAB). Study subjects were premenopausal, non-pregnant, and on no hormonal therapy, including oral contraceptives or insulin-sensitizing agents for > 3 months, and they met 1990 National Institutes of Health (NIH) criteria for PCOS [13]. Parameters for defining hirsutism, hyperandrogenemia, ovulatory dysfunction, and exclusion of related disorders were previously reported [14]. Supplementary Table 1 presents clinical characteristics.

Control subjects were healthy women, with regular menstrual cycles and no evidence of hirsutism, acne, alopecia, or endocrine dysfunction and who had not taken hormonal therapy (including oral contraceptives) for > 3 months. Controls were recruited by word of mouth and advertisements calling for "healthy women." Each of the subjects gave written informed consent. The study was performed according to the guidelines of the Institutional Review Boards of UAB and Cedars-Sinai Medical Center.

Replication cohort

Replication efforts were carried out in an independent cohort consisting of 526 Caucasian PCOS women and 3,585 unselected control subjects from the general population. The replication cohort included patients with oligomenorrhea (menstrual cycle > 35 days) or amenorrhea (absence of vaginal bleeding for > 6 months), serum FSH levels within the normal limits (1-10 U/L), i.e. normogonadotropic anovulation (classification according to the World Health Organization), and PCOS. PCOS was diagnosed according to the 2003 revised Rotterdam criteria [15]. Patients with non-Caucasian ethnic background were excluded.

Control subjects were drawn from the Rotterdam Study, of which the design was described previously [16]. In brief, this is a large population- based study of elderly subjects from a specific area near Rotterdam (Ommoord). All women with age of menopause > 45 years and available DNA were included in the present analysis, providing a reference group of allele frequencies reflective of the local general Caucasian population (rather than a control group wherein PCOS was specifically excluded). The study was approved by the Institutional Review Board at the Erasmus Medical Center, Rotterdam, and informed consent was obtained from each of the patients as well as control subjects.

Genotyping

SNPs were selected by using data from the Caucasian (CEU) subjects in the International HapMap database (release 24; http://hapmap.ncbi.nlm.nih.gov) with the aim of exploiting linkage disequilibrium (LD) for the study of each gene. For all 39 genes, we selected SNPs to tag ($r^2 > 0.8$) the majority of variation in each gene plus 10 kb upstream and 3 kb downstream or 5% upstream and 5% downstream (depending on gene size), resulting in a total of 341 SNPs. Supplementary Table 2 displays gene statistics; Supplementary Table 3 displays SNP statistics. The average SNP coverage was 74%. 338 of the 341 SNPs were genotyped by using the oligoligation assay (GoldenGate Chemistry) on an Illumina (San Diego, CA) BeadStation [17]. Six of the SNPs were not polymorphic, and one failed Hardy-Weinberg equilibrium ($P < .001$); these SNPs were not considered further. SNPs were clustered in a single project that included all samples and used Illumina's BeadStudio clustering algorithm. Each SNP was then manually reviewed to remove individual samples from each cluster plot that were not clearly assigned to one cluster (major or minor allele homozygotes or heterozygotes). SNPs that did not clearly form three clusters for each genotype group were removed from the analysis. Thirty-nine SNPs were removed because they failed this quality control. Therefore, 292 SNPs were ultimately analyzed of the 338 genotyped on this platform. The genotyping success rate for these 292 SNPs was 93.6%. Duplicate genotyping of ten subjects yielded 100% concordance.

The remaining three of the 341 SNPs (from the *IRS1* gene) were genotyped by using Applied Biosystems Taqman Assays-On-Demand (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. The overall genotyping success rate was 93.4%. Four *INSR* and three *IRS2* SNPs were genotyped in the replication study.

For the PCOS cases, Taqman Allelic Discrimination (Applied Biosystems) was used for the *INSR* SNPs. For the Rotterdam Study control subjects, genotypes from two INSR SNPs (rs12459488 and rs12971499) were extracted from the Genome-Wide Association Study (GWAS) imputed genotypes data- base, and INSR SNPs rs2252673 and rs10401628 were genotyped by using Taqman Allelic Discrimination. The *IRS2* SNPs (rs7997595, rs7987237, rs1865434) were extracted from the GWAS imputed genotypes database. We also genotyped 127 ancestry informative markers [18] (listed in Supplementary Table 4) in the discovery cohort and examined the cohort for population structure by generating principal components in Eigenstrat implemented in Golden Helix (Bozeman, MT). Plotting the first and second most informative principal components revealed that the case and control subjects were of similar ancestry (Supplementary Figure 1), with the exception of two control subjects, who were excluded from analysis. None of the principal components were associated with PCOS status.

Statistical Analysis

Unpaired t tests and chi-square tests were used to compare clinical characteristics between case and control subjects; quantitative traits were log or square root transformed as appropriate to reduce non-normality. Data are presented as median (interquartile range).

In the discovery cohort, allelic analyses were conducted using chi-square tests to compare allele frequencies between case and control subjects for 295 SNPs. The same allelic analyses were carried out in the replication cohort (seven SNPs). We targeted for replication seven SNPs in the genes with the most SNPs associated with PCOS at $P < 0.05$. In the replication stage, SNPs were considered to be significantly associated with PCOS if the P value was < 0.007 ($0.05/7=0.007$).

To assess whether the effects of the seven associated SNPs were independent of body mass index (BMI), adjusted analyses were conducted with logistic regression; the dependent variable was PCOS status, and the independent variables were genotype (dominant model) and BMI. Furthermore, we directly evaluated the seven SNPs for an effect on BMI by conducting linear regression wherein BMI was the dependent variable. Meta-analyses were conducted on the logistic regression results of the discovery and replication cohorts by using the Metal program (http://www.sph.umich.edu/csg/abecasis/metal/). The discovery cohort sample size had good power (≥ 80%) to detect association of risk alleles of frequency ≥ 0.2 with PCOS at odds ratio (OR) ≥ 1.75 and moderate power (40%-60%) to detect association at OR 1.5; the power to detect association of rare risk alleles (frequency ≤ 0.1) with PCOS at OR ≤1.5 is limited [19].

Table 1 | Allelic association results in the discovery cohort.

Name	Gene	Chr	Alleles (major/minor)	Associated Allele	Case Frequency	Control Frequency	Chi-square value	*P* value
rs7997595[a]	*IRS2*[a]	13	C:G	C	0.881	0.795	11.281	0.001
rs10401628[a]	*INSR*[a]	19	G:A	G	0.919	0.856	8.298	0.004
rs2274490	*SORBS1*	10	T:C	C	0.430	0.334	7.428	0.006
rs7987237[a]	*IRS2*[a]	13	C:T	C	0.884	0.817	7.398	0.007
rs12459488[a]	*INSR*[a]	19	C:G	C	0.760	0.674	7.335	0.007
rs12971499[a]	*INSR*[a]	19	T:C	T	0.635	0.543	6.932	0.009
rs2252673[a]	*INSR*[a]	19	C:G	G	0.178	0.112	6.479	0.011
rs1865434[a]	*IRS2*[a]	13	A:G	A	0.875	0.814	5.875	0.015
rs4952834	*RHOQ*	2	G:C	C	0.338	0.261	5.500	0.019
rs3753242	*PRKCZ*	1	C:T	C	0.953	0.913	5.432	0.020
rs2267922	*PIK3R2*	19	C:G	C	0.539	0.460	5.002	0.025
rs3736328	*PIK3R2*	19	C:G	G	0.523	0.444	4.978	0.026
rs7207345	*PRKCA*	17	T:C	C	0.281	0.212	4.902	0.027
rs6510949	*INSR*	19	G:T	T	0.070	0.034	4.858	0.028
rs7641983	*PIK3CA*	3	C:T	C	0.832	0.772	4.610	0.032
rs7646409	*PIK3CA*	3	T:C	T	0.833	0.774	4.527	0.033
rs3094127	*FLOT1*	6	T:C	C	0.262	0.199	4.328	0.038
rs6020608	*PTPN1*	20	C:T	C	0.764	0.702	3.923	0.048
rs13420857	*RHOQ*	2	C:G	G	0.391	0.323	3.905	0.048
rs6443624	*PIK3CA*	3	C:A	C	0.771	0.711	3.848	0.050
rs7651265	*PIK3CA*	3	A:G	A	0.906	0.862	3.832	0.050

Single-nucleotide polymorphisms (*SNPs*) are displayed in order of significance. *P* values are for allelic association with polycystic ovary syndrome (no covariate adjustment). *Chr* Chromosome.
[a] *INSR* and *IRS2* SNPs that were selected for genotyping in the replication cohort.

Results

In the discovery cohort (275 case subjects and 171 control subjects [2 control subjects excluded based on principal component analysis]), we analyzed 295 SNPs for association with PCOS (Supplementary Table 3). SNPs with allelic association with PCOS ($P < 0.05$) are listed in Table 1. The gene with the largest number of SNPs significantly associated with PCOS in the discovery phase was *INSR*. *IRS2* also contained several associated SNPs at the top of the list. These observations, and the key importance of *INSR* and *IRS2* in the initial steps of insulin signaling, prompted us to attempt replication of the four *INSR* SNPs (rs12459488, rs12971499, rs2252673, rs10401628) and three *IRS2* SNPs (rs7997595, rs7987237, rs1865434) most associated with PCOS in the discovery cohort. These seven SNPs were not associated with BMI,

either in the entire discovery cohort (linear regression, adjusted for age and diagnosis) or in the PCOS women only (adjusted for age).

Table 2 | Allelic association of four *INSR* and three *IRS2* SNPs with PCOS in the replication cohort.

SNP	Gene	Case MAF	Control MAF	Chi-square value	*P* value
rs12459488	*INSR*	0.292	0.288	0.58	0.81
rs12971499	*INSR*	0.432	0.412	1.38	0.24
rs2252673	*INSR*	0.209	0.170	9.56	0.002
rs10401628	*INSR*	0.127	0.125	0.02	0.89
rs7997595	*IRS2*	0.173	0.166	0.186	0.666
rs7987237	*IRS2*	0.151	0.153	0.039	0.842
rs1865434	*IRS2*	0.153	0.153	0.014	0.907

Allelic *P* values are derived from chi-square testing. *MAF,* minor allele frequency; *PCOS,* polycystic ovary syndrome; *SNP,* single nucleotide polymorphism.

These seven SNPs were genotyped in the replication cohort consisting of 526 Caucasian PCOS women and 3,585 control subjects. Association results are presented in Table 2. SNP rs2252673 replicated association with PCOS ($P = 0.002$). Logistic regression analyses revealed that the association of rs2252673 was independent of BMI (Table 3). Carriers of the minor G allele at rs2252673 had increased odds of PCOS in both the discovery cohort (BMI-adjusted OR 2.06, 95% confidence interval [CI] 1.21-3.52; $P < 0.008$) and the replication cohort (BMI-adjusted OR 1.32, 95% CI 1.08-1.60; $P = 0.006$). P value meta-analysis of the two cohorts for this SNP revealed a combined $P = 0.00058$ (Table 3). The significant associations using the dominant model (Table 3) exhibited significance similar to the additive model. We also analyzed the association of the *INSR* SNP in the replication cohort with inclusion of only the 286 cases meeting the NIH 1990 definition of PCOS. With the halving of the number of cases in this analysis, statistical significance was diminished, but the odds of association with PCOS for rs2252673 (BMI-adjusted OR 1.34, 95% CI 1.03-1.74; $P = 0.03$) were identical to that in the entire replication cohort.

Table 3 | Logistic regression association of four *INSR* and three *IRS2* SNPs in the discovery and replication cohorts.

		Discovery cohort			Replication cohort			Meta-analysis
SNP	**Gene**	**OR**	**95% CI**	***P* value**	**OR**	**95% CI**	***P* value**	***P* value**
rs12459488	*INSR*	0.62	0.39-0.98	0.04	1.11	0.92-1.34	0.30	0.71
rs12971499	*INSR*	0.68	0.42-1.11	0.12	1.05	0.86-1.28	0.067	0.21
rs2252673	*INSR*	2.06	1.21-3.52	0.008	1.32	1.08-1.60	0.006	0.00058
rs10401628	*INSR*	0.45	0.25-0.79	0.006	1.00	0.80-1.25	0.99	0.40
rs7997595	*IRS2*	0.54	0.33-0.89	0.015	1.09	0.81-1.41	0.54	0.84
rs7987237	*IRS2*	0.54	0.33-0.90	0.019	1.03	0.79-1.35	0.84	0.58
rs1865434	*IRS2*	0.58	0.35-0.95	0.032	1.05	0.80-1.37	0.73	0.72

Logistic regression models included PCOS status as the dependent variable, and genotype (dominant model) and body mass index as independent variables. The *P* value meta-analysis was conducted on logistic regression models in the discovery and replication cohorts. *CI,* confidence interval; *OR,* odds ratio; *MAF,* minor allele frequency; *PCOS,* polycystic ovary syndrome; *SNP,* single nucleotide polymorphism.

Discussion

We have replicated association with PCOS of an *INSR* variant that predisposes to the syndrome. This result sheds light on the origins of PCOS as an inherited condition characterized by insulin resistance in many affected individuals.

Genetic variation in *INSR* may affect PCOS risk via an effect on insulin signaling. This was recognized as a possibility early in the field of PCOS genetics. Initial studies screening exons of the *INSR* gene (typically by polymerase chain reaction and single-stranded conformational polymorphisms) found missense/nonsense mutations in only isolated cases of women with PCOS-like phenotypes [5]. Such screens commonly identified only a silent C/T variation in codon His1058, leading many to conclude that *INSR* variation was not a major factor in PCOS.

In subsequent years, several genetic association studies of *INSR* in PCOS case-control cohorts were published, most of which focused on His1058 C/T variant; a recent meta-analysis of these studies revealed a non-significant trend (OR 1.28, 95% CI 0.88-1.85) for association of this silent variant with PCOS [5]. In Chinese subjects, a Cys1008Arg variant in exon 17 was associated with PCOS and lower insulin sensitivity index [20]; however, the frequency of this variant in other populations has not been determined (not available in the Single-Nucleotide Polymorphism Database [dbSNP]; http:// www.ncbi.nlm.nih.gov/SNP/). A Korean study that genotyped nine variants (initially found by sequencing) found association of a novel intron 21 variant with PCOS; this is another variant not yet listed in dSNP [21]. Therefore, we cannot determine whether our replicated *INSR* SNP is in linkage disequilibrium (LD) with either of these variants. The microsatellite D19S884, a replicated PCOS locus, was initially selected

based on proximity to *INSR*; however, it resides in a different gene (*FBN3*) > 850 kb from *INSR* and therefore likely represents an independent susceptibility locus [22].

The Korean study genotyped rs2252673 and found that carriers of the G allele (major allele in Koreans, minor allele in Caucasians) exhibited a trend to increased PCOS (OR 1.18, 95% CI 0.53-2.26) [21], consistent with our findings for this SNP. This suggests that this variant might influence PCOS risk across three continents (North America, Europe, Asia). If PCOS is an ancient disorder, as recently proposed, one would predict shared susceptibility variants across the globe [23]. Further studies of this variant are needed to establish whether rs2252673 is such a universal risk factor for PCOS.

How the replicated variant affects *INSR* function and/or expression is unknown. SNP rs2252673 is not in LD ($r^2 < 0.8$) with any other SNP in the *INSR* gene in the HapMap Caucasian database [24]. As an intronic variant, it may be in LD with unknown functional (or not yet genotyped in HapMap) variants elsewhere in the gene. Sequencing may identify coding variants in LD with the associated variant. Alternatively, the variant itself may influence *INSR* transcription or mRNA splicing. In the Transcription Element Search System database (http://www.cbil.upenn.edu/tess/), rs2252673 results in alternative possible transcription factor-binding sites. Two forms of the insulin receptor arise from alternative splicing of exon 11 (excluded from the A isoform and included in the B isoform). The A isoform, expressed mainly during fetal development and the adult brain, has low affinity for insulin but high affinity for insulin-like growth factor 2; the B isoform has high affinity for insulin and is expressed in insulin-responsive tissues such as muscle, liver, and adipose [25]. Whether the replicated *INSR* SNP affects INSR splicing and/or transcription must be determined by biochemical studies. Tissue-specific effects (e.g., muscle, liver, ovary) could also be evaluated.

The large size of the *INSR* gene (> 180 kb) has made its investigation as a candidate gene challenging. Earlier studies only looked at one or a few variants in the gene, leading to incomplete coverage of its genetic variation. This is the first study that tagged the entire *INSR* region using information from HapMap [24]. This was critical to allow the discovery phase to detect *INSR* variants potentially associated with PCOS. Efforts to replicate genetic association results have been infrequent in the field of PCOS genetics. Examples include replication studies of *HSD17B5* [26], *HSD17B6* [27], *CYP11A1* [28], the insulin gene [29], and others [22,30]. A key factor in our successful replication was that the replication cohort was substantially larger than the discovery cohort. Genetic effect size is typically overestimated in initial studies, a phenomenon known as the "winner's curse" [31].

Therefore, replication cohorts should be larger than discovery cohorts to have sufficient power to detect the association at a lower effect size. This proved to be true in the present study, wherein the odds for association with PCOS were higher in the discovery cohort than in the replication cohort. Another possible explanation for the different effect sizes is that PCOS was excluded from the discovery cohort control subjects (increasing power for discovery), whereas the replication cohort control subjects consisted of an unselected sample

representative of the general population. The latter large (> 3,500 subjects) population-based sample may indicate broad generalizability of our findings.

Ideally, both the discovery cohort and replication cohort PCOS cases would have been diagnosed by using the same criteria. This was not the case, because these were preexisting cohorts. Fortunately, this did not hinder our ability to replicate at least one SNP association with PCOS. We ruled out BMI disparity between case and control subjects as a cause of spurious SNP association with PCOS by directly ruling out SNP association with BMI and by adjusting for BMI as a covariate.

We did not adjust the results in the discovery cohort for multiple testing (e.g., Bonferroni correction), because we planned to follow-up with a replication effort. Replication is the optimal solution to multiple testing. We corrected for multiple testing in the replication phase. Given their central roles in insulin signaling and preponderance of significant SNPs in the discovery phase (Table 1), we first attempted to replicate variants in *INSR* and *IRS2*. In the future, we will attempt to replicate SNPs in other loci associated with PCOS in the discovery cohort. We hope that this publication inspires other investigators to also examine these loci in their cohorts.

In conclusion, in a survey of 39 genes coding for the central components of metabolic insulin-signaling pathways, we identified SNPs in the *INSR* and *IRS2* genes that were associated with PCOS. One of these SNPs replicated association with PCOS in a larger independent cohort. These data should reinvigorate interest in INSR as a genetic susceptibility factor for PCOS.

References

1. DeUgarte CM, Bartolucci AA, Azziz R. Prevalence of insulin resistance in the polycystic ovary syndrome using the homeostasis model assessment. *Fertil Steril*. 2005; May; 83: 1454-60.
2. Palomba S, Falbo A, Zullo F, et al. Evidence-based and potential benefits of metformin in the polycystic ovary syndrome: a comprehensive review. *Endocr Rev*. 2009; Feb; 30: 1-50.
3. Goodarzi MO. Looking for polycystic ovary syndrome genes: rational and best strategy. *Semin Reprod Med*. 2008; Jan; 26: 5-13.
4. Legro RS, Bentley-Lewis R, Driscoll D, et al. Insulin resistance in the sisters of women with polycystic ovary syndrome: association with hyperandrogenemia rather than menstrual irregularity. *J Clin Endocrinol Metab*. 2002; May; 87: 2128-33.
5. Ioannidis A, Ikonomi E, Dimou NL, et al. Polymorphisms of the insulin receptor and the insulin receptor substrates genes in polycystic ovary syndrome: a Mendelian randomization meta-analysis. *Mol Genet Metab*. 2010; Feb; 99: 174-83.
6. Lizcano JM, Alessi DR. The insulin signalling pathway. *Curr Biol*. 2002; Apr 2; 12: R236-8.
7. Baumann CA, Saltiel AR. Spatial compartmentalization of signal transduction in insulin action. *Bioessays*. 2001; Mar; 23: 215-22.
8. Simoni M, Tempfer CB, Destenaves B, et al. Functional genetic polymorphisms and female reproductive disorders: Part I: Polycystic ovary syndrome and ovarian response. *Hum Reprod Update*. 2008; Sep-Oct; 14: 459-84.
9. Doble BW, Woodgett JR. GSK-3: tricks of the trade for a multi-tasking kinase. *J Cell Sci*. 2003; Apr 1; 116: 1175-86.
10. Chang W, Goodarzi MO, Williams H, et al. Adipocytes from women with polycystic ovary syndrome demonstrate altered phosphorylation and activity of glycogen synthase kinase 3. *Fertil Steril*. 2008; Dec; 90: 2291-7.
11. Goodarzi MO, Antoine HJ, Pall M, et al. Preliminary evidence of glycogen synthase kinase 3 beta as a genetic determinant of polycystic ovary syndrome. *Fertil Steril*. 2007; Jun; 87: 1473-6.
12. Goodarzi MO, Jones MR, Chen YD, et al. First evidence of genetic association between AKT2 and polycystic ovary syndrome. *Diabetes Care*. 2008; Dec; 31: 2284-7.
13. Zawadzki J., Dunaif A. Diagnostic criteria for polycystic ovary syndrome: toward a rational approach. Boston: Blackwell Scientific; 1992. 377-84 p.
14. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
15. ESHRE ASRM Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004; Jan; 81: 19-25.
16. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol*. 2007; 22: 819-29.
17. Gunderson KL, Steemers FJ, Ren H, et al. Whole-genome genotyping. *Methods Enzymol*. 2006; 410: 359-76.
18. Kosoy R, Nassir R, Tian C, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat*. 2009; Jan; 30: 69-78.
19. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003; Jan; 19: 149-50.
20. Jin L, Zhu XM, Luo Q, et al. A novel SNP at exon 17 of INSR is associated with decreased insulin sensitivity in Chinese women with PCOS. *Mol Hum Reprod*. 2006; Mar; 12: 151-5.
21. Lee EJ, Oh B, Lee JY, et al. A novel single nucleotide polymorphism of INSR gene for polycystic ovary syndrome. *Fertil Steril*. 2008; May; 89: 1213-20.

22. Stewart DR, Dombroski BA, Urbanek M, et al. Fine mapping of genetic susceptibility to polycystic ovary syndrome on chromosome 19p13.2 and tests for regulatory activity. *J Clin Endocrinol Metab*. 2006; Oct; 91: 4112-7.

23. Azziz R, Dumesic DA, Goodarzi MO. Polycystic ovary syndrome: an ancient disorder? *Fertil Steril*. 2011; Apr; 95: 1544-8.

24. The International HapMap Project. *Nature*. 2003; Dec 18; 426: 789-96.

25. Belfiore A, Frasca F, Pandini G, et al. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocr Rev*. 2009; Oct; 30: 586-623.

26. Goodarzi MO, Jones MR, Antoine HJ, et al. Nonreplication of the type 5 17beta-hydroxysteroid dehydrogenase gene association with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2008; Jan; 93: 300-3.

27. Jones MR, Mathur R, Cui J, et al. Independent confirmation of association between metabolic phenotypes of polycystic ovary syndrome and variation in the type 6 17beta-hydroxysteroid dehydrogenase gene. *J Clin Endocrinol Metab*. 2009; Dec; 94: 5034-8.

28. Gaasenbeek M, Powell BL, Sovio U, et al. Large-scale analysis of the relationship between CYP11A promoter variation, polycystic ovarian syndrome, and serum testosterone. *J Clin Endocrinol Metab*. 2004; May; 89: 2408-13.

29. Powell BL, Haddad L, Bennett A, et al. Analysis of multiple data sets reveals no association between the insulin gene variable number tandem repeat element and polycystic ovary syndrome or related traits. *J Clin Endocrinol Metab*. 2005; May; 90: 2988-93.

30. Ewens KG, Stewart DR, Ankener W, et al. Family-based analysis of candidate genes for polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2010; May; 95: 2306-15.

31. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology*. 2008; Sep; 19: 649-51; discussion 57-8.

Supplementary Tables

Supplementary Table 1 | Clinical characteristics of the discovery cohort.

	Control n=171	PCOS n=275
Age (yr)	34.0 (15.0)	27.6 (11.5)[a]
BMI (kg/m^2)	24.3 (6.3)	34.7 (13.0)[a]
WHR	0.78 (0.08)	0.83 (0.11)[a]
mFG score	0 (0)	7.0 (5.8)[a]
Hirsute (%)	0	74.2[a]
Total testosterone (ng/dl)	42.0 (27.3)	80.0 (31.0)[a]
Free testosterone (pg/ml)	0.36 (0.26)	0.84 (0.46)[a]
DHEAS (ng/ml)	966 (723)	2094 (1702)[a]
SHBG (nmol/l)[b]	210.0 (120.0)	150.0 (62.5)[a]
Insulin (µIU/ml)	6.9 (6.3)	18.0 (18.0)[a]
Glucose (mg/dl)	86.0 (10.8)	86.0 (12.8)
HOMA-IR	0.95 (0.81)	2.25 (1.92)[a]
HOMA-%B	106.6 (61.1)	175.0 (99.6)[a]

Data are presented as median (interquartile range). *BMI,* body mass index; *WHR,* waist-to-hip ratio; *mFG,* modified Ferriman-Gallwey hirsutism score; *HOMA-IR,* insulin resistance estimated by homeostatic model assessment; *HOMA-%B,* beta-cell function estimated by homeostatic model assessment; *PCOS,* polycystic ovary syndrome.

[a] $P < .001$ compared with control group.

[b] SHBG activity was measured by competitive binding analysis using Sephadex G-25 and [3H]T as the ligand; this assay gives values of ~100-300 nmol/L in normal adult women.

Supplementary Table 2 | Gene statistics and single-nucleotide polymorphism (SNP) coverage.

Gene	Gene name [alternate name]	Chr	Gene size (bp)	MAF of tagged SNPs	SNPs in region	SNPs tagged	Coverage (%)	Mean r^2	No. of SNPs genotyped
AKAP11	A kinase (PRKA) anchor protein 11 [AKAP220]	13	51,113	0.05	43	42	97.7%	0.98	10
AKT1	v-akt murine thymoma viral oncogene homologue 1	14	26,391	0.05	13	10	76.9%	0.97	4
CBL	Cas-Br-M (murine) ecotropic retroviral transforming sequence	11	101,868	0.05	35	34	97.1%	0.96	10
CRK	v-crk sarcoma virus CT10 oncogene homologue	17	34,084	0.05	11	11	100.0%	0.99	3
FLOT1	flotillin 1	6	14,94	0.05	7	6	85.7%	1.00	4
FLOT2	flotillin 2	17	18,357	0.05	8	7	87.5%	0.99	4
FRAT1	frequently rearranged in advanced T-cell lymphomas	10	2,65	0.05	6	5	83.3%	0.91	2
FRAT2	frequently rearranged in advanced T-cell lymphomas 2	10	2,203	0.05	3	2	66.7%	1.00	3
FYN	FYN oncogene related to SRC, FGR, YES	6	212,14	0.2	70	42	60.0%	0.95	11
GNA12	guanine nucleotide–binding protein (G protein) alpha 12	7	116,216	0.05	126	80	63.5%	0.93	6
GNA13	guanine nucleotide–binding protein (G protein), alpha 13	17	45,92	0.001	9	9	100.0%	1.00	5
GSK3A	glycogen synthase kinase 3 alpha	19	12,398	0.001	7	5	71.4%	0.98	4
GYS1	glycogen synthase 1 (muscle)	19	25,166	0.05	10	6	60.0%	0.97	5
GYS2	glycogen synthase 2 (liver)	12	68,658	0.05	98	9	9.2%	0.95	6
INPPL1	inositol polyphosphate phosphatase–like 1 [SHIP2]	11	14,323	0.05	4	4	100.0%	0.98	3
INSR	insulin receptor	19	181,745	0.05	146	82	56.2%	0.95	35
IRS1	insulin receptor substrate 1	2	64,537	0.01	53	17	32.1%	1.00	3
IRS2	insulin receptor substrate 2	13	32,73	0.05	41	39	95.1%	0.95	12
PDPK1	3-phosphoinositide–dependent protein kinase-1 [PDK1]	16	65,218	0.05	3	3	100.0%	1.00	5
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide [P110A]	3	86,184	0.05	39	38	97.4%	0.97	13
PIK3CB	phosphoinositide-3-kinase, catalytic, beta polypeptide [P110B]	3	103,952	0.001	53	8	15.1%	0.95	5

Supplementary Table 2 | *Continued*

Gene	Gene name [alternate name]	Chr	Gene size (bp)	MAF of tagged SNPs	SNPs in region	SNPs tagged	Coverage (%)	Mean r^2	No. of SNPs genotyped
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha) [P85A]	5	75,185	0.05	70	50	71.4%	0.97	21
PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta) [P85B]	19	17,328	0.05	5	5	100.0%	1.00	4
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme [PP1A]	11	3,724	0.05	6	5	83.3%	0.97	2
PPP1R3A	protein phosphatase 1, regulatory (inhibitor) subunit 3A	7	44,678	0.05	21	20	95.2%	0.97	7
PRKCA	protein kinase C, alpha	17	507,936	0.2	257	135	52.5%	0.94	32
PRKCI	protein kinase C, iota	3	83,539	0.05	21	21	100.0%	0.97	6
PRKCZ	protein kinase C, zeta	1	134,923	0.05	69	63	91.3%	0.94	10
PTEN	phosphatase and tensin homolog	10	105,336	0.05	25	25	100.0%	0.99	11
PTK2B	PTK2B protein tyrosine kinase 2 beta [PYK2]	8	147,904	0.2	142	138	97.2%	0.96	14
PTPN1	protein tyrosine phosphatase, non-receptor type 1 [PTP1B]	20	74,193	0.05	50	49	98.0%	0.94	10
RAPGEF1	Rap guanine nucleotide exchange factor (GEF) 1 [C3G]	9	160,767	0.05	172	108	62.8%	0.93	12
RHOQ	ras homologue gene family, member Q [TC10]	2	41,958	0.05	23	9	39.1%	0.95	4
RPS6KA1	ribosomal protein S6 kinase, 90 kd, polypeptide 1 [p90-RSK 1]	1	45,271	0.05	33	4	12.1%	1.00	4
RPS6KB1	ribosomal protein S69 kinase, 70 kd, polypeptide 1 [p70-S6K]	17	57,341	0.05	21	20	95.2%	1.00	9
SH2B2	SH2B adaptor protein 2 [APS]	7	33,731	0.05	13	12	92.3%	0.99	9
SLC2A4	solute carrier family 2 (facilitated glucose transporter), member 4 [GLUT4]	17	6,312	0.05	11	4	36.4%	1.00	4
SORBS1	sorbin and SH3 domain–containing 1 [CAP]	10	249,641	0.2	160	115	71.9%	0.95	27
TP53	tumor protein p53	17	19,197	0.05	9	4	44.4%	0.91	2

Chr, chromosome; *MAF,* minor allele frequency.

Supplementary Table 3 | 341 single-nucleotide polymorphisms (SNPs) initially selected for genotyping in the discovery cohort.

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs7997595	*IRS2*	13	0.152	C:G	C	0.881	0.795	11.281	.001
rs10401628	*INSR*	19	0.105	G:A	G	0.919	0.856	8.298	.004
rs2274490	*SORBS1*	10	0.392	T:C	C	0.430	0.334	7.428	.006
rs7987237	*IRS2*	13	0.142	C:T	C	0.884	0.817	7.398	.007
rs12459488	*INSR*	19	0.273	C:G	C	0.760	0.674	7.335	.007
rs12971499	*INSR*	19	0.400	T:C	T	0.635	0.543	6.932	.009
rs2252673	*INSR*	19	0.153	C:G	G	0.178	0.112	6.479	.011
rs1865434	*IRS2*	13	0.149	A:G	A	0.875	0.814	5.875	.015
rs4952834	*RHOQ*	2	0.308	G:C	C	0.338	0.261	5.500	.019
rs3753242	*PRKCZ*	1	0.062	C:T	C	0.953	0.913	5.432	.020
rs2267922	*PIK3R2*	19	0.492	C:G	C	0.539	0.460	5.002	.025
rs3736328	*PIK3R2*	19	0.493	C:G	G	0.523	0.444	4.978	.026
rs7207345	*PRKCA*	17	0.255	T:C	C	0.281	0.212	4.902	.027
rs6510949	*INSR*	19	0.056	G:T	T	0.070	0.034	4.858	.028
rs7641983	*PIK3CA*	3	0.191	C:T	C	0.832	0.772	4.610	.032
rs7646409	*PIK3CA*	3	0.190	T:C	T	0.833	0.774	4.527	.033
rs3094127	*FLOT1*	6	0.237	T:C	C	0.262	0.199	4.328	.038
rs6020608	*PTPN1*	20	0.260	C:T	C	0.764	0.702	3.923	.048
rs13420857	*RHOQ*	2	0.365	C:G	G	0.391	0.323	3.905	.048
rs6443624	*PIK3CA*	3	0.252	C:A	C	0.771	0.711	3.848	.050
rs7651265	*PIK3CA*	3	0.111	A:G	A	0.906	0.862	3.832	.050
rs2736627	*PTEN*	10	0.119	T:C	T	0.898	0.854	3.725	.054

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs932420	*PTPN1*	20	0.495	T:C	T	0.531	0.463	3.712	.054
rs11079656	*PRKCA*	17	0.315	A:G	A	0.709	0.646	3.636	.057
rs2906713	*SH2B2*	7	0.142	G:C	G	0.876	0.829	3.625	.057
rs7225452	*PRKCA*	17	0.332	T:C	T	0.692	0.630	3.392	.066
rs752993	*PTK2B*	8	0.448	C:T	T	0.473	0.409	3.188	.074
rs17375610	*PTK2B*	8	0.423	G:C	C	0.447	0.385	3.130	.077
rs10205	*FLOT2*	17	0.284	C:T	C	0.737	0.681	3.031	.082
rs2305268	*PIK3CB*	3	0.016	C:T	T	0.021	0.006	3.005	.083
rs919275	*INSR*	19	0.435	A:G	A	0.588	0.528	2.889	.089
rs7914	*CBL*	11	0.220	C:T	C	0.799	0.750	2.736	.098
rs2042902	*INSR*	19	0.499	A:G	G	0.521	0.463	2.729	.099
rs4804404	*INSR*	19	0.141	A:C	C	0.156	0.116	2.689	.101
rs734589	*PTPN1*	20	0.495	C:T	T	0.518	0.459	2.671	.102
rs712842	*SH2B2*	7	0.318	C:T	T	0.339	0.284	2.662	.103
rs1952422	*RAPGEF1*	9	0.458	G:T	T	0.480	0.422	2.656	.103
rs8104760	*GYS1*	19	0.044	C:T	C	0.965	0.941	2.652	.103
rs9387024	*FYN*	6	0.483	C:G	C	0.539	0.481	2.646	.104
rs7254921	*INSR*	19	0.482	T:C	C	0.504	0.447	2.545	.111
rs706716	*PIK3R1*	5	0.239	C:T	C	0.779	0.733	2.340	.126
rs1044364	*Irs2*	13	0.070	A:G	G	0.080	0.053	2.318	.128
rs17132719	*GNA12*	7	0.081	A:G	G	0.092	0.062	2.283	.131
rs7217337	*RPS6KB1*	17	0.082	C:T	C	0.929	0.900	2.261	.133

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs13041704	*PTPN1*	20	0.363	A:C	A	0.656	0.606	2.193	.139
rs750539	*PTK2B*	8	0.401	T:C	C	0.421	0.370	2.192	.139
rs9262155	*FLOT1*	6	0.083	A:G	G	0.094	0.065	2.121	.145
rs1234225	*PTEN*	10	0.350	G:A	A	0.369	0.320	2.109	.147
rs532678	*PTEN*	10	0.350	C:T	T	0.369	0.320	2.109	.147
rs262688	*PRKCZ*	1	0.283	A:C	C	0.301	0.255	2.073	.150
rs11135993	*PTK2B*	8	0.475	T:C	T	0.545	0.494	2.068	.150
rs12579540	*GYS2*	12	0.249	C:G	C	0.768	0.724	2.042	.153
rs10226417	*PPP1R3A*	7	0.041	T:C	T	0.967	0.947	1.994	.158
rs10896172	*PPP1CA*	11	0.060	T:C	T	0.949	0.925	1.979	.160
rs11618950	*Irs2*	13	0.192	G:A	A	0.207	0.168	1.972	.160
rs8081659	*FLOT2*	17	0.496	C:T	T	0.516	0.466	1.960	.162
rs706862	*FYN*	6	0.303	G:A	G	0.715	0.669	1.951	.163
rs7258382	*INSR*	19	0.173	T:C	T	0.842	0.804	1.941	.164
rs1862162	*PIK3R1*	5	0.264	T:C	T	0.753	0.709	1.921	.166
rs1536558	*SORBS1*	10	0.253	G:T	T	0.270	0.227	1.918	.166
rs1549616	*INSR*	19	0.083	G:A	G	0.928	0.901	1.915	.166
rs180531	*RPS6KB1*	17	0.254	T:C	C	0.271	0.228	1.879	.171
rs7080545	*SORBS1*	10	0.281	G:C	C	0.298	0.255	1.838	.175
rs10853078	*PRKCA*	17	0.275	G:A	A	0.291	0.248	1.798	.180
rs7641889	*PIK3CA*	3	0.070	C:T	C	0.939	0.915	1.791	.181
rs2677760	*PIK3CA*	3	0.448	T:C	C	0.467	0.419	1.781	.182

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs7248939	*INSR*	19	0.305	G:A	G	0.712	0.669	1.717	.190
rs7919271	*SORBS1*	10	0.406	G:A	G	0.611	0.566	1.705	.192
rs2974945	*PPP1R3A*	7	0.153	G:C	G	0.859	0.826	1.686	.194
rs803090	*SH2B2*	7	0.331	G:A	A	0.348	0.304	1.675	.196
rs3730084	*PIK3R1*	5	0.020	C:T	T	0.025	0.012	1.665	.197
rs12416299	*SORBS1*	10	0.180	C:T	C	0.834	0.799	1.654	.198
rs11986969	*PTK2B*	8	0.245	G:A	A	0.260	0.220	1.650	.199
rs4443935	*RPS6KA1*	1	0.296	G:A	A	0.312	0.270	1.639	.201
rs2078153	*PRKCA*	17	0.255	G:C	G	0.760	0.720	1.603	.206
rs3787345	*PTPN1*	20	0.413	T:C	T	0.604	0.559	1.582	.209
rs9303510	*PRKCA*	17	0.237	G:A	G	0.777	0.741	1.469	.226
rs10072475	*PIK3R1*	5	0.108	A:G	A	0.902	0.876	1.449	.229
rs1005273	*PDPK1*	16	0.448	C:T	T	0.465	0.422	1.442	.230
rs10882583	*SORBS1*	10	0.332	G:A	A	0.348	0.307	1.440	.230
rs208346	*GNA12*	7	0.408	G:T	T	0.424	0.382	1.433	.231
rs2302975	*PIK3R1*	5	0.461	A:G	A	0.556	0.511	1.430	.232
rs2276047	*INPPL1*	11	0.195	T:C	T	0.818	0.784	1.386	.239
rs2344706	*FYN*	6	0.498	A:T	T	0.514	0.472	1.376	.241
rs12659907	*PIK3R1*	5	0.114	G:A	G	0.896	0.870	1.372	.242
rs9903921	*PRKCA*	17	0.369	C:T	T	0.385	0.345	1.361	.243
rs10408374	*INSR*	19	0.106	C:T	T	0.115	0.090	1.327	.249
rs754204	*IRS2*	13	0.494	C:T	T	0.510	0.469	1.318	.251

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs3739497	*RAPGEF1*	9	0.361	C:T	C	0.654	0.615	1.314	.252
rs11597175	*SORBS1*	10	0.276	G:T	G	0.738	0.702	1.313	.252
rs3783192	*AKAP11*	13	0.358	T:C	T	0.657	0.618	1.296	.255
rs3787348	*PTPN1*	20	0.431	G:T	T	0.447	0.407	1.287	.257
rs7216404	*FLOT2*	17	0.285	C:T	C	0.729	0.693	1.254	.263
rs16822573	*IRS1*	2	0.053	A:C	A	0.953	0.936	1.244	.265
rs9515119	*IRS2*	13	0.317	A:C	C	0.331	0.295	1.203	.273
rs533480	*SORBS1*	10	0.305	G:A	A	0.318	0.283	1.193	.275
rs6784397	*PRKCI*	3	0.207	C:A	C	0.805	0.773	1.185	.276
rs7218480	*PRKCA*	17	0.429	T:C	C	0.443	0.406	1.138	.286
rs9913645	*PRKCA*	17	0.181	T:C	T	0.830	0.801	1.109	.292
rs481781	*PRKCI*	3	0.431	T:C	C	0.446	0.407	1.108	.292
rs2514256	*PPP1CA*	11	0.087	G:A	G	0.922	0.901	1.096	.295
rs9913908	*PRKCA*	17	0.231	T:C	T	0.781	0.750	1.083	.298
rs8103483	*INSR*	19	0.488	T:C	C	0.502	0.465	1.074	.300
rs11153317	*FYN*	6	0.388	G:A	A	0.402	0.366	1.071	.301
rs1887228	*SORBS1*	10	0.472	T:C	C	0.486	0.450	1.043	.307
rs2299939	*PTEN*	10	0.171	C:A	C	0.840	0.812	1.040	.308
rs17562384	*PTEN*	10	0.173	T:C	T	0.838	0.811	1.034	.309
rs7216019	*PRKCA*	17	0.361	G:A	G	0.652	0.618	1.010	.315
rs171649	*PIK3R1*	5	0.366	G:A	A	0.379	0.345	0.996	.318
rs17689224	*PRKCA*	17	0.224	C:T	C	0.787	0.758	0.979	.323

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs7974959	*GYS2*	12	0.126	T:C	T	0.883	0.860	0.915	.339
rs3848428	*PRKCA*	17	0.331	T:C	T	0.681	0.649	0.913	.339
rs1234219	*PTEN*	10	0.070	T:C	C	0.076	0.059	0.900	.343
rs10882899	*FRAT2*	10	0.433	C:G	C	0.580	0.547	0.884	.347
rs1367088	*PTK2B*	8	0.401	T:C	C	0.414	0.381	0.882	.348
rs4917766	*FRAT2*	10	0.433	A:G	A	0.580	0.547	0.866	.352
rs12705918	*PPP1R3A*	7	0.458	A:G	A	0.555	0.522	0.864	.353
rs1607237	*PIK3CA*	3	0.364	T:C	C	0.376	0.345	0.859	.354
rs406985	*RPS6KA1*	1	0.403	T:C	T	0.610	0.578	0.849	.357
rs4804428	*INSR*	19	0.180	T:C	T	0.829	0.804	0.839	.360
rs2494743	*AKT1*	14	0.080	T:C	T	0.927	0.909	0.835	.361
rs11655449	*CRK*	17	0.270	C:A	C	0.741	0.712	0.821	.365
rs7777484	*GNA12*	7	0.403	A:G	A	0.609	0.578	0.800	.371
rs3745550	*INSR*	19	0.189	G:A	A	0.199	0.174	0.797	.372
rs4938642	*CBL*	11	0.089	G:C	G	0.918	0.900	0.785	.376
rs712843	*SH2B2*	7	0.042	T:A	T	0.963	0.950	0.778	.378
rs17521782	*AKAP11*	13	0.161	T:C	C	0.170	0.147	0.774	.379
rs955760	*SORBS1*	10	0.422	A:G	G	0.434	0.403	0.750	.387
rs4938637	*CBL*	11	0.089	G:A	G	0.918	0.901	0.736	.391
rs2320273	*RAPGEF1*	9	0.297	C:T	C	0.714	0.686	0.709	.400
rs10509692	*SORBS1*	10	0.435	C:G	C	0.576	0.547	0.704	.401
rs4534095	*PTK2B*	8	0.459	T:C	C	0.471	0.441	0.703	.402

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs8112883	*INSR*	19	0.299	G:T	T	0.310	0.283	0.696	.404
rs10426094	*INSR*	19	0.239	C:T	T	0.248	0.224	0.650	.420
rs7981705	*IRS2*	13	0.171	C:T	T	0.180	0.158	0.631	.427
rs1536442	*SORBS1*	10	0.452	G:T	T	0.463	0.435	0.631	.427
rs2511837	*CBL*	11	0.452	T:C	C	0.463	0.435	0.616	.433
rs1801278	*IRS1*	2	0.077	A:C	C	0.082	0.068	0.607	.436
rs1375620	*AKAP11*	13	0.326	C:T	C	0.684	0.658	0.604	.437
rs238314	*AKAP11*	13	0.105	C:T	C	0.902	0.885	0.600	.439
rs6512652	*PTPN1*	20	0.058	G:A	A	0.062	0.050	0.598	.439
rs9916520	*PRKCA*	17	0.361	T:C	C	0.371	0.345	0.596	.440
rs6126034	*PTPN1*	20	0.094	A:T	T	0.100	0.084	0.579	.447
rs2459994	*PRKCZ*	1	0.186	C:T	C	0.822	0.801	0.577	.447
rs2296950	*RAPGEF1*	9	0.330	A:G	G	0.340	0.314	0.571	.450
rs7640662	*PIK3CA*	3	0.147	C:G	G	0.154	0.135	0.569	.451
rs11841502	*IRS2*	13	0.324	G:A	G	0.686	0.660	0.567	.451
rs7323055	*AKAP11*	13	0.322	C:G	C	0.688	0.662	0.564	.453
rs943851	*RAPGEF1*	9	0.209	T:C	C	0.217	0.196	0.563	.453
rs2024321	*PRKCA*	17	0.363	T:C	T	0.646	0.621	0.550	.458
rs10513055	*PIK3CB*	3	0.215	A:C	A	0.793	0.772	0.534	.465
rs3744405	*SLC2A4*	17	0.472	A:G	G	0.482	0.457	0.532	.466
rs11243471	*RAPGEF1*	9	0.077	G:A	A	0.082	0.068	0.524	.469
rs34308	*PIK3R1*	5	0.165	G:A	G	0.842	0.823	0.507	.477

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2278319	*PTK2B*	8	0.240	C:T	T	0.248	0.227	0.492	.483
rs2241652	*PTK2B*	8	0.148	A:G	A	0.859	0.842	0.464	.496
rs9594724	*AKAP11*	13	0.212	A:G	A	0.796	0.776	0.458	.498
rs10751199	*INPPL1*	11	0.215	A:G	A	0.793	0.773	0.454	.500
rs2418003	*GYS2*	12	0.234	T:C	T	0.773	0.753	0.453	.501
rs7218425	*PRKCA*	17	0.248	T:C	T	0.760	0.739	0.451	.502
rs17318918	*PIK3R1*	5	0.112	T:C	C	0.118	0.102	0.448	.503
rs9886	*INPPL1*	11	0.056	G:C	C	0.061	0.050	0.438	.508
rs8110116	*INSR*	19	0.112	C:T	T	0.117	0.102	0.431	.511
rs10509690	*SORBS1*	10	0.230	G:A	G	0.777	0.758	0.428	.513
rs10512513	*PRKCA*	17	0.464	G:T	G	0.545	0.522	0.427	.513
rs929457	*PDPK1*	16	0.135	T:C	C	0.141	0.125	0.413	.521
rs2860172	*INSR*	19	0.204	C:A	A	0.211	0.193	0.412	.521
rs7245562	*INSR*	19	0.066	C:T	T	0.070	0.059	0.410	.522
rs1327200	*FYN*	6	0.376	C:T	T	0.384	0.362	0.385	.535
rs890862	*INSR*	19	0.301	C:T	T	0.309	0.289	0.383	.536
rs8071475	*RPS6KB1*	17	0.252	T:C	T	0.755	0.736	0.373	.542
rs10500204	*INSR*	19	0.301	A:C	C	0.309	0.289	0.367	.544
rs1083	*CRK*	17	0.080	T:C	C	0.084	0.072	0.364	.546
rs7005183	*PTK2B*	8	0.428	A:G	A	0.580	0.559	0.359	.549
rs3107151	*PRKCZ*	1	0.400	G:T	T	0.408	0.388	0.352	.553
rs744313	*RAPGEF1*	9	0.069	G:A	G	0.936	0.925	0.343	.558

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs12980441	*GYS1*	19	0.388	C:T	T	0.396	0.376	0.343	.558
rs11656396	*GNA13*	17	0.101	C:T	T	0.105	0.093	0.330	.566
rs7219495	*PRKCA*	17	0.371	G:A	A	0.379	0.359	0.322	.571
rs1292034	*RPS6KB1*	17	0.457	C:T	C	0.551	0.531	0.310	.578
rs2322719	*PTK2B*	8	0.451	C:T	C	0.557	0.537	0.300	.584
rs10770837	*GYS2*	12	0.349	G:A	G	0.658	0.640	0.296	.586
rs11668751	*INSR*	19	0.275	T:G	G	0.281	0.264	0.296	.586
rs11243436	*RAPGEF1*	9	0.292	T:C	C	0.299	0.281	0.294	.588
rs2699905	*PIK3CA*	3	0.257	G:A	A	0.264	0.247	0.291	.590
rs9900205	*PRKCA*	17	0.209	A:T	A	0.797	0.781	0.291	.590
rs3730082	*PIK3R1*	5	0.205	G:A	G	0.801	0.786	0.275	.600
rs2699887	*PIK3CA*	3	0.257	G:A	A	0.263	0.247	0.265	.607
rs4330616	*PPP1R3A*	7	0.327	A:G	A	0.680	0.662	0.264	.607
rs4918940	*SORBS1*	10	0.248	G:A	G	0.758	0.742	0.257	.612
rs3849448	*PRKCI*	3	0.072	T:C	T	0.932	0.922	0.255	.614
rs4938638	*CBL*	11	0.474	A:G	G	0.480	0.463	0.249	.618
rs11963612	*FYN*	6	0.442	T:C	C	0.449	0.432	0.241	.624
rs2865084	*PIK3CA*	3	0.072	T:A	T	0.931	0.922	0.240	.625
rs7069574	*SORBS1*	10	0.266	A:G	A	0.740	0.725	0.234	.628
rs2322606	*PTK2B*	8	0.157	G:A	G	0.848	0.835	0.224	.636
rs3786680	*INSR*	19	0.100	G:A	G	0.904	0.894	0.216	.642
rs11667110	*INSR*	19	0.311	C:G	C	0.695	0.680	0.213	.644

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs6510960	*INSR*	19	0.118	A:G	A	0.886	0.876	0.209	.647
rs1001746	*RHOQ*	2	0.235	T:C	C	0.240	0.227	0.201	.654
rs12493155	*PIK3CB*	3	0.457	C:T	T	0.463	0.447	0.196	.658
rs11079667	*PRKCA*	17	0.438	G:A	G	0.568	0.553	0.195	.659
rs180535	*RPS6KB1*	17	0.165	A:G	G	0.170	0.158	0.191	.663
rs180522	*RPS6KB1*	17	0.165	T:C	C	0.170	0.158	0.191	.663
rs2270938	*GYS1*	19	0.381	T:A	A	0.387	0.372	0.184	.668
rs222842	*SLC2A4*	17	0.270	T:C	T	0.735	0.722	0.180	.672
rs12943037	*GNA13*	17	0.159	A:G	G	0.163	0.152	0.165	.684
rs9901261	*PRKCA*	17	0.294	G:C	C	0.299	0.286	0.164	.686
rs17254521	*INSR*	19	0.466	A:G	A	0.539	0.525	0.161	.689
rs1043526	*PIK3R1*	5	0.152	A:G	A	0.852	0.842	0.152	.697
rs706895	*FYN*	6	0.234	T:C	C	0.238	0.227	0.148	.701
rs17317833	*PIK3R1*	5	0.124	G:A	A	0.127	0.118	0.146	.702
rs4919093	*FRAT2*	10	0.014	G:C	C	0.016	0.012	0.143	.705
rs12944877	*GNA13*	17	0.118	A:G	G	0.121	0.112	0.140	.708
rs2322599	*PTK2B*	8	0.384	G:A	A	0.389	0.376	0.139	.709
rs273269	*PIK3R2*	19	0.124	C:T	T	0.127	0.119	0.137	.711
rs7096152	*SORBS1*	10	0.288	A:C	C	0.293	0.281	0.132	.717
rs17788127	*PTPN1*	20	0.089	G:T	G	0.914	0.907	0.128	.721
rs2503706	*PRKCZ*	1	0.305	C:T	T	0.310	0.298	0.127	.722
rs4740304	*RAPGEF1*	9	0.221	A:G	A	0.783	0.773	0.113	.737

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2960266	*SH2B2*	7	0.349	G:A	A	0.353	0.342	0.112	.738
rs2242584	*SH2B2*	7	0.354	A:G	G	0.358	0.347	0.111	.739
rs2968845	*PPP1R3A*	7	0.323	G:A	G	0.682	0.671	0.106	.745
rs173702	*PIK3R1*	5	0.401	A:G	G	0.406	0.394	0.101	.751
rs12447555	*PDPK1*	16	0.106	C:G	C	0.897	0.890	0.098	.755
rs1500111	*PRKCI*	3	0.353	C:T	T	0.357	0.347	0.096	.757
rs4381631	*PRKCA*	17	0.496	A:C	A	0.508	0.497	0.095	.757
rs251409	*PIK3R1*	5	0.157	A:G	G	0.160	0.152	0.095	.758
rs6584000	*SORBS1*	10	0.270	T:C	T	0.733	0.724	0.095	.758
rs11765492	*GNA12*	7	0.113	G:A	A	0.115	0.109	0.085	.771
rs6584004	*SORBS1*	10	0.447	C:T	T	0.451	0.441	0.083	.774
rs12026349	*PRKCZ*	1	0.108	G:T	G	0.895	0.888	0.082	.774
rs2276554	*RHOQ*	2	0.128	A:G	G	0.131	0.124	0.078	.780
rs9910301	*PRKCA*	17	0.367	A:C	A	0.637	0.627	0.075	.784
rs1042522	*TP53*	17	0.268	G:C	C	0.271	0.263	0.073	.788
rs7713645	*PIK3R1*	5	0.496	C:A	A	0.500	0.491	0.069	.793
rs2460002	*PRKCZ*	1	0.306	G:A	A	0.310	0.301	0.068	.794
rs4917696	*SORBS1*	10	0.269	C:T	C	0.734	0.727	0.059	.808
rs3808803	*RAPGEF1*	9	0.446	G:A	A	0.449	0.441	0.054	.816
rs9291926	*PIK3R1*	5	0.452	G:T	T	0.455	0.447	0.051	.821
rs10841855	*GYS2*	12	0.264	G:T	G	0.738	0.731	0.050	.823
rs7342847	*PRKCA*	17	0.348	C:T	T	0.351	0.344	0.048	.827

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs10736089	*SORBS1*	10	0.199	C:A	A	0.201	0.195	0.047	.828
rs11243480	*RAPGEF1*	9	0.290	C:T	T	0.293	0.286	0.042	.837
rs17208470	*IRS1*	2	0.096	A:C	C	0.097	0.093	0.041	.839
rs2844704	*FLOT1*	6	0.156	G:T	T	0.158	0.153	0.039	.844
rs11576136	*RPS6KA1*	1	0.052	C:T	T	0.053	0.050	0.037	.847
rs400288	*AKAP11*	13	0.217	T:G	T	0.785	0.780	0.037	.847
rs2270937	*GYS1*	19	0.324	G:A	A	0.326	0.320	0.036	.850
rs11188327	*SORBS1*	10	0.261	T:C	C	0.264	0.258	0.036	.850
rs9905351	*PRKCA*	17	0.439	G:A	A	0.441	0.435	0.035	.851
rs2302485	*GSK3A*	19	0.076	A:G	A	0.926	0.922	0.033	.856
rs7220127	*PRKCA*	17	0.434	T:C	T	0.568	0.562	0.031	.859
rs1799999	*PPP1R3A*	7	0.087	G:T	T	0.088	0.084	0.031	.861
rs831125	*PIK3R1*	5	0.101	A:G	A	0.900	0.896	0.031	.861
rs12569998	*PTEN*	10	0.149	T:G	G	0.150	0.146	0.031	.861
rs8080721	*PRKCA*	17	0.487	T:C	T	0.516	0.509	0.030	.862
rs6863431	*PIK3R1*	5	0.211	A:C	A	0.791	0.786	0.028	.868
rs9525605	*AKAP11*	13	0.067	G:A	A	0.068	0.066	0.023	.878
rs208358	*GNA12*	7	0.246	C:T	C	0.756	0.752	0.020	.888
rs11656099	*PRKCA*	17	0.283	G:A	G	0.719	0.714	0.019	.889
rs12572106	*PTEN*	10	0.149	T:C	C	0.150	0.147	0.019	.890
rs8078231	*PRKCA*	17	0.255	C:T	C	0.746	0.742	0.015	.901
rs13082485	*PIK3CA*	3	0.110	G:A	G	0.891	0.888	0.012	.913

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2677764	*PIK3CA*	3	0.111	G:A	G	0.890	0.888	0.010	.920
rs12452841	*GNA13*	17	0.060	A:T	T	0.061	0.059	0.009	.923
rs4983387	*AKT1*	14	0.083	A:G	A	0.918	0.916	0.009	.926
rs2498802	*AKT1*	14	0.362	C:G	G	0.363	0.360	0.008	.929
rs706711	*PIK3R1*	5	0.448	G:A	G	0.553	0.550	0.007	.931
rs4730603	*PPP1R3A*	7	0.085	G:T	T	0.086	0.084	0.006	.937
rs7768046	*FYN*	6	0.401	A:G	A	0.600	0.597	0.006	.938
rs939269	*PTK2B*	8	0.253	A:T	T	0.254	0.252	0.006	.939
rs5435	*SLC2A4*	17	0.362	C:T	C	0.639	0.637	0.006	.940
rs10882609	*SORBS1*	10	0.407	C:T	T	0.408	0.406	0.005	.942
rs13045492	*PTPN1*	20	0.095	A:T	A	0.906	0.904	0.005	.946
rs4804195	*INSR*	19	0.427	G:C	G	0.574	0.572	0.004	.947
rs746754	*RAPGEF1*	9	0.356	A:G	A	0.645	0.643	0.004	.948
rs7995222	*AKAP11*	13	0.483	A:G	A	0.518	0.516	0.003	.954
rs8064946	*TP53*	17	0.141	G:C	C	0.142	0.141	0.003	.956
rs1550805	*PIK3R1*	5	0.086	C:T	C	0.914	0.913	0.003	.959
rs2498804	*AKT1*	14	0.333	G:T	T	0.334	0.332	0.003	.960
rs8111710	*INSR*	19	0.213	G:T	G	0.787	0.786	0.002	.962
rs1534243	*SH2B2*	7	0.371	C:T	T	0.371	0.370	0.002	.965
rs1292033	*RPS6KB1*	17	0.203	T:A	A	0.203	0.202	0.002	.965
rs7089306	*SORBS1*	10	0.460	C:T	T	0.461	0.460	0.001	.971
rs1042265	*GYS1*	19	0.100	C:T	T	0.100	0.099	0.001	.977

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2509671	*CBL*	11	0.294	G:T	G	0.706	0.705	0.001	.978
rs804192	*FYN*	6	0.239	G:A	G	0.762	0.761	0.001	.978
rs884080	*PRKCZ*	1	0.451	A:G	G	0.451	0.450	0.001	.981
rs12979424	*INSR*	19	0.314	G:A	A	0.314	0.314	0.001	.981
rs1234223	*PTEN*	10	0.087	G:C	C	0.087	0.087	0.000	.986
rs11202596	*PTEN*	10	0.090	G:A	G	0.910	0.910	0.000	.991
rs10408844	*INSR*	19	0.198	T:C	C	0.198	0.198	0.000	.992
rs222847	*SLC2A4*	17	0.037	T:A	T	0.963	0.962	0.000	.994
rs706905	*FYN*	6	0.205	A:G	G	0.205	0.205	0.000	.997
rs4644888	PRKCA	17	0.486	G:A	failed HWE				
rs7994365	AKAP11	13	0.000	A:A	Nonpolymorphic				
rs927010	FYN	6	0.000	G:G	Nonpolymorphic				
rs3745233	GSK3A	19	0.000	T:T	Nonpolymorphic				
rs1182174	GNA12	7	0.000	T:T	Nonpolymorphic				
rs2227988	CBL	11	0.000	C:C	Nonpolymorphic				
rs2227986	CBL	11	0.000	C:C	Nonpolymorphic				
rs12585507	IRS2	13	n/a	n/a	failed QC				
rs12415527	SORBS1	10	n/a	n/a	failed QC				
rs3781373	FRAT1	10	n/a	n/a	failed QC				
rs16946807	CRK	17	n/a	n/a	failed QC				
rs11880337	INSR	19	n/a	n/a	failed QC				
rs2267630	GSK3A	19	n/a	n/a	failed QC				

Supplementary Table 3 | *Continued*

Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2526354	RPS6KB1	17	n/a	n/a	failed QC				
rs10786213	SORBS1	10	n/a	n/a	failed QC				
rs3972219	GNA13	17	n/a	n/a	failed QC				
rs7215315	FLOT2	17	n/a	n/a	failed QC				
rs6510975	INSR	19	n/a	n/a	failed QC				
rs12372504	GYS2	12	n/a	n/a	failed QC				
rs1705584	PRKCI	3	n/a	n/a	failed QC				
rs803073	SH2B2	7	n/a	n/a	failed QC				
rs3130660	FLOT1	6	n/a	n/a	failed QC				
rs908742	PRKCZ	1	n/a	n/a	failed QC				
rs262661	PRKCZ	1	n/a	n/a	failed QC				
rs7726	RPS6KA1	1	n/a	n/a	failed QC				
rs17548850	PIK3CB	3	n/a	n/a	failed QC				
rs3729774	PIK3CB	3	n/a	n/a	failed QC				
rs516741	PRKCI	3	n/a	n/a	failed QC				
rs251406	PIK3R1	5	n/a	n/a	failed QC				
rs2960257	SH2B2	7	n/a	n/a	failed QC				
rs12357281	PTEN	10	n/a	n/a	failed QC				
rs7900095	SORBS1	10	n/a	n/a	failed QC				
rs12414693	SORBS1	10	n/a	n/a	failed QC				
rs1334891	FRAT1	10	n/a	n/a	failed QC				
rs11217191	CBL	11	n/a	n/a	failed QC				

Supplementary Table 3 | *Continued*

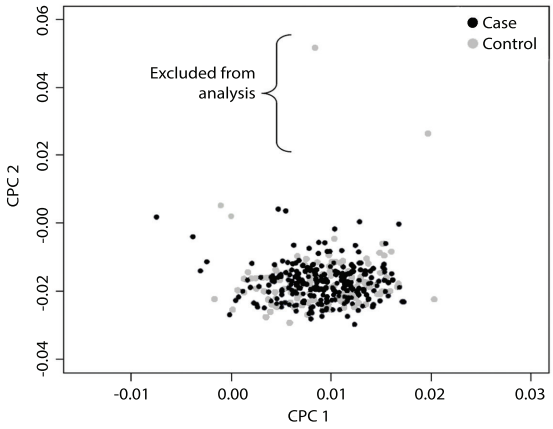
Name	Gene	Chr	MAF	Alleles (major:minor)	Associated allele or status	Case frequencies	Control frequencies	Chi-square value	*P* value
rs2229073	CBL	11	n/a	n/a	failed QC				
rs913949	IRS2	13	n/a	n/a	failed QC				
rs4771644	IRS2	13	n/a	n/a	failed QC				
rs758319	PDPK1	16	n/a	n/a	failed QC				
rs7201584	PDPK1	16	n/a	n/a	failed QC				
rs1292035	RPS6KB1	17	n/a	n/a	failed QC				
rs11867695	PRKCA	17	n/a	n/a	failed QC				
rs17710992	PRKCA	17	n/a	n/a	failed QC				
rs2860183	INSR	19	n/a	n/a	failed QC				
rs273276	PIK3R2	19	n/a	n/a	failed QC				
rs708598	GSK3A	19	n/a	n/a	failed QC				

Single-nucleotide polymorphisms (SNPs) are displayed in order of significance. *INSR* and *IRS2* SNPs that were selected for genotyping in the replication cohort are indicated in bold. *P* values are for allelic association with polycystic ovary syndrome (no covariate adjustment). *Chr,* chromosome; *MAF,* minor allele frequency

Supplementary Table 4 | Ancestry informative markers genotyped in the discovery cohort.

rs2986742	rs10007810	rs12544346	rs946918
rs6541030	rs1369093	rs7844723	rs200354
rs647325	rs385194	rs2001907	rs3784230
rs4908343	rs7657799	rs1408801	rs12439433
rs1325502	rs2702414	rs10511828	rs2899826
rs12130799	rs316598	rs3793451	rs8035124
rs3118378	rs870347	rs2306040	rs4984913
rs3737576	rs37369	rs10513300	rs4781011
rs7554936	rs6451722	rs2073821	rs2269793
rs1040404	rs12657828	rs3793791	rs818386
rs1407434	rs6556352	rs4746136	rs2966849
rs4951629	rs1500127	rs4918842	rs1879488
rs316873	rs6422347	rs4880436	rs2033111
rs798443	rs1040045	rs10839880	rs11652805
rs7421394	rs2504853	rs1837606	rs10512572
rs4666200	rs7745461	rs2946788	rs2125345
rs4670767	rs192655	rs11227699	rs4798812
rs13400937	rs4463276	rs948028	rs4800105
rs260690	rs4458655	rs2416791	rs7238445
rs10496971	rs1871428	rs1513056	rs881728
rs2627037	rs731257	rs214678	rs4891825
rs1569175	rs32314	rs772262	rs874299
rs10510228	rs2330442	rs2070586	rs8113143
rs4955316	rs4717865	rs1503767	rs3745099
rs9809104	rs10954737	rs9319336	rs2532060
rs6548616	rs705308	rs7997709	rs6104567
rs12629908	rs7803075	rs9530435	rs3907047
rs9845457	rs10236187	rs9522149	rs2835370
rs734873	rs6464211	rs1760921	rs1296819
rs2030763	rs10108270	rs2357442	rs4821004
rs1513181	rs3943253	rs1950993	rs5768007
rs9291090	rs1471939	rs8021730	

Supplementary Figure 1 | Principal component plot of the discovery cohort. The plot confirms that the cases and controls are well matched, except for two controls who were excluded from further analysis.



CPC Common Principal Component

Chapter 4.2

Variants in *SULT2A1* Affect the DHEA Sulphate to DHEA Ratio in Patients with Polycystic Ovary Syndrome, But Not the Hyperandrogenic Phenotype

Y.V. Louwers
F.H. de Jong
N.A.A. van Herwaarden
L. Stolk
B.C.J.M Fauser
A.G. Uitterlinden
J.S.E. Laven

Journal of Clinical Endocrinology & Metabolism 2013 Sep;98(9):3848-55

Abstract

Context: Because of the elevated dehydroepiandrosterone sulfate (DHEAS) levels in polycystic ovary syndrome (PCOS) and the heritability of DHEAS serum levels, genes encoding the enzymes that control the sulfation of dehydroepiandrosterone (DHEA) to DHEAS and vice versa are obvious candidate genes to explain part of the heritability of PCOS.

Objective: The objective of the study was to determine the role of genetic variants in sulfotransferase (*SULT2A1*), 3-phosphoadenosine 5-phosphosulfate synthase isoform 2 (*PAPSS2*) and steroid sulfatase (*STS*) in PCOS and in hormone levels related to the hyperandrogenic phenotype of PCOS.

Design: This was a candidate-gene study.

Patients: The discovery set consisted of 582 patients and 2017 controls.

Main outcome measures: A pruned subset of 28 single nucleotide polymorphisms (SNPs) in *SULT2A1*, *PAPSS2* and *STS* was generated based on pairwise genotypic correlation. Association with PCOS was tested and we studied whether the SNPs modulate DHEAS levels, DHEA levels and their ratio in PCOS. Significant SNPs were replicated in an independent sample of patients.

Results: None of the SNPs in *SULT2A1*, *PAPSS2* and *STS* constituted risk alleles for PCOS. SNP rs2910397 in *SULT2A1* decreased the DHEAS to DHEA ratio in PCOS by 5% in the discovery sample. Meta-analysis of discovery and replication sample resulted in a combined effect of -0.095 ($P = 0.027$). However, carrying the minor T-allele did not contribute to differences in hyperandrogenic phenotype, including levels of T and androstenedione, of PCOS patients.

Conclusions: Genetic variants in *SULT2A1*, *PAPSS2* and *STS* do not predispose to PCOS. Although a variant in *SULT2A1* decreased the DHEAS to DHEA ratio, no changes in other androgenic hormone levels were observed.

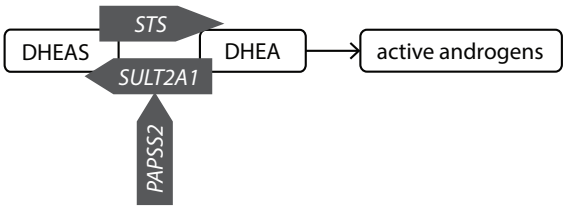
Introduction

Hyperandrogenism is one of the cardinal features of the polycystic ovary syndrome (PCOS). Overall prevalence of hyperandrogenism in PCOS patients amounts up to 75%, and increased levels of androgens persist in postmenopausal PCOS patients [1,2]. Not only PCOS itself, but also hyperandrogenemia within PCOS appears to be under genetic control. Familial aggregation of androgen excess in both brothers and sisters of patients with PCOS suggests that hyperandrogenemia is a highly heritable trait [3,4].

The ovaries are considered to be the main source of androgen production in PCOS. Nevertheless, in most patients the adrenal glands contribute to the hyperandrogenism as well, and adrenal androgen production remains high until menopause as reflected by elevated levels of dehydroepiandrosterone sulfate (DHEAS) [5,6]. Heritability of DHEAS levels in twin studies was estimated to be approximately 60% [7]. DHEAS is almost exclusively of adrenocortical origin, and its serum concentration is the most abundant of the steroids secreted by the adrenal glands [8]. Moreover, it is relatively stable throughout the day and the menstrual cycle [9,10]. Elevated DHEAS levels have been reported in up to 50% of the PCOS patients [11-13]. The biologically inactive DHEAS can be converted to dehydroepiandrosterone (DHEA), which is a principal precursor for the production of biologically active androgens and estrogens [14].

An increase in circulating DHEA would lead to an increase of the androgens downstream of DHEA leading to excess of active androgens, which is one of the key features of PCOS. As illustrated in Figure 1, conversion of DHEA to DHEAS and vice versa is controlled by the enzymes sulfotransferase 2A1 (SULT2A1) and steroid sulfatase (STS). The activity of the SULT2A1 enzyme depends on the sulfate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS), which is synthesized by PAPS synthatase (PAPSS). Of the 2 isoforms PAPSS1 and PAPSS2, PAPSS2 is highly expressed in the major sites of DHEA sulfation, ie, the liver and the adrenal glands [15,16].

Figure 1 | (De) sulfation of DHEA to DHEAS and vice versa.



The sulfation of DHEA takes place through the action of DHEA sulfotransferase (SULT2A1). The activity of the SULT2A1 enzyme depends on the sulfate donor PAPS, produced by the enzyme PAPSS2. In turn, the sulfate group in DHEAS is removed by steroid sulfatase (STS) regenerating DHEA that can then be metabolized to active androgens, such as androstenedione and T.

Genetic variants in genes encoding these enzymes might influence their activity and as such might contribute to the higher levels of androgens in patients with PCOS. Unlike genetic variants in *STS*, a potential role of the single nucleotide polymorphism (SNP) rs182420 in the *SULT2A1* gene contributing to the adrenal androgen excess among PCOS patients has previously been proposed [17]. However, no replication studies have been conducted to validate this finding. Mutations in the *PAPSS2* gene have been recently suggested to be a monogenic adrenocortical cause of androgen excess in a girl with premature adrenarche, which might be an early manifestation of PCOS [18,19]. Genetic variants in *PAPSS2* have never been studied in association with PCOS before.

The aim of this study was to test genetic variations in *SULT2A1*, *PAPSS2* and *STS* for association with PCOS. Moreover, we studied whether these genetic variants modulate DHEAS levels, DHEA levels and DHEAS to DHEA ratio, as a marker for DHEA sulfation capacity, in PCOS. Finally, we studied whether carrying these genetic variants would result in differences in phenotypic characteristics, including levels of T and androstenedione, in PCOS patients.

Materials and Methods

Study population

Discovery sample

A total of 582 Caucasian PCOS patients and 2017 Caucasian controls from the general population were included in the discovery cohort. All PCOS patients underwent a standardized physical examination and hormonal evaluation. Clinical work-up included menstrual history as well as current cycle length, anthropometric measurements (height and weight), and calculation of body mass index (BMI) as weight in kilograms divided by the squared height in meters. Transvaginal ultrasonography was performed to assess ovarian volume and total number of follicles in both ovaries. The diagnosis PCOS was established on the basis of the 2003 European Society for Human Reproduction and Embryology/American Society for Reproductive Medicine Rotterdam criteria [20,21].

Controls were derived from the Rotterdam Study, which is a large population based study of elderly men and women from Ommoord, a defined area in Rotterdam, The Netherlands. The design of the study has been described in detail previously [22]. All women of whom DNA was available, who were proven to be fertile, with an age of menopause > 45 years and age of menarche between 10-17 years and available BMI measurements were included as controls. These controls provide a reference group of the local general Caucasian population rather than a control group in which PCOS was specifically excluded. The study was approved by the Medical Ethical Review Board of the Erasmus Medical Center, Rotterdam, and written informed consent was obtained from all patients as well as from control subjects.

Replication sample

Replication efforts were carried out in an independent sample consisting of 412 Caucasian PCOS patients from the area near Utrecht, The Netherlands. PCOS was diagnosed according to the 2003 revised Rotterdam criteria [20,21]. These patients were screened according to the same protocol as the PCOS patients in the discovery sample including anthropometric measurements, transvaginal ultrasonography and extensive endocrine evaluation. The study was approved by the Institutional Review Board of the University Medical Center Utrecht and all patients gave written informed consent.

Hormone assays

Blood samples of all patients were drawn after an overnight fast before 10.00 am on the day of clinical examination and processed within 2h after withdrawal. Until assayed, serum was stored at -20°C. In Rotterdam serum levels of SHBG, androstenedione, insulin and DHEAS were measured by immunoluminometric assay (Immulite® platform, Siemens Diagnostic Products Corporation (DPC), Los Angeles, CA). Testosterone was determined by radioimmunoassay (RIA) (Siemens DPC, Los Angeles, CA). DHEA was measured using RIA (Diagnostic Systems Laboratories, Webster, TX). Glucose levels were measured using a Hitachi 917 analyzer (Roche Diagnostics, Almere, The Netherlands). Intra- and interassay coefficients of variation (CV) were less than 4 and 5% for SHBG, less than 8 and 11% for androstenedione, less than 6 and 8% for insulin, less than 9 and 11% DHEAS, less than 3 and 5% for testosterone, and, less than 5 and 15% for DHEA, respectively. The ratio of DHEAS to DHEA was calculated.

In Utrecht, SHBG and insulin levels were quantified using an immunoluminometric assay (Immulite® platform, Siemens Diagnostic Products Corporation). Androstenedione was measured after hexane-toluene extraction using an in-house RIA (with extraction). DHEAS was measured using the Coat-A-Count DHEA-SO4 RIA (TKDS2, Siemens Diagnostics, Breda, Netherlands). Testosterone measurement was performed using an in-house extraction RIA. DHEA was measured after diethylether extraction and Celite chromatography using an in-house RIA (with extraction). Glucose levels were measured using a VITROS Chemistry System (Ortho-Clinical Diagnostics, Strasbourg, France) and from November 2006 on the Beckman Coulter UniCell DxC 800 chemistry analyser (Beckman Coulter, Woerden, the Netherlands).

Intra- and interassay CVs were less than 10% and 6% for SHBG, less than 4% and 12% for androstenedione, less than 3% and 8% for insulin, less than 5% and 8% for DHEAS, less than 11%, for testosterone, less than 9% and 14% for DHEA, and less than 4% and 4% for glucose, respectively. The ratio of DHEAS to DHEA was calculated.

Genotyping and quality control

Genomic DNA was extracted from peripheral venous blood samples according to standard procedures. For the discovery cohort genotypes were extracted from our Genome Wide Associated imputed genotypes database. Genotyping was performed using the Illumina HumanHap 610K bead array (Illumina). Imputation of the genome-wide genotyped data

on all autosomes was performed using the MACH algorithm (http://www.sph.umich.edu/csg/abecasis/MACH/). Imputation of the X-chromosome was performed using IMPUTE (http://mathgen.stats.ox.ac.uk/impute/impute.html). SNPs in *SULT2A1*, *PAPSS2 and STS* plus 50kb upstream and 50 kb downstream of the genes were selected using data from the Caucasian (CEU) subjects in the International HapMap database (release 24; http://hapmap.ncbi.nlm.nih.gov). Of the total 594 SNPs, i.e. 129 SNPs in *SULT2A1*, 303 SNPs in *PAPSS2* and 162 SNPs in *STS*, 335 SNPs passed the quality thresholds of an imputation quality score of at least 0.90 and a minor allele frequency of 0.01 or greater. Based on pair-wise genotypic correlation with an r^2 threshold of 0.2, a pruned subset of SNPs in *SULT2A1*, *PAPSS2* and *STS* was generated using PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) [23]. This pair-wise genotypic correlation resulted in a subset consisting of 28 independent SNPs: 9 SNPs in *SULT2A1*, 11 SNPs in *PAPSS2* and 8 SNPs in *STS*. Replication genotyping was done using Taqman Allelic Discrimination (Applied Biosystems) according to the manufacturer's instructions.

Statistical analysis

Baseline characteristics of the PCOS patients from the discovery and replication set are presented as medians with interquartile ranges and percentages. In case of non-normality, parameters were log transformed to obtain a normal distribution. The statistical analyses were carried out using the IBM Statistical Package for the Social Sciences version 20 for Windows (IBM SPSS Inc).

First, to determine association of the SNPs in *SULT2A1*, *PAPSS2* and *STS* with PCOS, a case-control analysis was carried out in the discovery set using logistic regression in MACH2DAT [24] for the SNPs in *SULT2A1* and *PAPSS2* and in SNPTEST [25] for the SNPs in *STS*. These analyses were adjusted for the first four principal components obtained from multidimensional scaling to correct for population stratification. Secondly, SNPs in *SULT2A1*, *PAPSS2* and *STS* were tested in relation to DHEAS levels, DHEA levels and DHEAS to DHEA ratio in PCOS patients from the discovery set using MACHQTL [24] and SNPTEST [25]. Because the DHEAS and DHEA levels as well as the ratio were non-normally distributed we used the log transformed data in the analyses.

Analyses within the PCOS set were first only adjusted for age and the first four principal components. In a second run the analyses were adjusted for principle components, age, BMI and insulin levels. The threshold for significance was set to a $P \leq 8.9e\text{-}4$ [0.05/(28 SNPs*2 independent analyses)].

The discovery sample had power greater than 80% to detect an association of a risk allele with a frequency of 0.25 at an odds ratio of 1.5 with an alpha of 8.9e-4. SNPs reaching this threshold for significance were validated in the independent replication set of PCOS patients from Utrecht, the Netherlands using adjusted linear regression analysis. Replication analyses were carried out using SPSS. In the replication analysis Bonferroni adjustment was applied for the number of SNPs tested. Combined effect of the significant SNPs was tested using random

effects meta-analysis in R, statistical software for data analysis (http://www.R-project.org) [26]. To provide some insight in the role of the identified variants in SULT2A1, STS and PAPSS2 in the general population, association of significant SNPs with DHEAS levels, DHEA levels and their ratio will be tested in women from the Rotterdam Study. These data is only available in a subset of women included in the Rotterdam study, ie, for DHEAS levels n=852; for DHEA levels, n=303; and for their ratio, n=290. However, because there is only little overlap (n=95) between the subset of women with known DHEAS and DHEA levels in the Rotterdam Study and the women who were included as controls in the current study, and because a large age difference is present between PCOS cases (mean age 29.2 years) and controls (mean age 71.4 years), no direct comparisons will be made (Supplementary Note).

Finally, to determine the influence of the replicated SNPs on the hyperandrogenic phenotype of PCOS, including testosterone and androstenedione levels, we compared phenotypic characteristics of PCOS patients who carried one of both minor allele of the significantly associated SNP with phenotypic characteristics of the patients who were homozygous for the major allele. These data were presented as median and interquartile range, and a regression analysis was performed to calculate differences between the carriers and noncarriers using SPSS. Analyses were adjusted for age, BMI, and insulin levels as indicated.

Results

Baseline characteristics of the PCOS patients from Rotterdam and Utrecht are listed in Table 1. The PCOS patients from Rotterdam had a more severe PCOS phenotype, indicated by a higher percentage of amenorrhea, a higher follicle count, higher BMI levels, higher insulin levels and higher androstenedione levels compared to patients from Utrecht. First, we determined the association of SNPs in *SULT2A1*, *PAPSS2* and *STS* with PCOS. None of the SNPs in the genes constituted risk alleles for PCOS (Table 2).

As indicated in Table 3, none of the SNPs in *SULT2A1*, *PAPSS2* and *STS* were associated with logtransformed DHEAS levels in PCOS. SNP rs12011936 mapping *STS* showed a trend towards significance in relation to log-transformed DHEA levels as well as the ratio between DHEAS and DHEA. However, these variants did not reach the threshold for significance.

Carriers of the minor T-allele of SNP rs2910397 in *SULT2A1*, had a decreased log-transformed DHEAS to DHEA ratio compared to noncarriers (effect -0.052, se 0.014, $P = 1.24e\text{-}4$). Each extra copy of the T-allele at rs2910397 decreased the log-transformed DHEAS to DHEA ratio with 0.052, resulting in a 5.1% decrease in this ratio in patients suffering from PCOS. Additional adjustment of the association of rs2910397 for BMI and insulin levels did not change our results.

Table 1 | Phenotypic characteristics of the PCOS patients in the discovery and replication sample.

	Discovery sample	**Replication sample**
	n=582	**n=412**
Age, years	29.2 (25.5-32.0)	28.7 (25.0-31.9)
BMI, kg/m^2	24.7 (21.4-30.1)	23.4 (20.7-28.4)
Amenorrhea	32.5%	27.6%
Oligomenorrhea	66.2%	70.9%
Regular cycle	1.4%	1.5%
Total follicle count, n	36 (26-50)	30 (24-39)
PCOM, %	93.4%	90.8%
DHEA, nmol/l	13.0 (9.2-19.0)	17.0 (13.0-24.0)
DHEAS, µmol/l	5.0 (3.5-6.7)	4.7 (3.3-6.5)
DHEAS/DHEA ratio, µmol/nmol	0.36 (0.26-0.50)	0.26 (0.2-0.35)
Adione, nmol/l	10.5 (8.1-13.4)	6.7 (5.1-8.8)
T, nmol/l	1.8 (1.3-2.3)	2.0 (1.5-2.4)
SHBG, nmol/l	43.3 (29.6-63.5)	52 (35-72)
FAI	4.1 (2.2-6.8)	3.7 (2.3-6.0)
Insulin, pmol/l	50 (34-82)	35 (21-63)
Glucose, mmol/l	4.2 (3.8-4.6)	5.0 (4.8-5.3)

Adione, androstenedione; *FAI*, free androgen index [(T/SHBG)100]; *PCOM*, polycystic ovarian morphology. Data are presented as median (interquartile range: 25th to 75th percentile), unless otherwise indicated.

Similar to the results in the discovery set, replication of the rs2910397 SNP in the independent set of PCOS patients from Utrecht was highly significantly associated with the log-transformed DHEAS to DHEA ratio after correction for age (effect -0.138, se 0.013, $P = 0.005$) and nominally significantly associated with the log-transformed DHEAS levels (effect -0.090, se 0.015, $P = 0.05$). Random effect meta-analysis of the discovery and replication sample resulted in a combined effect of -0.095 ($P = 0.027$) of rs2910397 on the ratio. In the general population, SNP rs2910397 was associated with neither the DHEAS levels nor the DHEA levels ($P = 0.212$ and $P = 0.373$, respectively), however, the SNP was nominally significantly associated with the DHEAS to DHEA ratio (effect = -0.036, $P = 0.04$, see also supplementary note). Table 4 depicts the effect of carrying the minor T-allele of rs2910397 on several phenotypic characteristics of PCOS. We aimed to determine whether carrying this minor allele would result in differences in phenotype within PCOS. Obviously, the DHEAS to DHEA ratio was significantly decreased in the carriers of one or both minor T-alleles at rs2910397 compared to the noncarriers. However, all other parameters were similar in the carriers compared with the noncarriers.

Table 2 | Case-control association analysis of genetic variants in *SULT2A1*, *PAPSS2* and *STS* in PCOS cases versus controls from the general population.

Gene	Chr	SNP	Position	Alleles	OR (95% CI)	SE	*P* value
SULT2A1	19	rs10410589	53016534	C,A	1.06 (0.92-1.23)	0.075	0.43
		rs17272610	53020551	T,A	0.96 (0.83-1.12)	0.077	0.58
		rs3760818	53028770	A,T	1.00 (0.88-1.14)	0.068	0.99
		rs8100405	53034176	C,G	0.97 (0.85-1.11)	0.069	0.68
		rs12462534	53036382	A,G	1.08 (0.95-1.24)	0.069	0.28
		rs4802397	53053724	T,C	1.07 (0.88-1.30)	0.099	0.46
		rs3936308	53078612	C,T	0.83 (0.47-1.49)	0.296	0.54
		rs2932766	53082236	T,G	1.05 (0.91-1.21)	0.073	0.52
		rs2910397	53089929	T,C	1.02 (0.88-1.18)	0.074	0.83
PAPSS2	10	rs10887732	89362213	T,G	1.00 (0.87-1.14)	0.068	0.99
		rs17110864	89364066	C,T	1.03 (0.82-1.29)	0.114	0.80
		rs1555433	89408867	G,A	0.93 (0.78-1.10)	0.086	0.38
		rs17430359	89418502	T,G	0.87 (0.66-1.15)	0.142	0.33
		rs17774221	89421320	G,A	1.17 (0.81-1.71)	0.191	0.40
		rs7903516	89431082	T,C	1.09 (0.95-1.24)	0.068	0.22
		rs12570024	89432677	C,T	1.08 (0.79-1.48)	0.160	0.62
		rs941830	89447141	G,A	0.96 (0.84-1.09)	0.068	0.50
		rs7919607	89459569	T,C	0.99 (0.86-1.13)	0.069	0.86
		rs2302404	89464052	C,T	1.04 (0.69-1.58)	0.212	0.84
		rs1321934	89485344	A,G	1.01 (0.88-1.16)	0.070	0.92
STS	X	rs17331438	7012290	G,A	1.59 (1.14-2.20)	0.168	0.006
		rs12011936	7177230	G,T	1.04 (0.89-1.22)	0.079	0.59
		rs5934670	6982708	A,G	0.87 (0.75-1.02)	0.078	0.08
		rs4830732	7140734	C,T	1.23 (1.04-1.47)	0.090	0.02
		rs802891	6959714	G,A	0.69 (0.49-0.97)	0.173	0.03
		rs5934937	7100818	C,G	1.00 (0.87-1.14)	0.069	0.94
		rs5933907	7176576	G,A	1.02 (0.86-1.22)	0.090	0.79
		rs802901	6947743	G,A	1.00 (0.88-1.15)	0.069	0.96

Chr, chromosome; *CI*, confidence interval; *OR*, odds ratio; *SE*, standard error. The first mentioned allele is the effect allele.

Table 3 | Association of genetic variants in *SULT2A1*, *PAPSS2* and *STS* in relation to log DHEAS, log DHEA and their log ratio in PCOS patients from the discovery sample.

						log DHEAS			log DHEA			log DHEAS to DHEA ratio		
	Chr	SNP	Position	Alleles	Freq 1	Effect 1	SE	*P* value	Effect 1	SE	*P* value	Effect 1	SE	*P* value
SULT2A1	19	rs10410589	53016534	C,A	0.73	0.002	0.015	0.88	0.003	0.016	0.84	-0.002	0.013	0.87
		rs17272610	53020551	T,A	0.25	-0.031	0.015	0.03	0.007	0.016	0.64	-0.038	0.014	0.006
		rs3760818	53028770	A,T	0.45	-0.036	0.013	0.01	-0.009	0.015	0.56	-0.029	0.013	0.02
		rs8100405	53034176	C,G	0.45	-0.015	0.013	0.23	0.010	0.014	0.49	-0.026	0.012	0.03
		rs12462534	53036382	A,G	0.41	-0.030	0.014	0.03	0.003	0.015	0.84	-0.033	0.013	0.01
		rs4802397	53053724	T,C	0.86	-0.016	0.019	0.41	-0.013	0.020	0.53	-0.003	0.017	0.84
		rs3936308	53078612	C,T	0.99	-0.045	0.056	0.42	-0.020	0.061	0.75	-0.030	0.052	0.57
		rs2932766	53082236	T,G	0.68	-0.005	0.014	0.72	0.020	0.015	0.18	-0.027	0.013	0.04
		rs2910397	53089929	T,C	0.29	-0.022	0.015	0.13	0.030	0.016	0.06	-0.052	0.014	1.24e-4
PAPSS2	10	rs10887732	89362213	T,G	0.47	-0.007	0.013	0.59	-0.013	0.014	0.36	0.005	0.012	0.66
		rs17110864	89364066	C,T	0.90	-0.013	0.022	0.55	0.009	0.024	0.71	-0.022	0.021	0.29
		rs1555433	89408867	G,A	0.81	0.002	0.017	0.90	0.009	0.018	0.63	-0.006	0.015	0.68
		rs17430359	89418502	T,G	0.95	-0.034	0.027	0.120	-0.003	0.029	0.91	-0.033	0.025	0.20
		rs17774221	89421320	G,A	0.96	0.022	0.038	0.56	0.044	0.041	0.29	-0.020	0.035	0.56
		rs7903516	89431082	T,C	0.50	-0.016	0.014	0.25	-0.003	0.015	0.85	-0.013	0.013	0.31
		rs12570024	89432677	C,T	0.95	-0.002	0.031	0.95	0.018	0.034	0.60	-0.018	0.029	0.54
		rs941830	89447141	G,A	0.60	-0.029	0.013	0.03	-0.005	0.014	0.70	-0.023	0.012	0.06
		rs7919607	89459569	T,C	0.42	-0.013	0.013	0.32	-0.004	0.015	0.80	-0.010	0.012	0.41
		rs2302404	89464052	C,T	0.97	-0.061	0.042	0.14	0.038	0.045	0.41	-0.098	0.039	0.01
		rs1321934	89485344	A,G	0.35	0.003	0.014	0.85	-0.001	0.015	0.94	0.004	0.013	0.77

Table 3 | *Continued*

						log DHEAS			log DHEA			log DHEAS to DHEA ratio		
	Chr	SNP	Position	Alleles	Freq 1	Effect 1	SE	*P* value	Effect 1	SE	*P* value	Effect 1	SE	*P* value
STS	X	rs17331438	7012290	G,A	0.96	0.271	0.16	0.10	0.208	0.170	0.22	0.063	0.169	0.71
		rs12011936	7177230	G,T	0.76	0.031	0.07	0.64	0.219	0.069	0.001	-0.225	0.070	0.001
		rs5934670	6982708	A,G	0.65	-0.027	0.06	0.68	-0.165	0.067	0.01	-0.153	0.068	0.02
		rs4830732	7140734	C,T	0.78	0.070	0.08	0.38	0.178	0.079	0.02	-0.131	0.081	0.11
		rs802891	6959714	G,A	0.92	0.071	0.13	0.57	0.224	0.131	0.09	-0.199	0.139	0.15
		rs5934937	7100818	C,G	0.61	-0.106	0.06	0.07	-0.063	0.060	0.29	-0.041	0.060	0.49
		rs5933907	7176576	G,A	0.84	0.046	0.08	0.55	-0.071	0.079	0.36	0.136	0.080	0.09
		rs802901	6947743	G,A	0.60	-0.089	0.06	0.13	-0.055	0.059	0.36	-0.033	0.060	0.58

Chr, chromosome; *SE*, standard error. All analyses were adjusted for age and the first 4 principal components. Effect 1 is the effect of the first mentioned allele.

Table 4 | Comparison of phenotypic characteristics of PCOS patients based on genotype of rs2910397 in PCOS patients from Rotterdam and Utrecht.

	PCOS patients Rotterdam (n=582)					PCOS patients Utrecht (n=407[a])				
	CC	CT	TT	*P* value[b]	*P* value[c]	CC	CT	TT	*P* value[b]	*P* value[c]
	n=276	n=262	n=44			n=197	n=167	n=43		
Age y	29.1 (26.0-31.8)	29.3 (25.5-32.4)	28.5 (25.4-31.7)	0.70	0.69	28.7 (24.8-28.3)	28.7 (20.5-29.2)	28.1 (24.6-31.3)	0.73	0.77
BMI kg/m^2	24.7 (21.2-29.9)	24.9 (21.8-30.6)	24.0 (20.5-27.9)	0.22	0.39	23.7 (21.0-28.3)	23.4 (20.5-29.2)	22.7 (20.8-25.9)	0.83	0.59
Amenorrhea, %	34.1	30.9	31.8	0.49	0.67	25.5	28.7	31.7	0.55	0.50
Total follicle number	36 (26-51)	37 (27-51)	30 (24-46)	0.76	0.06	30 (24-40)	30 (25-39)	28 (24-32)	0.52	0.49
Mean ovarian volume, ml	8.2 (6.0-10.7)	8.4 (6.5-10.9)	8.1 (5.1-11.1)	0.18	0.37	7.4 (4.9-10.2)	6.8 (5.4-9.0)	7.0 (5.6-10.0)	0.49	0.49
T, nmol/l	1.7 (1.3-2.2)	1.9 (1.2-2.4)	1.8 (1.2-2.2)	0.35	0.96	2.0 (1.6-2.4)	1.9 (1.4-2.4)	2.0 (1.5-2.7)	0.32	0.98
SHBG, nmol/l[d]	46.2 (29.6-66.7)	40.5 (29.3-59.0)	43.8 (29.9-63.7)	0.06	0.46	51 (34-74.5)	51.5 (36-71.5)	59 (36-72)	0.59	0.96
FAI[d]	4.0 (6.2)	4.3 (2.4-7.3)	4.4 (2.1-6.5)	0.06	0.50	3.4 (2.3-5.9)	3.7 (2.3-6.1)	3.2 (2.6-6.1)	0.65	0.89
Adione, nmol/l	10.2 (8.0-12.8)	10.8 (8.2-13.7)	11.5 (9.0-14.2)	0.32	0.33	6.9 (5.4-8.8)	6.3 (4.9-8.7)	7.1 (5.3-9.7)	0.31	0.52
DHEA, nmol/l[d,e]	12.7 (9.3-18.3)	13.1 (9.0-19.0)	16.4 (10.2-23.8)	0.70	0.02	17.0 (12.5-24.0)	17.0 (13-24)	18.0 (13-27)	0.57	0.30
DHEAS, µmol/l[d,e]	5.1 (3.6-6.8)	4.7 (3.4-6.6)	5.2 (3.8-6.6)	0.07	0.36	5.0 (3.4-6.9)	4.2 (3.2-6.4)	4.2 (2.8-6.4)	0.12	0.12
DHEAS/DHEA, µmol/nmol[d,e]	0.39 (0.29-0.52)	0.36 (0.25-0.47)	0.27 (0.24-0.37)	0.01	0.001	0.28 (0.21-0.36)	0.26 (0.20-0.34)	0.23 (0.17-0.34)	0.35	0.002
Insulin, pmol/l[d]	49 (35-81)	51 (34-82.3)	38 (22-75)	0.27	0.04	34 (21-63)	42 (21-76)	35 (21-48)	0.30	0.80
Glucose, mmol/l[d]	4.2 (3.8-4.6)	4.2 (3.7-4.6)	4.1 (4.0-4.5)	0.42	0.19	5.0 (4.8-5.3)	5.0 (4.7-5.3)	4.9 (4.6-5.3)	0.71	0.30

Adione, androstenedione; *FAI*, free androgen index. All analyses were adjusted for age.

[a] Genotypes of 5 patients were not determined.

[b] *P* values were calculated between carriers of CC and CT genotypes.

[c] *P* values were calculated between carriers of the CC and TT genotypes.

[d] Adjusted for BMI.

[e] Adjusted for insulin levels.

Discussion

None of the investigated SNPs in *SULT2A1*, *PAPSS2,* and *STS* were associated with PCOS. We did identify and replicate one SNP in *SULT2A1* associated with the DHEAS to DHEA ratio in patients with PCOS. Although the minor allele of this SNP was associated with lower ratios, it did not affect the hyperandrogenic phenotype of PCOS.

Because of the elevated DHEAS levels in PCOS patients [11] and given the evidence that adrenal hyperandrogenism is an inherited phenotype in PCOS [3,4], genes involved in the interconversion between DHEAS and DHEA are obvious candidate genes for PCOS. Remarkably, literature evaluating these genes in PCOS is scarce. Previously, only Goodarzi et al. [17] studied *SULT2A1* and *STS* in relation to PCOS. They found no association of genetic variants in *SULT2A1* and *STS* with PCOS. However, no replication cohort was included to validate the findings, and the study was only adequately powered to detect risk alleles with large effects. In our sample, which was twice as large and adequately powered, we could confirm that indeed both *SULT2A1* and *STS* do not seem to be associated with PCOS. To our knowledge the present study is the first evaluating the role of *PAPSS2* as a risk gene for PCOS. Availability of the sulfate donor PAPS regulates the activity of *SULT2A1* [16]. Impaired sulfation of DHEA would result in elevated DHEA levels and thus increase levels of active androgens, which is an important feature of PCOS [18,19]. However, we found no association of SNPs in *PAPSS2* with PCOS suggesting that *PAPSS2* as well as *SULT2A1* and *STS*, do not constitute risk genes for PCOS.

PCOS encompasses a broad spectrum of signs and symptoms. Phenotypic characteristics, such as DHEAS and DHEA levels, vary largely between patients diagnosed with PCOS. Only half of the PCOS patients encounter elevated DHEAS levels [11-13]. Identification of modifiers of these characteristics can provide insight in the etiology of PCOS. Genetic variants mapping to *SULT2A* have been identified previously to be associated with lower DHEAS levels in PCOS patients [17]. We found that genetic variants in *SULT2A1*, *PAPSS2* and *STS* were neither associated with DHEAS nor with DHEA levels in PCOS patients. However, one variant, ie, rs2910397, was highly associated with the DHEAS to DHEA ratio in PCOS patients from Rotterdam as well as with the ratio in the independent sample of PCOS patients from Utrecht. This genetic variant rs2910397 is mapping to *SULT2A1*. Previous studies have shown that the action of *SULT2A1* is the rate-limiting step regulating the equilibrium between DHEAS and DHEA. In other words, sulfation appears to be the predominant direction of the interconversion between DHEAS and DHEA [27,28].

To further explore the role of rs2910397 as a potential disease modifier in PCOS, we compared the phenotype of carriers of one or two minor alleles with the phenotype of patients who carried both copies of the major allele. Carriers of this genetic variant in *SULT2A1* however, demonstrated no change in levels of active androgens or ovarian morphology compared to noncarriers. It is possible that SNP rs2910397 might only be an enhancing factor for PCOS or its phenotypic characteristics in association with other genetic variants. A genetic variant identified in Chinese PCOS patients was associated with hyperandrogenism in PCOS patients

from Northern European ancestry [29,30]. Interestingly, the variant was not related to a specific androgen level. The authors suggested it might be reflected a more complex relationship with serum androgen levels or the clinical features of hyperandrogenism [30].

Because PCOS is a complex genetic disorder, interaction of several genetic variants in genes encoding enzymes in the steroidogenesis might contribute to the variation in adrenal androgen levels in PCOS. Increased expression of 3β-hydroxysteroid dehydrogenase (3β-HSD) type 1 and type 2 in adipose tissue has been observed in PCOS patients [31]. These enzymes are responsible for the conversion of DHEA to androstenedione in peripheral and adrenal tissue respectively. Increased expression of 3β-HSD 1 and 2 might lead to an increase of the conversion of DHEA to androstenedione and, subsequently, an elevation of the DHEAS to DHEA ratio [32]. In addition, an increased 5α-reductase activity as well as dysregulation of 11β-HSD has been described in PCOS patients [33-36]. Both enzymes are involved in cortisol metabolism: 5α-reductase is responsible for 5alpha-reduction of cortisol to 5α-hydrocortisol in liver and 11β-HSD catalyzes the conversions of cortisone in cortisol and vice versa. Increased peripheral cortisol metabolism caused by either an increased 5α-reductase activity or dysregulation of 11β-HSD leads to a compensatory increase of ACTH secretion via a decrease in the negative feedback signal. As a result adrenal androgen levels will increase. Genetic variation in the *11β-HSD1* gene was observed to enhance cortisol clearance, and resulted in, increased DHEAS levels independently of BMI in PCOS patients [37]. On the contrary, a decrease in DHEAS levels was observed in PCOS patients carrying *CYP3A7*1C* variants [38]. The enzyme CYP3A7 metabolizes DHEA and DHEAS, facilitating their elimination [39]. Decreased levels of DHEAS levels were found in individuals carrying the *CYP3A7*1C* variant [40], which causes CYP3A7 expression to persist after birth [41]. Amongst all these factors, SNP rs2910397 in *SULT2A1* might contribute to the variation in adrenal androgens in PCOS.

Because we also observed an association of rs2910397 in *SULT2A1* with the DHEAS to DHEA ratio in controls, this SNP might be related to the genetic basis of steroidogenesis in general. Ideally, we would have also evaluated the effect of the genetic variants on DHEAS, DHEA levels as well as their ratio in PCOS patients compared with the effect in controls, but unfortunately a large age-matched control group with DHEAS as well as DHEA levels was not available. Therefore, whether the effect size of rs2910397 on the DHEAS/DHEA ratio is larger in PCOS cases compared to controls remains to be revealed.

Genetic variants in *SULT2A1,* including rs2910397, have been found to be associated with serum DHEAS in a large meta-analysis of female Caucasian subjects from the general population [42]. Probably due to lack of power, we did not observe an association with DHEAS levels in our PCOS patients.

In conclusion, genetic variants in *SULT2A1*, *PAPSS2* and *STS* may not predispose to PCOS. One genetic variant in *SULT2A1* was associated with an increase in the DHEAS to DHEA ratio in PCOS. However, clinical implications of this finding for PCOS diagnosis and etiology seem limited, since this SNP did not modulate the hyperandrogenic phenotype in PCOS and was also found be related to the DHEAS to DHEA ratio in the general population.

References

1. Huang A, Brennan K, Azziz R. Prevalence of hyperandrogenemia in the polycystic ovary syndrome diagnosed by the National Institutes of Health 1990 criteria. *Fertil Steril*. 2010; Apr; 93: 1938-41.
2. Markopoulos MC, Rizos D, Valsamakis G, et al. Hyperandrogenism in women with polycystic ovary syndrome persists after menopause. *J Clin Endocrinol Metab*. 2011; Mar; 96: 623-31.
3. Legro RS, Kunselman AR, Demers L, et al. Elevated dehydroepiandrosterone sulfate levels as the reproductive phenotype in the brothers of women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2002; May; 87: 2134-8.
4. Yildiz BO, Goodarzi MO, Guo X, et al. Heritability of dehydroepiandrosterone sulfate in women with polycystic ovary syndrome and their sisters. *Fertil Steril*. 2006; Dec; 86: 1688-93.
5. Carmina E, Lobo RA. Prevalence and metabolic characteristics of adrenal androgen excess in hyperandrogenic women with different phenotypes. *Journal of endocrinological investigation*. 2007; Feb; 30: 111-6.
6. Puurunen J, Piltonen T, Jaakkola P, et al. Adrenal androgen production capacity remains high up to menopause in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2009; Jun; 94: 1973-8.
7. Nestler JE, Whitfield JB, Williams TY, et al. Genetics of serum dehydroepiandrosterone sulfate and its relationship to insulin in a population-based cohort of twin subjects. *J Clin Endocrinol Metab*. 2002; Feb; 87: 682-6.
8. Lobo RA, Paul WL, Goebelsmann U. Dehydroepiandrosterone sulfate as an indicator of adrenal androgen function. *Obstet Gynecol.* 1981; Jan; 57: 69-73.
9. Abraham GE. Ovarian and adrenal contribution to peripheral androgens during the menstrual cycle. *J Clin Endocrinol Metab*. 1974; Aug; 39: 340-6.
10. Migeon CJ, Keller AR, Lawrence B, et al. Dehydroepiandrosterone and androsterone levels in human plasma: effect of age and sex; day-to-day and diurnal variations. *J Clin Endocrinol Metab*. 1957; Sep; 17: 1051-62.
11. Kumar A, Woods KS, Bartolucci AA, et al. Prevalence of adrenal androgen excess in patients with the polycystic ovary syndrome (PCOS). *Clin Endocrinol (Oxf)*. 2005; Jun; 62: 644-9.
12. Carmina E, Gonzalez F, Chang L, et al. Reassessment of adrenal androgen secretion in women with polycystic ovary syndrome. *Obstet Gynecol.* 1995; Jun; 85: 971-6.
13. Gonzalez F. Adrenal involvement in polycystic ovary syndrome. *Semin Reprod Endocrinol*. 1997; May; 15: 137-57.
14. Labrie F. Intracrinology. *Mol Cell Endocrinol*. 1991; Jul; 78: C113-8.
15. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocrine reviews*. 2011; Feb; 32: 81-151.
16. Strott CA. Sulfonation and molecular action. *Endocr Rev*. 2002; Oct; 23: 703-32.
17. Goodarzi MO, Antoine HJ, Azziz R. Genes for enzymes regulating dehydroepiandrosterone sulfonation are associated with levels of dehydroepiandrosterone sulfate in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2007; Jul; 92: 2659-64.
18. Ibanez L, Dimartino-Nardi J, Potau N, et al. Premature adrenarche--normal variant or forerunner of adult disease? *Endocr Rev*. 2000; Dec; 21: 671-96.
19. Noordam C, Dhir V, McNelis JC, et al. Inactivating PAPSS2 mutations in a patient with premature pubarche. *N Eng J Med.* 2009; May 28; 360: 2310-8.
20. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004; Jan; 81: 19-25.
21. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.

22. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol*. 2007; 22: 819-29.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; Sep; 81: 559-75.
24. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010; Dec; 34: 816-34.
25. Ferreira T, Marchini J. Modeling interactions with known risk loci-a Bayesian model averaging approach. *Ann Hum Genet*. 2011; Jan; 75: 1-9.
26. R Core Team. A Language and Environment for Statistical Computing. Vienna, Austria 2013.
27. Hammer F, Subtil S, Lux P, et al. No evidence for hepatic conversion of dehydroepiandrosterone (DHEA) sulfate to DHEA: in vivo and in vitro studies. *J Clin Endocrinol Metab*. 2005; Jun; 90: 3600-5.
28. Arlt W, Hammer F, Sanning P, et al. Dissociation of serum dehydroepiandrosterone and dehydroepiandrosterone sulfate in septic shock. *J Clin Endocrinol Metab*. 2006; Jul; 91: 2548-54.
29. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011; Jan; 43: 55-9.
30. Welt CK, Styrkarsdottir U, Ehrmann DA, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. *J Clin Endocrinol Metab*. 2012; Jul; 97: E1342-7.
31. Wang L, Li S, Zhao A, et al. The expression of sex steroid synthesis and inactivation enzymes in subcutaneous adipose tissue of PCOS patients. *J Steroid Biochem Mol Biol*. 2012; Oct; 132: 120-6.
32. Doi SA, Al-Zaid M, Towers PA, et al. Steroidogenic alterations and adrenal androgen excess in PCOS. *Steroids*. 2006; Sep; 71: 751-9.
33. Fassnacht M, Schlenz N, Schneider SB, et al. Beyond adrenal and ovarian androgen generation: Increased peripheral 5 alpha-reductase activity in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2003; Jun; 88: 2760-6.
34. Rodin A, Thakkar H, Taylor N, et al. Hyperandrogenism in polycystic ovary syndrome. Evidence of dysregulation of 11 beta-hydroxysteroid dehydrogenase. *N Eng J Med*. 1994; Feb 17; 330: 460-5.
35. Stewart PM, Shackleton CH, Beastall GH, et al. 5 alpha-reductase activity in polycystic ovary syndrome. *Lancet*. 1990; Feb 24; 335: 431-3.
36. Tsilchorozidou T, Honour JW, Conway GS. Altered cortisol metabolism in polycystic ovary syndrome: insulin enhances 5alpha-reduction but not the elevated adrenal steroid production rates. *J Clin Endocrinol Metab*. 2003; Dec; 88: 5907-13.
37. Gambineri A, Vicennati V, Genghini S, et al. Genetic variation in 11beta-hydroxysteroid dehydrogenase type 1 predicts adrenal hyperandrogenism among lean women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2295-302.
38. Goodarzi MO, Xu N, Azziz R. Association of CYP3A7*1C and serum dehydroepiandrosterone sulfate levels in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2008; Jul; 93: 2909-12.
39. Miller KK, Cai J, Ripp SL, et al. Stereo- and regioselectivity account for the diversity of dehydroepiandrosterone (DHEA) metabolites produced by liver microsomal cytochromes P450. *Drug Metab Dispos*. 2004; Mar; 32: 305-13.
40. Smit P, van Schaik RH, van der Werf M, et al. A common polymorphism in the CYP3A7 gene is associated with a nearly 50% reduction in serum dehydroepiandrosterone sulfate levels. *J Clin Endocrinol Metab*. 2005; Sep; 90: 5313-6.
41. Kuehl P, Zhang J, Lin Y, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet*. 2001; Apr; 27: 383-91.
42. Zhai G, Teumer A, Stolk L, et al. Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. *PLoS Genet*. 2011; Apr; 7: e1002025.

Supplementary note

To provide some insight in the role of the identified variants in *SULT2A1*, *STS* and *PAPSS2* in the general population, association of significant SNPs with DHEAS levels, DHEA levels and their ratio will be tested in women from the general population.

Methods

Study population

The Rotterdam Study is a large population-based study of elderly men and women from a defined area in Rotterdam the Netherlands. Data on DHEAS and DHEA levels were available in a randomly selected subset of women included in the Rotterdam Study, i.e. for DHEAS levels n=852; for DHEA levels, n=303; and for their ratio, n=290. DHEA and DHEAS were measured by double-antibody radioimmunoassays (RIA), Diagnostic Systems Laboratories, Webster, TX, USA. The ratio of DHEAS to DHEA was calculated.

Statistical analysis

SNPs mapping to the *SULT2A1*, *STS* and *PAPSS2* genes that were significantly associated with DHEAS, DHEA levels or their ratio in PCOS patients, were also tested in controls for association with DHEAS, DHEA levels or their ratio in subject from the general population. Linear regression analysis was carried out to determine the association of these SNPs in relation to DHEAS, DHEA levels and their ratio in control women from the Rotterdam Study. Bonferroni adjustment was applied for the number of SNPs tested. The statistical analyses were carried out using the IBM Statistical Package for the Social Sciences version 20 for Windows (IBM SPSS Inc.). Because there was only little overlap (n=95) between the subset of women with known DHEAS and DHEA levels in the Rotterdam Study and the women who were included as controls in the current study, and because a large age difference is present between PCOS cases (mean age 29.2 years) and controls (mean age 71.4 years) no direct comparisons will be made.

Results

In the controls from the general population mean levels of DHEAS were 2.7 µmol/l (± 2.1), DHEA 11.2 nmol/l (± 7.0) and the DHEAS to DHEA ratio was 0.30 (± 0.18). Of these women, n=95 were included as controls in the current study. In this set mean levels of DHEAS were 2.6 µmol/l (± 2.0), DHEA 11.4 nmol/l (± 6.4) and the DHEAS to DHEA ratio was 0.29 (± 0.19). SNP rs2910397 was highly significantly associated with the DHEAS to DHEA ratio in PCOS patients and was therefore the role of rs2910397 in controls from the general population was tested. These association results are depicted in Table 1 and 2.

Table 1 | Association of rs2910397 in *SULT2A1* in relation to DHEAS, DHEA and their ratio in controls form a population-based cohort.

	DHEAS (n=852)			DHEA (n=303)			DHEAS to DHEA ratio (n=290)		
	Effect	**SE**	***P* value**	**Effect**	**SE**	***P* value**	**Effect**	**SE**	***P* value**
rs2910397	-0.14	0.11	0.21	0.584	1.0	0.37	-0.04	0.02	0.04

DHEA, dehydroepiandrosterone; *DHEAS,* dehydroepiandrosterone sulfate; *SE*, standard error.

Table 2 | Association of rs2910397 in *SULT2A1* in relation to DHEAS, DHEA and their ratio in controls included in the current study.

	DHEAS (n=298)			DHEA (n=98)			DHEAS to DHEA ratio (n=95)		
	Effect	**SE**	***P* value**	**effect**	**SE**	***P* value**	**Effect**	**SE**	***P* value**
rs2910397	-0.04	0.18	0.84	-0.06	1.0	0.95	-0.04	0.03	0.25

DHEA, dehydroepiandrosterone; *DHEAS,* dehydroepiandrosterone sulfate; *SE*, standard error.

Chapter 4.3

BMI-associated Alleles do not Constitute Risk Alleles for Polycystic Ovary Syndrome Independent of BMI: A Case-control Study

Y.V. Louwers
N.W. Rayner
B.M. Herrera
L. Stolk
C.J. Groves
T.M. Barber
A.G. Uitterlinden
S. Franks
J.S.E. Laven*
M.I. McCarthy*

* contributed equally

Plos One 2014, accepted

Abstract

Introduction: Polycystic Ovary Syndrome (PCOS) has a strong genetic background and the majority of patients with PCOS have elevated BMI levels. The aim of this study was to determine to which extent BMI-increasing alleles contribute to risk of PCOS when contemporaneous BMI is taken into consideration.

Methods: Patients with PCOS and controls were recruited from the United Kingdom (563 cases and 791 controls) and The Netherlands (510 cases and 2720 controls). Cases and controls were of similar BMI. SNPs mapping to 12 BMI-associated loci which have been extensively replicated across different ethnicities, i.e. *BDNF*, *FAIM2*, *ETV5*, *FTO*, *GNPDA2*, *KCTD15*, *MC4R*, *MTCH2*, *NEGR1*, *SEC16B*, *SH2B1*, and *TMEM18,* were studied in association with PCOS within each cohort using the additive genetic model followed by a combined analysis. A genetic allelic count risk score model was used to determine the risk of PCOS for individuals carrying increasing numbers of BMI-increasing alleles.

Results: None of the genetic variants, including *FTO* and *MC4R*, was associated with PCOS independent of BMI in the meta-analysis. Moreover, no differences were observed between cases and controls in the number of BMI-risk alleles present and no overall trend across the risk score groups was observed.

Conclusion: In this combined analysis of over 4,000 BMI-matched individuals from the United Kingdom and the Netherlands, we observed no association of BMI risk alleles with PCOS independent of BMI.

Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder affecting up to 10% of females of reproductive age [1]. PCOS itself as well as its separate phenotypic characteristics demonstrate familial aggregation [2,3] and its heritability has been estimated as high as 65% [4]. Key features of the syndrome include ovulatory dysfunction, hyperandrogenism and polycystic ovarian morphology. Moreover, the majority of patients with PCOS have overweight or obesity [5]. It has been well established that obesity worsens the phenotype of PCOS. Lifestyle interventions and weight-loss not only improve body composition and insulin resistance in patients with PCOS, they also ameliorate the reproductive phenotype [6-8]. Although this close relationship between PCOS and obesity clearly exists, underlying mechanisms are far from being understood.

Obesity is also known to be a highly heritable trait [9,10] and its genetics have been widely and successfully investigated using genome wide association studies (GWAS) [11-15]. Since patients with PCOS have increased BMI levels compared to controls, genetic variants influencing PCOS might well include BMI risk alleles such that GWAS signals identified influencing PCOS might in fact be driven by primary effects on BMI. Therefore, it is of importance to evaluate whether adjustment or matching for BMI would eradicate the potential for variants influencing BMI to have an apparent effect on PCOS risk. If these effects remain after taking case-control differences in BMI into consideration, it might suggest either that a single BMI measurement is not an adequate proxy for lifelong BMI when it comes to specifying the effects of BMI on PCOS, or that the BMI risk alleles have pleiotropic effects on BMI as well as PCOS. The latter has been suggested for SNPs mapping to the *FTO* gene in association with obesity and type 2 diabetes mellitus in Asians [16,17].

Previous studies observed association of risk-alleles mapping to the *FTO* and *MC4R* gene with PCOS and its phenotypic characteristics [18-23]. However, these studies did not include BMI-matched case-control sets and had relatively small sample sizes (number of cases ranging from 65 to 800 and less than 1000 controls).

Therefore, we studied twelve BMI-associated loci in BMI-matched case-control sets from two large university medical centers to determine the effect on PCOS-susceptibility independently of current BMI.

Materials and Methods

Ethics statement

All clinical investigations were conducted according to the guidelines in the Declaration of Helsinki. The study was approved by the medical ethics committee from the Erasmus MC University Medical Centre. Approval for the UK study was obtained from the North Thames Multicenter Research Ethics Committee [MREC/99/2/45]). All subjects provided fully written informed consent.

Subjects

Independent northern European PCOS populations from the United Kingdom (UK) and the Netherlands were included in this study. The UK case-control set included a total of 1354 women, of whom 563 were diagnosed with PCOS and 791 served as controls. The case-control set from the Netherlands consisted of 510 patients diagnosed with PCOS and 2720 control women from the general population. BMI levels between cases and controls in both studies were similar ($P > 0.05$). Patients in both cohorts were diagnosed according to 2003 Rotterdam criteria [24]. The controls from the UK were population-based and recruited as part of the UK Blood Services (UKBS) set up by the Wellcome Trust Case Control Consortium (WTCCC) [25]. Control women from the Netherlands were derived from the Rotterdam Study, a population-based prospective cohort study [26]. In brief, this is a large population-based study of elderly subjects from a specific area near Rotterdam (Ommoord). All women aged 45 years or older at onset of menopause and with available DNA were included in the present analyses. These population-based control groups provided reference groups of allele frequencies which reflect the local general Northern European population, rather than being control groups wherein PCOS specifically was excluded. Patients and controls were of European descent.

Genotyping and quality control

Supplementary Table 1 summarizes the studied SNPs mapping to BMI-associated loci as identified by Frayling et al. [11], Loos et al. [13], Thorleifsson et al. [14] and Willer et al. [15]. These 12 loci were established as genome wide significant between the years 2007-2009 during the first waves of GWAS and have been replicated across several ethnic populations ever since [27]. The studied SNPs were the lead SNPs mapping to the BMI-associated loci, as described in the aforementioned papers [11,13-15]. Genotyping was carried out using Taqman "on demand"-assays (Applied Biosystems, Warrington, UK). The genotyping success rate was > 95%. The Rotterdam Study controls were genotyped using the Illumina 550k array and imputed using HapMap2 CEU reference panel [28]. Genotypes of six of the SNPs were derived from this data: rs4074134 mapping to *BDNF*, rs7138803 mapping to *FAIM2*, rs7647305 mapping to *ETV5*, rs10838738 mapping to *MC4R*, rs10913469 mapping to *SEC16B*, rs7498665 mapping to *SH2B1*. SNP rs11084753 had an imputation quality of 94%, rs6548238 of 95% and all other SNPs had an imputation quality of > 99%. The other SNPs were genotyped using Taqman "on demand"-assays (Applied Biosystems, Warrington, UK). None of the SNPs deviated from Hardy-Weinberg equilibrium (HWE).

Statistical analysis and power calculation

Association analyses were initially carried out within each case-control set separately. The additive genetic model was tested using PLINK (v.1.07) [29] and IBM SPSS version 20 (IBM Statistical Package for the Sociological Sciences Inc., Chicago, USA). The combined effect of the BMI-increasing alleles in the two populations was evaluated using a fixed-effects meta-analysis in GWAMA [30]for SNPs with heterogeneity (I^2) less than 25%. When I^2 exceeded 25%,

a random effect meta-analysis was performed using statistical software package R (http://www.r-project.org) [31]. Using Genetic Power Calculator software we determined that with the sample size of the total case control set (cases: n=1073; controls: n=3511), we reached approximately 95% power to detect association of a risk allele of frequency ≥ 0.2 having an odds ratio of ≥ 1.3 and an alpha of 0.05 (http://pngu.mgh.harvard.edu/~purcell/gpc/) [32]. Since the genetic variants were selected and not randomly tested, we did not correct for multiple testing and a $P < 0.05$ was considered statistically significant.

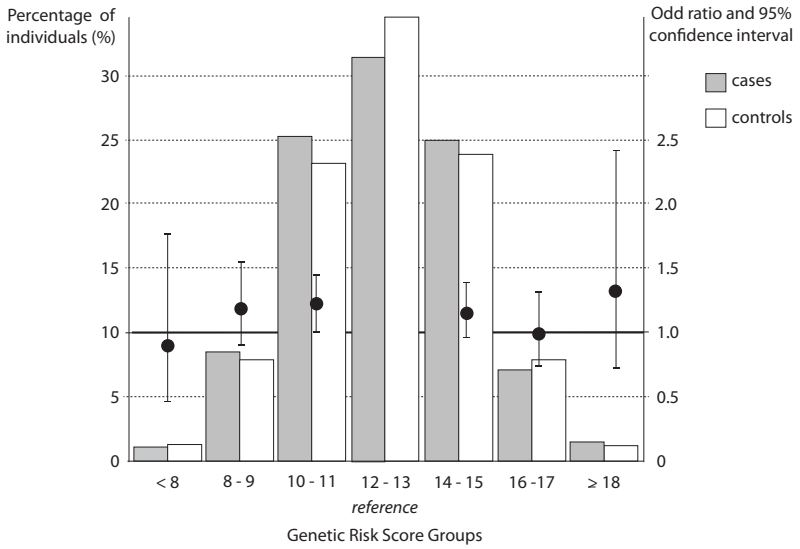
To test for the combined effect of all the BMI-associated alleles on PCOS susceptibility and to estimate the genetic risk of having PCOS for these women dependent on the number of BMI-increasing alleles present, we calculated the Genetic Risk Score (GRS). The GRS was modeled as a continuous variable and the calculation was carried out using R (http://www.r-project.org/) [31]. Using the GRS we assume that each SNP in the panel contributes equally to PCOS risk and that each individual allele has an equal and additive effect on risk. To obtain accurate counts of BMI-increasing alleles, only individuals with genotypes for at least 90% of SNPs (11 out of 12) were included. Based on this criterion, a total of 1264 individuals, i.e., 512 cases and 752 controls, from the UK and 3150 individuals from the Netherlands, i.e. 502 cases and 2648 controls, were included in the GRS-analysis. This method was described previously [33,34]. Missing genotypes were replaced with the average risk score for each SNP in the total population. The maximum attainable score was 24 BMI-increasing alleles (12 SNPs * 2 alleles). The reference group was defined as 12 to 13 BMI-increasing alleles, which was the mean number of BMI-increasing alleles present in the controls. Analyses were carried out within the separate case-control sets as well as in the combined set. Finally, we calculated the overall trend across the GRS-groups using the Kruskal Wallis trend test (IBM SPSS version 20).

Results

The UK cases had a mean BMI of 26.0 kg/m^2 (± 10.9 SD) and the controls had a BMI of 25.9 kg/m^2 (± 4.57 SD), whereas the cases from the Netherlands had a median BMI of 26.4 kg/m^2 (IQR 22.4-31.7) and the controls had a BMI of 26.3 kg/m^2 (IQR 23.97-29.1). Allele frequencies of the BMI-risk alleles in the cases and controls of both case-control sets are reported in supplementary Table 2. First, we tested whether carrying BMI-associated alleles influenced PCOS susceptibility in cases and controls who were of similar BMI (Table 1). SNP rs7498665 mapping to the *SH2B1* locus was significantly associated with a decreased risk of having PCOS (OR = 0.79; 95% CI 0.69-0.90, $P = 0.001$) in the case-control set from the Netherlands. No such association was evident in the case-control set from the UK (OR = 1.04; 95% CI 0.88-1.22, $P = 0.64$) or in the random effects meta-analysis. Moreover, none of the other SNPs was significantly associated with PCOS, neither in the separate UK and Dutch analysis nor in the meta-analysis.

To determine whether the overall burden of BMI-increasing alleles was associated with PCOS case-control status in these samples, a GRS was constructed (Figure 1, Table 2). The individuals carrying less than 8 BMI-associated alleles and individuals carrying over 18 BMI-associated alleles together account for a very small proportion of the total population, i.e. 2.6%. None of the GRS-groups was attributed more often to patients with PCOS compared with the reference GRS-group. Moreover, no overall trend in carrying an increasing number of BMI-associated alleles on PCOS susceptibility was observed neither for the total case-control set ($P = 0.44$), nor for the separate case-control sets from the UK ($P = 0.97$) or the Netherlands ($P = 0.17$).

Figure 1 | Combined impact of risk alleles on the risk of having PCOS compared to the reference risk group.



Along the X-axis the risk categories are shown based on the number of BMI-increasing alleles. The histogram (Y-axis on the left) indicates the percentage of individuals for each risk-score group. The odds ratio and confidence intervals calculated based on the risk of having PCOS compared to the reference risk group are plotted on the Y axis on the right.

Table 1 | Genetic association results for BMI-increasing risk alleles with PCOS in the United Kingdom and The Netherlands.

					United Kingdom			The Netherlands			Meta-analysis	
					cases n=563; controls n=791			cases=510; control=2720			cases=1073; control=3511	
SNP	Chr	Position	Locus name	Risk allele	Overall freq risk allele	OR per risk allele (95% CI)	*P*	Overall freq risk allele	OR per risk allele (95% CI)	*P*	OR per risk allele (95% CI)	*P*
rs4074134	11	27603861	*BDNF*	G	0.79	1.10 (0.91-1.33)	0.34	0.79	0.86 (0.74-1.02)	0.08	0.97 (0.76-1.23)*	0.79
rs7138803	12	48533735	*FAIM2*	A	0.38	1.05 (0.90-1.23)	0.53	0.37	1.02 (0.89-1.17)	0.80	1.03 (0.93-1.14)	0.54
rs7647305	3	187316984	*ETV5*	C	0.78	1.09 (0.90-1.31)	0.38	0.80	0.96 (0.82-1.14)	0.66	1.01 (0.90-1.15)	0.82
rs9939609*	16	52378028	*FTO*	A	0.42	1.17 (1.00-1.37)	0.05	0.36	1.01 (0.88-1.16)	0.87	1.08 (0.93-1.25)*	0.29
rs10938397	4	44877284	*GNPDA2*	G	0.45	0.99 (0.85-1.15)	0.87	0.42	1.15 (1.00-1.31)	0.05	1.06 (0.92-1.24)*	0.35
rs11084753	19	39013977	*KCTD15*	G	0.68	0.95 (0.80-1.11)	0.44	0.66	0.99 (0.86-1.14)	0.91	0.97 (0.87-1.08)	0.61
rs17782313	18	56002077	*MC4R*	C	0.24	1.06 (0.89-1.27)	0.51	0.25	1.12 (0. 96-1.31)	0.14	1.09 (0.97-1.23)	0.13
rs10838738	11	47619625	*MTCH2*	G	0.36	1.00 (0.85-1.17)	0.99	0.33	1.09 (0.95-1.26)	0.20	1.05 (0.94-1.17)	0.37
rs2815752	1	72585028	*NEGR1*	A	0.60	0.99 (0.85-1.17)	0.93	0.61	1.04 (0. 91-1.19)	0.56	1.02 (0.92-1.13)	0.72
rs10913469	1	176180142	*SEC16B*	C	0.21	0.91 (0.75-1.11)	0.37	0.20	0.91 (0.77-1.08)	0.30	0.90 (0.80-1.03)	0.15
rs7498665	16	28790742	*SH2B1*	G	0.39	1.04 (0.88-1.22)	0.64	0.40	0.79 (0.69-0.90)	0.001	0.90 (0.73-1.18)*	0.45
rs6548238	2	624905	*TMEM18*	C	0.84	1.05 (0.85-1.30)	0.64	0.83	1.00 (0.83-1.19)	0.95	1.02 (0.89-1.17)	0.77

* Random effect meta-analysis (I^2 > 25%), otherwise fixed effect meta-analysis was performed. *Chr,* chromosome; *Freq,* frequency; *SNP,* single nucleotide polymorphism; *OR,* odds ratio;, *CI,* confidence interval; *P, P* value. Risk allele is the BMI-increasing allele.

Table 2 | Combined impact of BMI-increasing alleles on the risk of having PCOS.

	United Kingdom			The Netherlands			Combined		
GRS	**Cases n (%)**	**Controls n (%)**	**OR (95%CI)**	**Cases n (%)**	**Controls n (%)**	**OR (95%CI)**	**Cases n (%)**	**Controls n (%)**	**OR (95%CI)**
<8	7 (1.4)	14 (1.9)	0.73 (0.29-1.86)	4 (0.8)	31 (1.2)	0.82 (0.28-2.34)	11 (1.1)	45 (1.3)	0.90 (0.46-1.76)
8 to 9	49 (9.6)	78 (10.4)	0.92 (0.61-1.38)	37 (7.4)	191 (7.2)	1.22 (0.83-1.81)	86 (8.5)	269 (7.9)	1.18 (0.90-1.55)
10 to 11	121 (23.6)	177 (23.5)	1.00 (0.74-1.36)	136 (27.1)	611 (23.1)	1.41 (1.09-1.81)	257 (25.3)	788 (23.2)	1.20 (1.00-1.45)
12 to 13	173 (33.8)	254 (33.8)	*Reference*	146 (29.1)	922 (34.8)	*Reference*	319 (31.5)	1176 (34.6)	*Reference*
14 to 15	125 (24.4)	171 (22.7)	1.07 (0.79-1.45)	129 (25.7)	641 (24.2)	1.27 (0.98-1.64)	254 (25.0)	812 (23.9)	1.15 (0.96-1.39)
16 to 17	32 (6.3)	52 (6.9)	0.90 (0.56-1.46)	40 (8.0)	216 (8.2)	1.17 (0.80-1.71)	72 (7.1)	268 (7.9)	0.99 (0.74-1.32)
> 18	5 (1.0)	6 (0.8)	1.22 (0.37-4.01)	10 (2.0)	36 (1.4)	1.75 (0.85-3.61)	15 (1.5)	42 (1.2)	1.32 (0.72-2.41)

Data are presented for pools of BMI-increasing alleles as odds ratio and confidence intervals. Only individuals with data of > 90% of the SNPs available were included. The mean number of BMI-increasing alleles in the controls was used as the reference group. *GRS*, genetic risk score; *n*, number of individuals; *OR*, odds ratio.

Discussion

In this combined analysis including > 4,000 patients with PCOS and controls from the United Kingdom and the Netherlands, we observed no association of genetic BMI-risk loci with PCOS when contemporaneous BMI is similar in cases and controls.

The last two years, genome wide association studies (GWAS) have emerged to identify genetic risk factors for PCOS. Two large studies identifying such PCOS-susceptibility loci in Han-Chinese patients have been published [35,36], while GWAS in patients from Northern European ancestry are in progress. In the current study, we observed no systematic effect of the BMI-associated alleles on PCOS susceptibility in our BMI-matched case-control set, indicating that BMI-associated alleles do not convey pleiotropic effects on PCOS risk. This infers that adjustment or matching for BMI will disentangle BMI-associated genetic signals to show up in PCOS GWAS and seems therefore not a genuine concern in the previous and upcoming GWASs in PCOS.

Presence of an increasing number of BMI-raising alleles is associated with an increased genetic predisposition to obesity. To determine the overall burden of BMI-associated alleles on risk for PCOS we calculated a counted genetic risk score and compared PCOS risk across such BMI risk groups. By doing so, we assumed that each allele has an equal and additive effect on PCOS risk. In practice some SNPs will have stronger effects than others. However, when the ORs are small as they are in our study, using a counted genetic risk model is appropriate [33,37]. None of the GRS-groups was attributed more often to the patients diagnosed with PCOS than controls compared to the reference GRS-group. Moreover, no overall trend across the consecutive GRS groups was observed, strengthening the results from the allelic-association analysis that BMI-associated alleles seem not to have pleotropic effects on PCOS risk. Increased BMI levels and weight in PCOS seem to be mediated by other genetic factors determining an individual's susceptibility to become obese or through modifying environmental effects. It has been shown that although women with PCOS reported a better, more healthy, dietary intake they seem to have an increased energy intake in combination with an increased sitting time without any discernible differences in total physical activity compared to women without PCOS [38]. Moreover, also in patients with PCOS, higher energy intake and glycaemic index and lower physical activity, as well as age, smoking, alcohol intake, occupation, education and country of birth, were independently associated with BMI [39].

A potential limitation of the current study is that we matched BMI based on a single measurement. In general the BMI increases in men as well as in women throughout life. Therefore, a single BMI measurement may not be an appropriate proxy for lifetime BMI and might be a poor estimate of the long-standing effects of BMI on PCOS risk. However, since we observed no systematic effect of BMI-associated alleles on PCOS risk in our matched case-control set, this seems not to have influenced our results tremendously. It has been previously observed that the association of genetic variants in *FTO* and *MC4R* with BMI and weight strengthen during childhood up to age 20 years and then become weaker with increasing age during adulthood [40]. Therefore, as has been suggested for phenotypes associated with type

2 diabetes mellitus [17], longitudinal studies are needed to adequately explore the complex and dynamic nature of BMI-associated alleles on cardiometabolic characteristics in PCOS.

In conclusion, we have shown in two independent large PCOS case-control sets matched for BMI, that there is no systematic effect of BMI-associated alleles on PCOS risk suggesting that these alleles do not have a pleiotropic effect on PCOS susceptibility. Hence, adjusting for BMI in PCOS case-control GWAS studies should be an effective strategy for removing confounding effects of BMI on the association of other genetic variants and PCOS.

References

1. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
2. Lunde O, Magnus P, Sandvik L, et al. Familial clustering in the polycystic ovarian syndrome. *Gynecol Obstet Invest*. 1989; 28: 23-30.
3. Kahsar-Miller MD, Nixon C, Boots LR, et al. Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. *Fertil Steril*. 2001; Jan; 75: 53-8.
4. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2100-4.
5. Barber TM, McCarthy MI, Wass JA, et al. Obesity and polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2006; Aug; 65: 137-45.
6. Moran LJ, Hutchison SK, Norman RJ, et al. Lifestyle changes in women with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2011: CD007506.
7. Holte J, Bergh T, Berne C, et al. Restored insulin sensitivity but persistently increased early insulin secretion after weight loss in obese women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 1995; Sep; 80: 2586-93.
8. Kiddy DS, Hamilton-Fairley D, Bush A, et al. Improvement in endocrine and ovarian function during dietary treatment of obese women with polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 1992; Jan; 36: 105-11.
9. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997; Jul; 27: 325-51.
10. Schousboe K, Willemsen G, Kyvik KO, et al. Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res*. 2003; Oct; 6: 409-21.
11. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007; May 11; 316: 889-94.
12. Lindgren CM, Heid IM, Randall JC, et al. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet*. 2009; Jun; 5: e1000508.
13. Loos RJ, Lindgren CM, Li S, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet*. 2008; Jun; 40: 768-75.
14. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*. 2009; Jan; 41: 18-24.
15. Willer CJ, Speliotes EK, Loos RJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*. 2009; Jan; 41: 25-34.
16. Li H, Kilpelainen TO, Liu C, et al. Association of genetic variation in FTO with risk of obesity and type 2 diabetes with data from 96,551 East and South Asians. *Diabetologia*. 2012; Apr; 55: 981-95.
17. Meyre D. Is FTO a type 2 diabetes susceptibility gene? *Diabetologia*. 2012; Apr; 55: 873-6.
18. Barber TM, Bennett AJ, Groves CJ, et al. Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia*. 2008; Jul; 51: 1153-8.
19. Kowalska I, Adamska A, Malecki MT, et al. Impact of the FTO gene variation on fat oxidation and its potential influence on body weight in women with polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2012; Jul; 77: 120-5.
20. Kowalska I, Malecki MT, Straczkowski M, et al. The FTO gene modifies weight, fat mass and insulin sensitivity in women with polycystic ovary syndrome, where its role may be larger than in other phenotypes. *Diabetes Metab*. 2009; Sep; 35: 328-31.
21. Tan S, Scherag A, Janssen OE, et al. Large effects on body mass index and insulin resistance of fat mass and obesity associated gene (FTO) variants in patients with polycystic ovary syndrome (PCOS). *BMC Med Genet*. 2010; 11: 12.

22. Wehr E, Schweighofer N, Moller R, et al. Association of FTO gene with hyperandrogenemia and metabolic parameters in women with polycystic ovary syndrome. *Metabolism*. 2010; Apr; 59: 575-80.
23. Yan Q, Hong J, Gu W, et al. Association of the common rs9939609 variant of FTO gene with polycystic ovary syndrome in Chinese women. *Endocrine*. 2009; Dec; 36: 377-82.
24. ESHRE ASRM Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.
25. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; Jun 7; 447: 661-78.
26. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol*. 2009; 24: 553-72.
27. Day FR, Loos RJ. Developments in obesity genetics in the era of genome-wide association studies. *J Nutrigenet Nutrigenomics*. 2011; 4: 222-38.
28. International HapMap Consortium. The International HapMap Project. *Nature*. 2003; Dec 18; 426: 789-96.
29. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; Sep; 81: 559-75.
30. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*. 2010; 11: 288.
31. R Development Core Team. R: A language and environment for statistical computing. Vienna Austria: R Foundation for Statistical Computing; 2013.
32. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003; Jan; 19: 149-50.
33. Lango H, Consortium UKTDG, Palmer CN, et al. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes*. 2008; Nov; 57: 3129-35.
34. Cornelis MC, Qi L, Zhang C, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Intern Med*. 2009; Apr 21; 150: 541-50.
35. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011; Jan; 43: 55-9.
36. Shi Y, Zhao H, Cao Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012; Sep; 44: 1020-5.
37. Janssens AC, Moonesinghe R, Yang Q, et al. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med*. 2007; Aug; 9: 528-35.
38. Teede HJ, Joham AE, Paul E, et al. Longitudinal weight gain in women identified With polycystic ovary syndrome: Results of an observational study in young women. *Obesity (Silver Spring)*. 2013; Aug; 21: 1526-32.
39. Moran LJ, Ranasinha S, Zoungas S, et al. The contribution of diet, physical activity and sedentary behaviour to body mass index in women with and without polycystic ovary syndrome. *Hum Reprod*. 2013; Aug; 28: 2276-83.
40. Hardy R, Wills AK, Wong A, et al. Life course variations in the associations between FTO and MC4R gene variants and body size. *Human molecular genetics*. 2010; Feb 1; 19: 545-52.

Supplementary Tables

Supplementary Table 1 | Studied BMI-associated loci.

Nearest gene	SNP	Chromosome	Position	Paper in which the SNP was first cited
BDNF	rs4074134	11	27603861	Thorleifsen et al.
FAIM2	rs7138803	12	48533735	Thorleifsen et al.
ETV5	rs7647305	3	187316984	Thorleifsen et al.
FTO	rs9939609	16	52378028	Frayling et al.
GNPDA2	rs10938397	4	44877284	Willer et al.
KCTD15	rs11084753	19	39013977	Willer et al.
MC4R	rs17782313	18	56002077	Willer et al.
MTCH2	rs10838738	11	47619625	Loos et al.
NEGR1	rs2815752	1	72585028	Willer et al.
SEC16B	rs10913469	1	176180142	Thorleifsen et al.
SH2B1	rs7498665	16	28790742	Willer et al.
TMEM18	rs6548238	2	624905	Willer et al.

SNP, Single Nucleotide Polymorphism.

Supplementary Table 2 | Allele frequencies in cases and controls from the United Kingdom and the Netherlands.

				United Kingdom				The Netherlands			
				Cases		Controls		Cases		Controls	
SNP	Nearest gene	allele 1	allele 2	frequency allele 1	frequency allele 2	frequency allele 1	frequency allele 2	frequency allele 1	frequency allele 2	frequency allele 1	frequency allele 2
rs4074134	*BDNF*	G	A	0.80	0.20	0.78	0.22	0.77	0.23	0.80	0.20
rs7138803	*FAIM2*	G	A	0.61	0.39	0.62	0.38	0.62	0.38	0.62	0.38
rs7647305	*ETV5*	C	T	0.79	0.21	0.77	0.23	0.79	0.21	0.80	0.20
rs9939609	*FTO*	T	A	0.54	0.46	0.58	0.42	0.62	0.38	0.63	0.37
rs10938397	*GNPDA2*	A	G	0.54	0.46	0.54	0.46	0.55	0.45	0.59	0.41
rs11084753	*KCTD15*	G	A	0.66	0.34	0.67	0.33	0.66	0.34	0.66	0.34
rs17782313	*MC4R*	T	C	0.75	0.25	0.76	0.24	0.73	0.27	0.75	0.25
rs10838738	*MTCH2*	A	G	0.64	0.36	0.64	0.36	0.64	0.36	0.67	0.33
rs2815752	*NEGR1*	A	G	0.60	0.40	0.61	0.39	0.61	0.39	0.60	0.40
rs10913469	*SEC16B*	T	C	0.80	0.20	0.79	0.21	0.81	0.19	0.79	0.21
rs7498665	*SH2B1*	A	G	0.61	0.39	0.62	0.38	0.63	0.37	0.58	0.42
rs6548238	*TMEM18*	C	T	0.85	0.15	0.84	0.16	0.83	0.17	0.83	0.17

SNP, Single Nucleotide Polymorphism

Chapter 5

Genome-Wide Association Study in Patients with PCOS from Northern European Descent

Y.V. Louwers
L. Stolk
C.B. Lambalk
D. Boomsma
B.C.J.M. Fauser
A.G. Uitterlinden
J.S.E. Laven

Preliminary results

Chapter 6

Cross-Ethnic Meta-analysis of Genetic Variants for Polycystic Ovary Syndrome

Y.V. Louwers
L. Stolk
A.G. Uitterlinden
J.S.E. Laven

Journal of Clinical Endocrinology & Metabolism. 2013 Dec;98(12):E2006-12.

Abstract

Context: Genome Wide Association Studies (GWAS) have revealed new susceptibility loci for Chinese patients with Polycystic Ovary Syndrome (PCOS). Because ethnic background adds to phenotypic diversities in PCOS, it seems plausible that genetic variants associated with PCOS act differently in various ethnic populations.

Objective: We studied cross-ethnic effects of Chinese PCOS loci (ie, *LHCGR, THADA, DENND1A, FSHR, c9orf3, YAP1, RAB5B/SUOX, HMGA2, TOX3, INSR, and SUMO1P)* in patients of Northern European descent.

Design: Genetic association study conducted at an University Medical Center.

Patients: Association was studied in 703 Dutch PCOS patients and 2164 Dutch controls. To assess the cross-ethnic effect, we performed a meta-analysis of the Dutch data combined with results of previously published studies in PCOS patients from China (n=2254) and the United States (n=2618) (US). Adjusted for multiple testing, a *P* value < 3.1×10^{-3} was considered statistically significant.

Results: Meta-analysis of the Chinese, US, and Dutch data resulted 12 significant variants mapping to the *YAP1* (*P* value = 1.0×10^{-9}), *RAB5B/SUOX* (*P* value = 3.8×10^{-11}), *LHCGR* (*P* value = 4.1×10^{-4}), *THADA* (*P* value = 2.2×10^{-4} and *P* value = 1.3×10^{-3}), *DENND1A* (*P* value = 2.3×10^{-3} and *P* value = 2.5×10^{-3}), *FSHR* (*P* value = 3.8×10^{-5} and *P* value = 3.6×10^{-4}), *c9orf3* (*P* value = 2.0×10^{-6} and *P* value = 9.2×10^{-6}), *SUMO1P1* (*P* value = 2.3×10^{-3}) loci with ORs ranging from 1.19 to 1.45 and 0.79 to 0.87.

Conclusions: Overall, we observed for 12 of 17 genetic variants mapping to the Chinese PCOS loci similar effect size and identical direction in PCOS patients from Northern European ancestry, indicating a common genetic risk profile for PCOS across populations. Therefore, it is expected that large GWAS in PCOS patients from Northern European ancestry will partly identify similar loci as the GWAS in Chinese PCOS patients.

Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrinopathy in women affecting approximately 6-10% of all reproductive-aged women. Patients with PCOS suffer from ovulatory dysfunction, have signs of hyperandrogenism and have polycystic morphology of the ovaries. Associated symptoms of PCOS include obesity, insulin resistance, dyslipidemia and type 2 diabetes leading to major health implications. It is a life-long condition with considerable variation of symptomatology between and within patients. PCOS itself and associated symptoms cluster in families [1]. Based on twin studies the heritability of PCOS is estimated to be around 65% [1]. Although, genetic variants in up to one hundred candidate genes have been found to be associated with PCOS in previous studies, the vast majority of these have not been replicated. Hitherto, only Genome Wide Association Studies (GWAS) of Han-Chinese PCOS patients have revealed new susceptibility loci for PCOS [2,3]. The top-signals of a recent Korean GWAS in PCOS patients did not reach the genome wide significance (gws) threshold [4]. The first GWAS on PCOS was conducted in a Han-Chinese population and identified variants mapping to three loci (ie, 2p16.3, 2p21 and 9q33.3) [2].

Subsequently, association of some of the variants in these susceptibility loci were confirmed in PCOS patients from the United States [5,6]. Recently, a second GWAS in a larger cohort of Han-Chinese PCOS patients discovered variants mapping to eight new risk loci (ie, 2p16.3, 9q22.32, 11q22.1, 12q13.2, 12q14.3, 16q12.1, 19p13.3 and 20q13.2) for PCOS [3]. Among the genes located in these loci were the *FSRH* gene and the *INSR* gene which both have been previously associated with PCOS in candidate gene studies (5,7]. Following the GWAS by Shi et al. [3], fine-mapping of the 2p16.3 region in PCOS patients from the United States reported evidence of this region for association with PCOS [8]. This is in line with what we and others previously demonstrated [7,9]. Moreover, genetic variation in the *FSHR* gene has been found to predict ovulation after clomiphene citrate treatment in infertile women [10].

It has been well established that ethnic background adds to the phenotypic diversities in PCOS patients [11-14]. Chinese patients with PCOS present themselves more often with an amenorrhea and hyperandrogenism compared with PCOS patients of Northern European ancestry [12]. Moreover, they seem to be at increased risk of metabolic complications in later life [12]. Differences in genetic risk profile might contribute to these phenotypic differences in PCOS patients from different ethnic populations. Thus, it seems plausible that genetic variants associated with PCOS act differently in diverse ethnic populations.

Therefore, the aim of this study was to assess the association of genetic variants identified in the studies by Chen et al. [2] and Shi et al. [3] in our Dutch PCOS cases and controls from Northern European ancestry. We performed a meta-analysis of the Dutch data with the data from the two previously published US studies. Finally, to determine the cross-ethnic effects of the genetic variants, we conducted a meta-analysis with the Chinese, US, and Dutch data combined.

Material and Methods

Subjects and phenotyping

All 703 Dutch PCOS patients of Northern European ancestry included in this study were recruited from a large University hospital in Rotterdam, the Netherlands. The patients visited the outpatient clinic and underwent a standardized physical examination and hormonal evaluation after an overnight fast. Clinical work-up included menstrual history as well as current cycle length, anthropometric measurements (height and weight) and calculation of body mass index (BMI) as weight in kilograms divided by the squared height in meters. Transvaginal ultrasonography was performed to assess ovarian volume and number of follicles in both ovaries. PCOS was diagnosed according to the 2003 Rotterdam consensus criteria [15]. In agreement with these criteria two of the following three criteria should be present: oligo-ovulation and/or anovulation with gonadotropins levels within the normal limits, biochemical and/or clinical hyperandrogenism and polycystic morphology of the ovaries. Oligomenorrhea was defined as a cycle length over 35 days and amenorrhea as absence of menstrual bleeding. Biochemical hyperandrogenism was determined by calculation of the Free Androgen Index as: 100 x T (testosterone) (nmol/L) / SHBG (nmol/L). A Free Androgen Index exceeding 4.5 was used as a cut-off. Clinical hirsutism was assessed using the modified Ferriman-Gallwey score and defined as a score of at least 8. Polycystic morphology of the ovaries was assessed by transvaginal ultrasound and defined as the presence of at least 12 follicles in one or both ovaries and/or increased ovarian volume > 10 ml. Exclusion criteria were presence of related disorders with similar presentations such as Cushing's disease and congenital adrenal hyperplasia.

Control women were recruited from a local general Dutch population of Northern European ancestry. The controls were drawn from the Rotterdam Study, a population-based prospective cohort study [16]. Briefly, it is a large population-based study of the elderly from an area near Rotterdam. All women of whom DNA was available, with an age of menarche between 10-17 years, age of menopause > 40 years and who had had at least two children were included. Data on menstrual cycle was available in only 20% of these women. Women who had reported an irregular menstrual cycle history were excluded. The included controls provide a reference group of the local general population rather than a control group in which PCOS was specifically excluded. This study was approved by the institutional medical ethical review board, and written informed consent was obtained from all participants.

Genotyping and quality control

Genomic DNA was extracted from peripheral venous blood samples according to standard procedures. Genotyping of the cases was performed using the Illumina HumanHap 610K beadarray. Controls were genotyped using the Illumina Infinium HumanHap 550K beadarray. Imputation of the genome-wide genotyped data of both cases and controls on all autosomes was performed using the MACH algorithm (http://www.sph.umich.edu/csg/abecasis/MACH)

based on phased autosomal chromosomes of the HapMap CEU Phase II panel (release September 2006 build 126). The genotypes of the 17 single nucleotide polymorphisms (SNPs) associated with PCOS in the Chinese as identified by Chen et al and Shi et al. [2,3], mapping to the *LHCGR*, *THADA*, *DENND1A*, *FSHR*, *c9orf3*, *YAP1*, *RAB5B/SUOX*, *HMGA2*, *TOX3*, *INSR*, *SUMO1P1* loci were retrieved from this genome-wide imputed genotypes database. Imputation quality of rs2349415 was 0.86; the imputation quality of rs6022786 was 0.87, and the imputation quality of rs2272046 was 0.92. All other SNPs were imputed with a quality above 0.95. Northern European ancestry of cases and controls was determined by Principal Component Analysis.

Statistical analysis and power calculation

Baseline characteristics of the PCOS cases are displayed as medians and interquartile ranges which were calculated using IBM SPSS Statistics version 20. To test for association with PCOS in our Northern European case-control sample we performed adjusted allelic association analysis using MACH2DAT software (http://www.sph.umich.edu/csg/abecasis/MaCH) implemented in BCSNPmax.

The first four Principal Components retrieved from principal component analysis, were used to correct for potential population stratification. Because a high correlation (i.e. $r^2 = 0.826$, $D' = 1$) between rs10986105 and rs10818854 exist in the Caucasian population, we adjusted for 16 independent SNPs. A *P* value < 3.1×10^{-3} (0.05/16) was considered statistically significant.

Genetic Power Calculator package (http://pngu.mgh.harvard.edu/~purcell/gpc/) was used to calculate the power to detect an association between the tested SNPs and PCOS [17]. We assumed effect estimates similar to those observed in the Chinese sample and used allele frequencies from the Hapmap CEU cohort [18]. As indicated in Supplementary Table 1, the power to detect effects of similar magnitude to the Chinese PCOS case-control sample was limited. To improve the statistical power to detect effects, we performed a meta-analysis with previously published data. SNPs mapping to the *THADA*, *DENND1A* and *LHCGR* loci identified by Chen et al. [2] have been previously tested in the Northern European case-control samples [5,6,19,20]. The studies by Eriksen et al. [19] and Lerchbaum et al. [20] only reported allele frequencies and *P* values in their published data, and were therefore not included in the Northern European meta-analysis. The GWAS in Korean PCOS patients by Hwang et al. [4] also evaluated the significance of the previously identified Chinese associated variants in their population. They describe that six of the seven tested genetic variants, had a significant *P* values between 2×10^{-2} and 8×10^{-4} [4]. However, neither in their manuscript, nor in the supplementary data, they report more detailed information. Therefore, this data was not included in the meta-analysis. Supplementary Table 2 summarizes the SNPs identified by Chen et al. [2] and Shi et al. [3] and replicated by Goodarzi et al. [5], Welt et al. [6] as well as the current study.

To assess the cross-ethnic effects of the Chinese susceptibility genetic variants for PCOS, we performed a meta-analysis of the Dutch data combined with results of previously published studies in PCOS patients from China, that is, the previously published data by Chen et al. [2], Shi

et al. [3], and the data from the United States, the studies by Goodarzi et al. [5], Welt et al. [6]. A fixed effect meta-analysis was conducted using GWAMA, a software tool for meta-analysis [21] for SNPs with heterogeneity (I^2) less than 25%. When I^2 exceeded 25%, a random effect meta-analysis was performed using statistical software package R (http://www.r-project.org). Because the study by Shi et al. [3] is an extension of the study by Chen et al. [2], we included for the meta-analysis only the results of the discovery phase to avoid overlap of patients.

Results

Baseline characteristics of the PCOS patients from the Dutch data and the previously published studies from China and the United States [2,3,5,6] are displayed in Table 1. The allele frequencies of each SNP in the Chinese, US and Dutch control samples as well as the allele frequencies from the Hapmap CEU and CHB populations [22] are depicted in Table 2. The majority of the studied SNPs differ in allele frequency in individuals from Chinese ancestry compared with individuals included in the US and Dutch control-samples, which were of Northern European ancestry. Table 3 summarizes the results of the Dutch study, the results of the previously published Chinese and US studies as well as the results of the meta-analyses. The strongest PCOS association in the Dutch data was with rs1894116 mapping to the *YAP1* locus (odds ratio [OR] = 1.37, *P* value = 1.85×10^{-3}). SNP rs705702 mapping to the *RAB5B/SUOX* locus (OR = 1.21, *P* value = 4.31×10^{-3}) was also associated with PCOS. However, most likely due to lack of power, this signal did not reach the threshold for significance. In the current study rs12468395 and rs12478601 mapping to the *THADA* locus, rs10986105 mapping to the *DENND1A* locus and rs4385527 mapping to the *c9orf3* locus were nominally significantly associated with PCOS susceptibility. The magnitude of the effects as well as the direction of the effect of these five SNPs was similar to those observed in the GWAS on Chinese PCOS patients.

In the fixed effect meta-analysis of the Dutch data and the United States, the studies by Goodarzi et al, Welt et al, rs10818854 mapping to the *DENND1A* locus was significantly (OR = 1.53, *P* value = 1.68×10^{-8}) associated with PCOS. However, heterogeneity of 0.82 was present amongst these studies for this SNP (Supplementary Table 3). In the random effect meta-analysis rs10818854 was no longer significant (*P* value = 4.9×10^{-3}). Two SNPs mapping to the *THADA* locus (ie, rs12468394 [OR = 0.87, *P* value = 1.01×10^{-5}] and rs12478601 [OR = 0.88, *P* value = 1.77×10^{-4}], as well as rs10986105 mapping to the *DENND1A* locus [OR = 1.56, *P* value = 1.53×10^{-5}]) were significantly associated with PCOS in the meta-analysis of the studies with samples from Northern European ancestry.

Table 1 | Baseline characteristics of PCOS patients from the study by Chen et al., Shi et al., Goodarzi et al., Welt et al. and the current study.

	China		United States					The Netherlands
	Chen et al.	Shi et al.	Goodarzi et al.		Welt et al.			Dutch data
	n=744	n=1510	cohort A: n=939	cohort B: n=535	cohort A: n=565	cohort B: n= 203	cohort C: n=376	n=703
Age, y	28.85 (± 3.62)	28.12 (± 2.75)	28 (24-32)	27 (22-32)	NA	NA	NA	29.1 (5.73)
BMI, kg/m^2	24.55 (± 3.99)	24.59 (± 3.17)	35.0 (29.1-42.5)	31.7 (24.9-39.6)	NA	NA	NA	24.8 (5.6)
T, nmol/L	2.81 (± 0.73)	2.03 (± 0.69)	2.5 (2.0-3.1)	2.3 (1.6-3.1)	NA	NA	NA	1.9 (1.1)
Insulin, pmol/L	NA	NA	132 (90-198)	84 (42-144)	NA	NA	NA	52 (53)
Glucose, mmol/L	NA	NA	4.9 (4.6-5.3)	4.8 (4.5-5.1)	NA	NA	NA	4.2 (0.8)

For the studies of Chen et al and Shi et al variables are presented as mean ± SD. For the studies of Goodarzi et al., Welt et al. and the current study, variables are presented as medians and interquartile ranges (IQR), unless otherwise indicated. *PCOS*, polycystic ovary syndrome; *BMI*, Body Mass Index; *T*, Testosterone. *NA*, not available.

Table 2 | Frequencies of risk alleles derived from the Hapmap CHB and Hapmap CEU as well as the allele frequencies control samples from the GWAS by Chen et al., the GWAS by Shi et al. and the current study.

		Chinese population			Northern European population	
SNP	**Allele**	**Hapmap CHB**	**Chen et al.**	**Shi et al.**	**Hapmap CEU**	**Dutch data**
		(n=90)	**(n=895)**	**(n=2106)**	**(n=90)**	**(n=2164)**
rs13405728	G	0.26	0.27	-	0.06	0.05
rs12468394	A	0.28	0.31	-	0.51	0.50
rs13429458	C	0.19	0.21	-	0.10	0.11
rs12478601	T	0.32	0.31	-	0.60	0.57
rs10818854	A	0.04	0.08	-	0.05	0.04
rs2479106	G	0.15	0.22	-	0.30	0.30
rs10986105	C	0.06	0.07	-	0.04	0.04
rs2268361	T	0.43	-	0.50	0.75	0.64
rs2349415	T	0.24	-	0.18	0.28	0.32
rs4385527	A	0.18	-	0.22	0.41	0.39
rs3802457	A	0.14	-	0.10	0.01	0.01
rs1894116	G	0.18	-	0.19	0.07	0.09
rs705702	G	0.30	-	0.25	0.28	0.30
rs2272046	C	0.08	-	0.09	0.03	0.03
rs4784165	G	0.30	-	0.33	0.24	0.25
rs2059807	G	0.33	-	0.30	0.64	0.61
rs6022786	A	0.35	-	0.34	0.43	0.42

SNP, single nucleotide polymorphism; *NA,* Not Available

Table 3 | Association analysis of the Dutch data and results of the cross-ethnic meta-analysis of the studies from China, the Netherlands and the USA.

			Asian ancestry						Northern European ancestry									
			China				Meta-analysis		United States				The Netherlands		Meta-analysis			
			Chen et al.		Shi et al.				Goodarzi et al.		Welt et al.		Dutch data				Cross-ethnic meta-analysis	
			PCOS n=744; controls n= 895		PCOS n= 1510; controls n= 2106		PCOS=2254; controls n=3001		PCOS n=1474; controls n= 1802		PCOS n=1144; controls n=17,619		PCOS n=703; controls n=2164		PCOS n=3321; controls n=21,585		PCOS n=5575; controls n=24,586	
SNP	nearby genes	All	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value
rs13405728	*LHCGR*	G	0.61	$2.5x10^{-7}$	0.81	$8.3x10^{-4}$	0.74	$3.8x10^{-9}$	0.83	0.10	0.87	0.34	0.92	0.58	0.86*	0.04	0.79**	$4.1x10^{-4}$
rs12468394	*THADA*	A	0.60	$1.2x10^{-9}$	0.78	$6.5x10^{-5}$	0.71	$2.5x10^{-12}$	0.84	$6.0x10^{-4}$	0.91	0.077	0.86	0.02	0.87*	$1.01x10^{-5}$	0.80**	$2.2x10^{-4}$
rs13429458	*THADA*	C	0.59	$1.1x10^{-7}$	0.71	$1.0x10^{-6}$	0.67	$4.2x10^{-13}$	0.93	0.39	0.95	0.60	0.86	0.15	0.91*	0.09	0.79**	0.01
rs12478601	*THADA*	T	0.61	$5.6x10^{-9}$	0.82	$6.2x10^{-4}$	0.75	$3.4x10^{-10}$	0.89†	0.02	0.92	0.18	0.88	0.04	0.88*	$1.77x10^{-4}$	0.82**	$1.3x10^{-3}$
rs10818854	*DENND1A*	A	1.80	$1.2x10^{-6}$	1.07	$4.2x10^{-1}$	1.30	$2.5x10^{-4}$	1.87	$9.8x10^{-8}$	1.53	$1.9x10^{-3}$	1.15	0.32	1.50**	$4.9x10^{-3}$	1.45**	$2.3x10^{-3}$
rs2479106	*DENND1A*	G	1.51	$5.1x10^{-7}$	1.26	$9.1x10^{-5}$	1.35	$5.1x10^{-10}$	1.04	0.51	1.05	0.45	0.97	0.68	1.02*	0.50	1.14**	0.05
rs10986105	*DENND1A*	C	2.08	$6.1x10^{-9}$	1.15	$1.2x10^{-1}$	1.42	$1.4x10^{-6}$	-	-	1.68	$3.3x10^{-4}$	1.45	0.01	1.56*	$1.53x10^{-5}$	1.54**	$2.5x10^{-3}$
rs2268361	*FSHR*	T	0.83	$6.7x10^{-3}$	0.84	$4.2x10^{-4}$	0.84	$8.8x10^{-6}$	-	-	-	-	0.94	0.33	-	-	0.87*	$3.8x10^{-5}$
rs2349415	*FSHR*	T	1.18	$1.6x10^{-1}$	1.37	$3.7x10^{-7}$	1.33	$2.6x10^{-7}$	-	-	-	-	1.15	0.05	-	-	1.24**	$3.6x10^{-4}$
rs4385527	*c9orf3*	A	0.82	$4.4x10^{-2}$	0.77	$6.8x10^{-5}$	0.78	$9.6x10^{-6}$	-	-	-	-	0.87	0.04	-	-	0.82*	$2.0x10^{-6}$
rs3802457	*c9orf3*	A	0.68	$4.3x10^{-3}$	0.70	$2.7x10^{-4}$	0.69	$3.8x10^{-6}$	-	-	-	-	0.90	0.77	-	-	0.70*	$9.2x10^{-6}$
rs1894116	*YAP1*	G	1.45	$1.4x10^{-5}$	1.23	$5.9x10^{-4}$	1.30	$1.1x10^{-7}$	-	-	-	-	1.37	$1.85x10^{-3}$	-	-	1.32*	$1.0x10^{-9}$
rs705702	*RAB5B,SUOX*	G	1.41	$1.9x10^{-5}$	1.28	$7.9x10^{-6}$	1.32	$1.1x10^{-9}$	-	-	-	-	1.21	$4.31x10^{-3}$	-	-	1.28*	$3.8x10^{-11}$
rs2272046	*HMGA2*	C	0.71	$6.8x10^{-3}$	0.65	$5.4x10^{-6}$	0.67	$1.4x10^{-7}$	-	-	-	-	1.19	0.36	-	-	0.79**	0.14
rs4784165	*TOX3*	G	1.09	$5.6x10^{-1}$	1.28	$2.0x10^{-6}$	1.26	$2.8x10^{-6}$	-	-	-	-	1.09	0.27	-	-	1.18**	$8.1x10^{-3}$
rs2059807	*INSR*	G	1.34	$1.2x10^{-4}$	1.19	$1.7x10^{-3}$	1.24	$1.6x10^{-6}$	-	-	-	-	0.93	0.27	-	-	1.13**	0.23
rs6022786	*SUMO1P1*	A	1.32	$1.7x10^{-4}$	1.21	$3.8x10^{-4}$	1.24	$4.1x10^{-7}$	-	-	-	-	1.06	0.38	-	-	1.19**	$2.3x10^{-3}$

All effect allele. Odds ratios (*OR*) and *P* values of the SNPs nearby *LHCGR, THADA, DENND1A, FSHR, c9orf3, YAP1, RAB5B/SUOX, HMGA2, TOX3, INSR and SUMO1P1* are depicted as reported in the discovery samples of the Chinese PCOS GWAS by Chen et al and Shi et al. *CI*, Confidence Interval. *P* value threshold was set to 3.1×10^{-3}. † Results of association with rs6544661 which is a perfect proxy for rs12478601 ($r^2=1$). * Fixed effect meta-analysis was performed when $I^2 < 25\%$. ** Random effect meta-analysis was conducted when $I^2 > 25\%$.

Meta-analysis of the Dutch, US and Chinese data resulted in 12 significant variants: rs13405728 (OR = 0.86, *P* value = 4.1x10-4) mapping to the *LHCGR* locus, rs12468394 (OR = 0.80, *P* value = $2.2x10^{-4}$) and rs12478601 (OR = 0.82, *P* value = 1.3x10-3) mapping to the *THADA* locus, rs10818854 (OR=1.45, *P* value = $2.3x10^{-3}$) and rs10986105 (OR = 1.54, *P* value = $2.5x10^{-3}$) mapping to the *DENND1A* locus, rs2268361 (OR = 0.87, *P* value = $3.8x10^{-5}$) and rs2349415 (OR = 1.24, *P* value = $3.6x10^{-4}$) mapping to the *FSHR* locus, rs4385527 (OR = 0.82, *P* value =2.0x10-6) and rs3802457 (OR = 0.70, *P* value = $9.2x10^{-6}$), mapping to the *c9orf3* locus, rs1894116 (OR = 1.32, *P* value = $1.0x10^{-9}$) mapping to the *YAP1* locus and rs705702 (OR = 1.28, *P* value = $3.8x10^{-11}$) mapping to the *RAB5B*/*SUOX* locus, rs6022786 (OR = 1.19, *P* value = $2.3x10^{-3}$) mapping to the *SUMO1P1* locus.

Discussion

For 12 genetic variants, mapping to the *LHCGR*, *THADA*, *DENND1A*, *FSHR*, *c9orf3*, *YAP1*, *RAB5B*/*SUOX* and *SUMO1P1* loci, identified by GWAS in Chinese PCOS patients we observed similar effect sizes and identical direction in patients of Northern European ancestry from the Netherlands and the United States. Since the publication of the first GWAS in Han-Chinese PCOS patients by Chen et al. [2], several research groups aimed to replicate the three identified loci (ie, *LHGCR*, *THADA* and *DENND1A)* in PCOS patients from Northern European ancestry. Adequately powered studies succeeded in validating several of the identified variants [5,11], whereas others were not able to replicate any of the Chinese findings [19,20]. Shi et al revealed eight new risk for PCOS [3], which hitherto have not been assessed in PCOS patients from Northern European ancestry. This is the first study replicating the association of rs1894116 mapping to the *YAP1* locus with PCOS in a case-control set from Northern European ancestry. After identification by Shi et al. [3], this variant is replicated in an independent sample of Han-Chinese PCOS patients [23]. The *YAP1* gene encodes for the protein YAP1 (Yes-associated protein 1) which is one of the transcriptional targets of the highly conserved Hippo pathway which controls organ size by regulating cell growth, proliferation and apoptosis [24]. Dysregulation of the Hippo pathway and overexpression of YAP resulted in tumor formation [25,26]. *YAP1* was also found to be expressed in porcine ovaries [27]. Because increased ovarian volume is an important clinical feature of PCOS, this might be regulated by the Hippo signaling pathway. However, translational studies are needed to confirm this hypothesis.

Another important key feature of PCOS is ovulatory dysfunction. Factors that play key regulatory roles in follicular growth, ovulation and luteinization are the pituitary gonadotropins FSH and LH as well as their receptors, *FSHR* and *LHCGR* respectively. In this cross-ethnic assessment, we observed that genetic variants mapping to the *FSHR* were associated with PCOS. Genetic variation in the *FSHR* gene, encoding the FSH receptor, has been extensively studied in association with ovulatory dysfunction. Previously, a higher prevalence of genetic variants in the *FSHR* gene in PCOS patients from Italy and the United States has been described

[7,8]. Moreover, it has been shown in studies with infertile patients from various ethnic ancestry that variants mapping to the *FSHR* influence FSH levels and can even predict ovarian response after ovulation induction treatment [7,9,10,28,29].

LHCGR encodes a G protein-coupled receptor for luteinizing hormone and human chorionic gonadotropin (hCG). In the ovary, *LHCGR* is expressed in granulosa cells at the later stages of pre-ovulatory follicles. Notably, the induction of *LHCGR* during granulosa cell differentiation allows the pre-ovulatory follicle to respond to the mid-cycle surge of luteinizing hormone, resulting in ovulation and release of the mature oocyte. In women, inactivating mutations of *LHCGR* are associated with increased luteinizing hormone, enlarged ovaries, oligomenorrhea, resistance to luteinizing hormone or HCG, and infertility [30]. In the current study, we were also able to relate genetic variants mapping to *LHCGR* to PCOS. Whether this gene, besides that it constitutes a risk allele for PCOS, shares a similar relationship with LH as well as androgen levels remains to be determined. Genetic variants mapping to *LHCGR* did not influence the characteristic clinical features, i.e. BMI, testosterone and homeostasis model assessment–insulin resistance (HOMA-IR), of PCOS [2]. Hence, in contrast to variants mapping *FSHR*, this variant seems not to account for other clinical phenotypes frequently found in PCOS.

Allele frequencies of the majority of risk alleles differ remarkably in the samples from Northern European ancestry compared with the sample from Asian ancestry. Previously, Ioannidis and colleagues [31] observed that genetic markers indeed vary in frequency across populations, however that their biological impact on the risk for common diseases may usually be consistent across these populations. Obviously, even if their biological impact on the risk is similar, larger samples sizes are needed when the allele frequency of the risk allele is low. We increased the power to detect associations by combining the results from US studies with the Dutch data and we confirmed the association of genetic variants mapping to *THADA* and *DENND1A*. In the cross-ethnic meta-analysis of the Chinese, US and Dutch data combined, 12 out of the 17 genetic variants were associated with PCOS. Genetic variants mapping to the *RAB5B/SUOX* and *YAP1* loci even reached the genome wide significant level. Another important issue when replicating GWAS findings from one population in other populations is dissimilarity of linkage disequilibrium (LD) across these populations. LD describes the correlation among neighboring alleles on the same chromosome [32]. It has been well established that these patterns of LD are different across populations [18,32]. The tested genetic variant will only give a significant association with the disease if it is strongly correlated with the functional genetic variant. When the functional genetic variant is located in another LD block because of population differences, it will not be detected with the tested variant. LD structure needs to be taken into account when studying association among different populations. However, while LD structure of the *THADA* locus is different for Northern Europeans compared with the Chinese, we were able replicate the Chinese finding. For the other loci the LD patterns were only slightly different.

Obviously, besides genetic factors, also the potential influence of non-genetic factors, such as dietary intake and environmental factors like exogenously derived compounds, might

be involved in the etiology of PCOS. Is has been well established that over-nutrition leading to obesity aggravates reproductive as well as metabolic components of the PCOS phenotype [33]. Moreover, high intake of dietary advanced glycation end products may also contribute to the pathogenesis of PCOS (34). A potential exogenously derived compound is Bisphenol A, which has been found to be associated with androgen levels and insulin resistance in PCOS [35]. Evidently, these non-genetic factors differ between ethnicities and might also underlay diversities in phenotype.

In this cross-ethnic assessment, we observed similar effect size and identical direction in PCOS patients from Northern European ancestry for 12 of the genetic variants mapping to the Chinese PCOS loci, indicating a common genetic risk profile for PCOS across populations. Because allele frequencies as well as phenotypic characteristics differ between PCOS patients from various ancestry, analyses in a large PCOS consortium, which we are conducting now, are needed to confirm and further explore ethnic-specific genetic profile. It is expected though, that large GWAS in PCOS patients from Northern European ancestry will at least partly identify similar loci as the GWAS in Chinese PCOS patients.

References

1. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2100-4.
2. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011; Jan; 43: 55-9.
3. Shi Y, Zhao H, Cao Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012; Aug 12; 44: 1020-5.
4. Hwang JY, Lee EJ, Jin Go M, et al. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet*. 2012; Oct; 57: 660-4.
5. Goodarzi MO, Jones MR, Li X, et al. Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *J Med Genet*. 2012; Feb; 49: 90-5.
6. Welt CK, Styrkarsdottir U, Ehrmann DA, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. *J Clin Endocrinol Metab*. 2012; Jul; 97: E1342-7.
7. Dolfin E, Guani B, Lussiana C, et al. FSH-receptor Ala307Thr polymorphism is associated to polycystic ovary syndrome and to a higher responsiveness to exogenous FSH in Italian women. *J Assist Reprod Genet*. 2011; Sep; 28: 925-30.
8. Mutharasan P, Galdones E, Penalver Bernabe B, et al. Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. *J Clin Endocrinol Metab*. 2013; Jan; 98: E185-90.
9. Laven JS, Mulders AG, Suryandari DA, et al. Follicle-stimulating hormone receptor polymorphisms in women with normogonadotropic anovulatory infertility. *Fertil Steril*. 2003; Oct; 80: 986-92.
10. Overbeek A, Kuijper EA, Hendriks ML, et al. Clomiphene citrate resistance in relation to follicle-stimulating hormone receptor Ser680Ser-polymorphism in polycystic ovary syndrome. *Hum Reprod*. 2009; Aug; 24: 2007-13.
11. Welt CK, Arason G, Gudmundsson JA, et al. Defining constant versus variable phenotypic features of women with polycystic ovary syndrome using different ethnic groups and populations. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4361-8.
12. Guo M, Chen ZJ, Eijkemans MJ, et al. Comparison of the phenotype of Chinese versus Dutch Caucasian women presenting with polycystic ovary syndrome and oligo/amenorrhoea. *Hum Reprod*. 2012; May; 27: 1481-8.
13. Valkenburg O, Lao O, Schipper I, et al. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *J Clin Endocrinol Metab*. 2011; Jul; 96: E1181-7.
14. Carmina E, Koyama T, Chang L, et al. Does ethnicity influence the prevalence of adrenal hyperandrogenism and insulin resistance in polycystic ovary syndrome? *Am J Obstet Gynecol*. 1992; Dec; 167: 1807-12.
15. ESHRE ASRM Sponsored PCOS Consensus Workshop Group.. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.
16. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol*. 2011; Aug; 26: 657-86.
17. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003; Jan; 19: 149-50.
18. de Bakker PI, Burtt NP, Graham RR, et al. Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet*. 2006; Nov; 38: 1298-303.
19. Eriksen MB, Brusgaard K, Andersen M, et al. Association of polycystic ovary syndrome susceptibility single nucleotide polymorphism rs2479106 and PCOS in Caucasian patients with PCOS or hirsutism as referral diagnosis. *Eur J Obstet Gynecol Reprod Biol*. 2012; Jul; 163: 39-42.

20. Lerchbaum E, Trummer O, Giuliani A, et al. Susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21, and 9q33.3 in a cohort of Caucasian women. *Horm Metab Res*. 2011; Oct; 43: 743-7.
21. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*. 2010; 11: 288.
22. International HapMap Consortium. The International HapMap Project. *Nature*. 2003; Dec 18; 426: 789-96.
23. Li T, Zhao H, Zhao X, et al. Identification of YAP1 as a novel susceptibility gene for polycystic ovary syndrome. *J Med Genet*. 2012; Apr; 49: 254-7.
24. Edgar BA. From cell structure to transcription: Hippo forges a new path. *Cell*. 2006; Jan 27; 124: 267-73.
25. Zhao B, Wei X, Li W, et al. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev*. 2007; Nov 1; 21: 2747-61.
26. Hall CA, Wang R, Miao J, et al. Hippo pathway effector Yap is an ovarian cancer oncogene. *Cancer Res*. 2010; Nov 1; 70: 8517-25.
27. Agca C, Ries JE, Kolath SJ, et al. Luteinization of porcine preovulatory follicles leads to systematic changes in follicular gene expression. *Reproduction*. 2006; Jul; 132: 133-45.
28. Sudo S, Kudo M, Wada S, et al. Genetic and functional analyses of polymorphisms in the human FSH receptor gene. *Mol Hum Reprod*. 2002; Oct; 8: 893-9.
29. Falconer H, Andersson E, Aanesen A, et al. Follicle-stimulating hormone receptor polymorphisms in a population of infertile women. *Acta Obstet Gynecol Scand*. 2005; Aug; 84: 806-11.
30. Arnhold IJ, Lofrano-Porto A, Latronico AC. Inactivating mutations of luteinizing hormone beta-subunit or luteinizing hormone receptor cause oligo-amenorrhea and infertility in women. *Horm Res*. 2009; 71: 75-82.
31. Ioannidis JP, Ntzani EE, Trikalinos TA. 'Racial' differences in genetic effects for complex diseases. *Nat Genet*. 2004; Dec; 36: 1312-8.
32. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005; Oct 27; 437: 1299-320.
33. Lim SS, Norman RJ, Davies MJ, et al. The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis. *Obes Rev*. 2013; Feb; 14: 95-109.
34. Diamanti-Kandarakis E, Katsikis I, Piperi C, et al. Increased serum advanced glycation end-products is a distinct finding in lean women with polycystic ovary syndrome (PCOS). *Clin Endocrinol (Oxf)*. 2008; Oct; 69: 634-41.
35. Kandaraki E, Chatzigeorgiou A, Livadas S, et al. Endocrine disruptors and polycystic ovary syndrome (PCOS): elevated serum levels of bisphenol A in women with PCOS. *J Clin Endocrinol Metab*. 2011; Mar; 96: E480-4.

Chapter 7

The Impact of Self-reported Ethnicity Versus Genetic Ancestry in Patients with Polycystic Ovary Syndrome (PCOS)

Y.V. Louwers
O. Lao
B.C.J.M. Fauser
M. Kayser
J.S.E. Laven

Manuscript submitted.

Abstract

Context: It has been well established that ethnicity is associated with the phenotype of Polycystic Ovary Syndrome (PCOS). However, self-reported ethnicity has been shown to be an inaccurate proxy for ethnic origin in other disease traits, and it is unclear how in PCOS patients self-reported ethnicity compares with a biological proxy such as genetic ancestry.

Objective: We analyzed the association of self-reported ethnicity in patients with PCOS to genetic ancestry based on genome-wide array data. Moreover, we determined which of these classifications better predicts the variability in endocrine and metabolic characteristics of PCOS better.

Patients: 1499 Dutch PCOS patients from the Netherlands, comprised of 11 self-reported ethnic groups of European, African, American and Asian descent were genotyped with the Illumina 610K Quad BeadChip and merged with the Illumina data available from HGDP-CEPH reference panel including 53 worldwide populations for ancestry inference.

Main outcome measures: Algorithms for inferring genetic relationships among individuals, including Multi-Dimensional-Scaling (MDS) and ADMIXTURE, were applied to describe and recover the genetic ancestry for each individual. Regression analysis was used to determine the best predictor for the variability in PCOS phenotypic and sub-phenotypic characteristics.

Results: The association between self-reported ethnicity and genetic ancestry as derived from genome-wide data was moderate, i.e. Cramer's V = 0.77. For amenorrhea, total follicle count, BMI, SHBG, DHEAS and insulin, mainly genetic ancestry clusters ended up in the final models (*P* value < 0.004), indicating that they explain a larger proportion of variability of these PCOS characteristics compared to self-reported ethnic background. Especially the variability of insulin levels seems predominantly influenced by genetic ancestry (hypergeometric probability 0.02).

Conclusions: Self-reported ancestry is not a perfect proxy for genetic ancestry in patients with PCOS, emphasizing that by using genetic ancestry instead of self-reported ethnicity PCOS-relevant misclassification can be avoided. Moreover, as genetic ancestry explained a larger proportion of phenotypic variability associated with PCOS than self-reported ethnicity, future studies should also focus on genetic ancestry verification of PCOS patients for research questions, treatment as well as preventive strategies in these women.

Introduction

Polycystic Ovary Syndrome (PCOS) is the most common endocrine disorder in women of reproductive age [1]. The broad spectrum of symptoms of PCOS encompasses reproductive endocrine as well as metabolic features [2,3]. Reproductive characteristics include anovulation, hyperandrogenism and polycystic ovarian morphology. Patients with PCOS also suffer from obesity, dyslipidemia and insulin resistance. Moreover, they have an increased risk of developing type 2 diabetes mellitus and cardiovascular disease later in life.

It has been well established that ethnic origin influences the variability of the phenotype [4,5]. For instance, Mexican American patients suffer more often from insulin resistance and have higher BMI levels compared to European American patients with PCOS [6]. A higher prevalence of hirsutism was observed among PCOS patients from the Middle East compared to those of European ancestry [7]. Asian PCOS patients have more often insulin resistance than European PCOS patients but seem less likely to become obese than European PCOS patients and European American PCOS patients [6,8]. However, another study observed a higher incidence of increased BM and waist circumference in Chinese patients with PCOS compared to European patients [9]. African American and Hispanic patients with PCOS have elevated BMI levels and an increased waist circumference compared to European and Asian patients [10,11]. Because of such observations, it has been previously suggested that there is need for development of ethnicity-specific guidelines for identifying anthropometric thresholds for appropriate screening, diagnosis and treatment in high-risk ethnic groups [4,5]. However, the definition of ethnicity is complex including non-genetic elements such as diet, language and other cultural elements as well as elements that are expected to have a genetic component such as bio-geographic origin (i.e. genetic ancestry) or the genetic background in general, which may or may not affect the variability of the PCOS phenotype including sub-phenotypes [10,12,13].

Usually in medical studies, including those on PCOS, ethnicity is assessed by self-reported data. However, self-reported ethnicity can be biased and does not necessarily capture all existing population substructure [14,15]. For instance, we have previously shown that a set of 10 ancestry informative markers is effective to map phenotypic differences in anovulatory patients at a continental level [13]. However, a more accurate resolution of the geographic ancestry of the patients would be desirable. Moreover, one of the proposed future directions for research during the consensus meeting on women's health aspects of PCOS was the role of genetic and environmental factors in explaining differences in PCOS incidence as well as in PCOS associated characteristics among different populations [8].

In the current study, we determine genetic ancestry of a multi-ethnic set of PCOS patients by means of high-resolution genome-wide microarray analysis and by using a worldwide reference dataset (HGDP-CEPH) – overall 152,375 independent single nucleotide polymorphisms (SNPs). We applied two clustering algorithms on the genome-wide SNP data and correlated self-reported ethnicity with derived genetic ancestry. Moreover, we compared

PCOS-associated sub-phenotypic differences among the various groups as defined by self-reported ethnicity as well as by genetic ancestry. Finally, we determined whether classification based on self-reported data or genetic ancestry explains more variation in endocrine and metabolic characteristics of PCOS patients.

Materials and Methods

PCOS diagnosis and phenotyping

All patients with PCOS included in this study were recruited from a large University hospital in Rotterdam, the Netherlands. The patients underwent a standardized physical examination and hormonal evaluation after an overnight fast. Clinical work-up included menstrual history as well as current cycle length, anthropometric measurements (height and weight) and calculation of body mass index (BMI) as weight in kilograms divided by squared height in meters. Self-reported ethnicity and country of birth were registered prior to the clinical investigation via a questionnaire. This information was verified by the research investigator at the day of investigation. Transvaginal ultrasonography was performed to assess ovarian volume and number of follicles in both ovaries.

PCOS was diagnosed according to the 2003 Rotterdam consensus criteria. In agreement with these criteria two of the following three criteria should be present: oligo-ovulation and/or anovulation with gonadotropins levels within the normal limits, biochemical and/or clinical hyperandrogenism and polycystic morphology of the ovaries (PCOM). Oligomenorrhea was defined as a cycle length over 35 days and amenorrhea as absence of menstrual bleeding. Biochemical hyperandrogenism was determined by calculation of the Free Androgen Index (FAI) as: 100 x [T (nmol/L) / SHBG (nmol/L)]. A FAI exceeding 4.5 was used as a cut-off. Clinical hirsutism was assessed using the modified Ferriman-Gallwey score and defined as an FG-score of at least 8. PCOM was assessed by transvaginal ultrasound and defined as the presence of at least 12 follicles in one or both ovaries and/or increased ovarian volume > 10 ml. Exclusion criteria were presence of related disorders with similar presentations such as Cushing's disease and congenital adrenal hyperplasia. This study was approved by the institutional medical ethical review board, and written informed consent was obtained from all patients.

Hormone assays

Blood samples were drawn after an overnight fast, before 10.00 am on the day of clinical examination and processed within 2h after withdrawal. Until assayed, serum was stored at -20° C. Serum levels of SHBG, androstenedione, insulin and DHEAS were measured by immunoluminometric assay (Immulite® platform, Siemens Diagnostic Products Corporation (DPC), Los Angeles, CA). Testosterone was determined by radioimmunoassay (RIA) (Siemens DPC, Los Angeles, CA). DHEA was measured using RIA (Diagnostic Systems Laboratories, Webster, TX). Glucose levels were measured using a Hitachi 917 analyzer (Roche Diagnostics,

Almere, The Netherlands). Intra- and interassay coefficients of variation (CV) were less than 4 and 5% for SHBG, less than 8 and 11% for androstenedione, less than 6 and 8% for insulin, less than 9 and 11% DHEAS, and, less than 3 and 5% for testosterone.

Genome-wide data and quality control

Genomic DNA was extracted from peripheral venous blood samples according to standard procedures. Each patient with PCOS was genotyped with the Illumina HumanHap 610K beadarray. SNPs with more than 5% of missing genotypes were excluded. Subsequently, we pruned for linkage disequilibrium (LD) by means of ascertaining markers that showed low LD at a distance of < 500 kb using the plink software [16] option plink --indep 50 5 2. This command prunes based on the variance inflation factor (VIF). The VIF is $1/(1-r^2)$ where r^2 is the multiple correlation coefficient for a SNP being regressed on all other SNPs simultaneously. A VIF of 1 would imply that the SNP is completely independent of all other SNPs. The parameters for --indep are: window size in SNPs, i.e. 50, the number of SNPs to shift the window at each step, i.e. 5, and the VIF threshold, i.e. 2.

In total, 166,194 Linkage Disequilibrium-pruned autosomal SNPs passed these quality steps and were included for further analysis. We included genome-wide data of 938 samples of CEPH's Human Genetic Diversity Cell Line Panel (HGDP-CEPH) from 53 populations with known bio-geographic ancestry as a reference [17]. After merging with the data publicly available for the HGDP-CEPH samples, a total number of 152,375 autosomal SNPs was used for the analyses.

Statistical analysis

Genetic ancestry

In order to determine the genetic ancestry of patients with PCOS based on the genome-wide SNP data, two different analyses were performed. The first method was algorithm ancestry based [18]. Plink software was used to compute an Identical by State (IBS) matrix between all the individuals. The cmdscale transformation function in R software [19] was used to iteratively perform a classical Multi-Dimensional Scaling (MDS) with the IBS distance matrix of the HGDP-CEPH panel and each patient independently, retaining the first 10 dimensions. These dimensions were then used in the MCLUST software implemented in R [19] to define groups of individuals by fitting multivariate normal distributions on the proposed dimensions (Supplementary Figure 1). MCLUST has been previously used (http://www.maths.bris.ac.uk/~madjl/finestructure/Lawson2012-GeneticSimilarityClustering.pdf) in the output of other multivariate techniques [20].

The second analysis of genetic ancestry was model-based ancestry estimation [18] and attempted to estimate ancestry proportions of each individual from a set of putative ancestral populations. ADMIXTURE software was used with the complete dataset setting the number of clusters (K) from 1 to 12 [18]. The best K was estimated by the cross-error estimation implemented in the ADMIXTURE software. For the best K a ten times rerun was performed. In

order to get a categorical classification of the PCOS individuals, each patient was assigned to the cluster with the highest proportion of admixture.

Genetic ancestry assignations were compared to self-reported ethnicity using the Cramer's V-statistic in R [19]. In order to quantify the cluster diversity within each self-reported population, we computed the entropy of each population *i* estimated as [21].

$$H_i = -\sum_{k=1}^{K} p_k \log(p_k)$$

Where p_{ik} is the proportion of individuals of population *i* that are assigned to cluster *k*. The value of entropy ranges from 0, which indicates no cluster diversity within i (i.e. all the individuals of a self-reported population belongs to the same cluster), to the log(K), which is the maximum value of diversity.

Association of PCOS phenotypic characteristics with ethnicity and genetic ancestry

Analysis of covariance adjusted for age was used to test for significant differences between the PCOS phenotypic characteristics of the ethnic subgroups based on genetic data. Chi-square test was used to test for differences of categorical variables. Variables were checked for normal distributions with the one-sample Kolmogorov-Smirnov test and log- or square-transformation were applied. Extreme outliers were excluded. Patients using oral contraceptives or other hormonal medication were excluded from the phenotypic analyses.

To determine which of the two ethnicity assignations, i.e. assignation based on self-reported data versus assignation based on genome-wide array data, explains the largest variability in phenotypical characteristics of PCOS, we computed a step-wise forward linear regression model for continuous variables and a step-wise forward logistic regression model for categorical variables. Age, all self-reported ethnicity clusters and all genetic ancestry clusters based on MCLUST cluster algorithm were included as independent variables in the model. The PCOS characteristics were included as dependent variables. Adjusted for multiple testing, a p-value of < 0.004 (0.05/12 variables) was considered to be statistically significant. If predominantly genetic clusters will end up in the final models, one can state that genetic clusters explain a larger proportion of variability compared to self-reported ethnicity.

Subsequently, using multivariate hypergeometric distribution, we calculated the probability of finding, at random, an equal or larger number of genetic clusters assuming a similar number of significant observations as resulted from the regression analysis. A one-tailed p-value < 0.025 is considered significant. Finally, a combined p-value was computed with the Fishers' method [22].

Results

All patients with PCOS were classified according to their self-reported ethnic origin: 66.7% (n=1000) Europeans, 8.8% (n=132) Turkish, 3.7% (n=55) Moroccan, 3.2% (n=48) Caribbean, 2.9% (n=44) Surinam Creole, 5.5% (n=83) Surinam Hindustani, 3.7% (n=55) Sub-Saharan Africans, 2.0% (n=30) Central-East Asians, 1.2% (n=18) Middle East, 0.9% (n=14) South Americans, and 1.3% (n=20) admixed. Countries of birth accompanying the self-reported ethnic origin are listed in Supplementary Table 1.

Genetic clustering: MCLUST

The first two dimensions of the 1499 patients computed by classical MDS analysis controlling for the HGDP-CEPH samples are plotted in Figure 1. In the first cluster analysis, we applied the MCLUST cluster algorithm to the first 10 dimensions of the MDS. This resulted in identification of 8 clusters based on the genomic data (Table 1). Cluster 1 comprises primarily self-reported European patients (90%, n=900). Most self-reported Turkish patients (n=124, 93.9%) are allocated to cluster 2, whereas most self-reported patients from Morocco (n=52, 94.5%) are assigned to cluster 5. Self-reported Surinam Creole patients are distributed over cluster 6 (n=18, 40.9%) and cluster 8 (n=17, 38.6%). Cluster 6 includes the majority of self-reported Surinam Hindustani patients with PCOS (n=76, 91.6%), whereas most of the self-reported Caribbean patients (n=41, 85.4%) and of the self-reported Sub-Sahara African patients (n=40, 72.7%) are allocated to cluster 8. Cluster 7 includes the majority of self-reported Central-East Asian patients (n=23, 76.7%). Cluster 4, and to a lesser extent cluster 3, include patients from nearly all self-reported ethnicities. Comparison of self-reported ethnicity with the genetic ancestry assignations as derived from MDS and MCLUST resulted in a Cramer's V of 0.77, indicating a moderate association between the self-reported ethnicity and genetic ancestry. Patients self-reported from the Middle East, the Surinam Creoles and those of admixed self-reported ethnicity showed the highest measure of genomic diversity as indicated by the high entropy values (entropy = 0.60, 0.51 and 0.45 respectively), which may be expected given their group definition.

Genetic clustering: ADMIXTURE

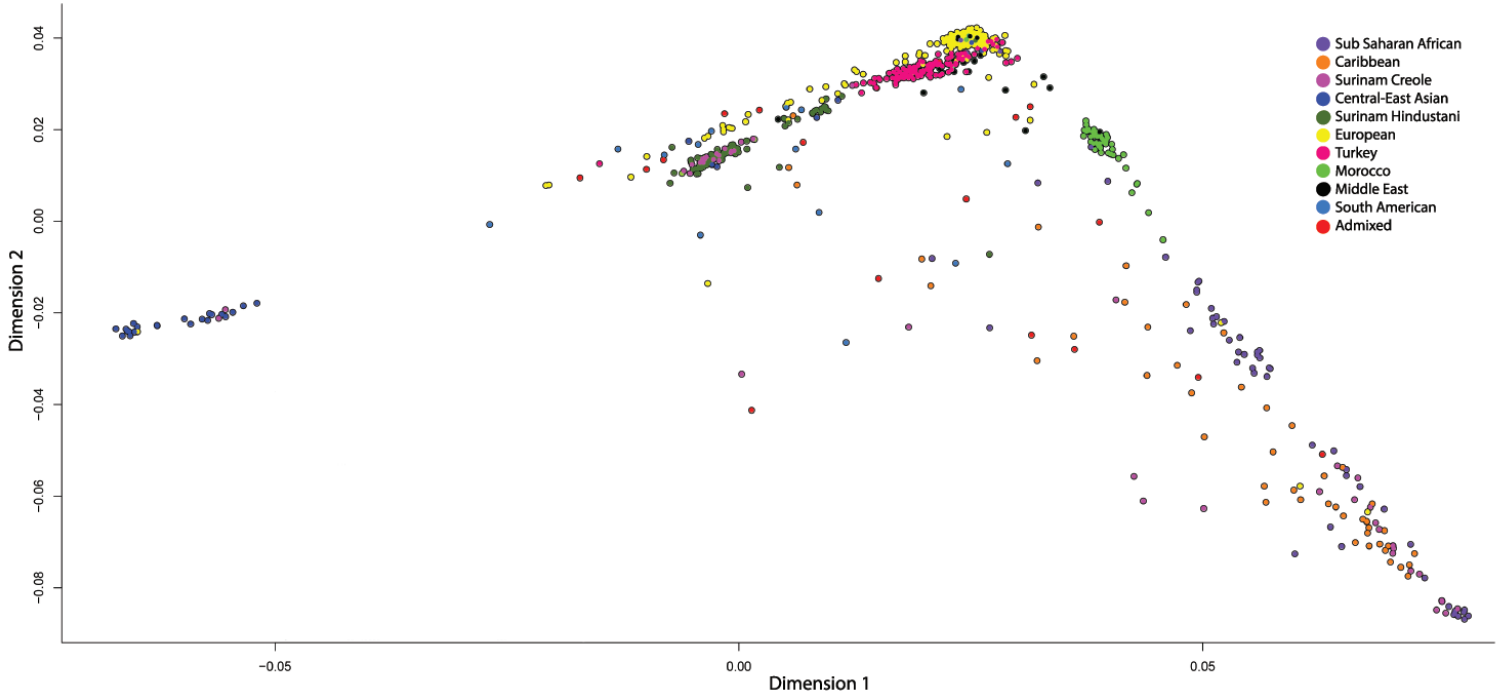
For the second cluster analysis, we used the cluster algorithm implemented in ADMIXTURE to estimate the genetic ancestry proportions of each individual according to K putative ancestral populations. By performing a cross-validation error analysis, we observed that at K=10 the cross-validation error was the smallest indicating that K = 10 reflects the most statistically supported number of ancestral populations (Supplementary Figure 2). Individual ancestry as estimated by ADMIXTURE using K = 10 is plotted in Figure 2. Patients were assigned to one of the clusters based on the highest proportion of detected ancestry. None of the patients were assigned to cluster 1 or cluster 8 (Table 2). Almost all self-reported Turkish patients (n=128, 97.0%) are allocated to cluster 2. All self-reported Moroccan patients, except for one (n=54,

98.2%) are gathered in cluster 3 and nearly all self-reported Surinam Hindustani are in cluster 4 (n=78, 94.0%). Cluster 5 is a mix of self-reported Caribbean, Surinam Creole and Sub Saharan Africans. Only one patient, a self-reported Turkish, was assigned to cluster 6. Self-reported Central-East Asians are primarily assigned to cluster 7 (n=23, 76.7%). Cluster 10 encompassed almost all self-reported European patients (n=987, 98.7%). However, this cluster also includes patients from all other self-reported ethnicities. The association of the self-reported ethnicity and the genetic ancestry assignations as derived from ADMIXTURE was with 0.77 (Cramer's V) the same as obtained when comparing self-reported ethnicity with genetic ancestry derived from MDS/MCLUST indicating a moderate association. As expected and as noted in the MDS/MCLUST analysis, the self-reported Surinam Creole, the admixed and self-reported Middle East group also in the ADMIXTURE analysis showed the highest degree of genetic diversity based on the entropy values (entropy = 0.43; 0.53 and 0.37, respectively), whereas the self-reported Europeans and Turkish patients were both genetically very homogeneous as reflected by low entropy values (entropy = 0.02 and 0.07, within both groups respectively).

PCOS phenotype, self-reported ethnicity, and genetic ancestry

Then, we determined whether self-reported ethnicity or genetic ancestry explained a larger proportion of variability of different PCOS characteristics. Besides age, the 11 self-reported ethnicities and the 8 genetic ancestry clusters obtained via MDS/MCLUST analysis were included as independent variables in the regression models. Table 3 and Figure 3 display the statistically significant results. For seven PCOS associated characteristics (amenorrhea, TFC, BMI, FAI, SHBG, DHEAS, and insulin) we revealed significant effects for at least one genetic cluster, while significant effects of self-reported ethnicities were seen in only four of these PCOS characteristics(amenorrhea, FAI, SHBG, and adion). For three characteristics (amenorrhea, FAI, and SHBG) we observed significant effects for more genetic clusters than for self-reported ethnicities, respectively. For two characteristics (BMI, insulin) only genetic clusters but no self-reported ethnicity revealed significant effects. Variability of adion was significantly influenced by age and one self-reported ethnicity. Notably and not unexpected, age had a significant effects on all characteristics except on insulin. For three characteristics (waist, LH, and testosterone) it was the only significant finding. Overall, predominantly genetic ancestry clusters ended up in the final models, indicating that they explain a larger proportion of variability of the PCOS characteristics compared to self-reported ethnicity. To add statistical value to these findings, we calculated per characteristics the probability of finding, at random, an equal or larger number of genetic clusters as observed in the regression analysis using multivariate hypergeometric distribution. For insulin, the probability of randomly selecting at least four genetic clusters was significant (*P* value = 0.02). The hypergeometric probability for the other characteristics revealed no significant results (*P* value > 0.05). Although, generally spoken, genetic ancestry clusters predominantly ended up in the final models, combining all the probabilities resulted in an overall non-significant *P* value of 0.09.

Figure 1 | The first two dimensions of the classical MDS analysis controlled by the genetic diversity present in the HGDP-CEPH reference population of the 1499 PCOS patients based on 152,375 genome-wide SNPs.



Every dot represents a single PCOS patient and the color coding is according to self-reported ethnicity. A classical multidimensional scaling (MDS) with the IBS distance matrix of the HGDP-CEPH panel and each patient independently, retaining the first 10 dimensions, was performed. The first two dimension, i.e. dimension 1 and dimension 2, are plotted on the X axis and Y axis, respectively.

Table 1 | Comparison of self-reported ethnic origin and assigned clusters based on MCLUST clustering algorithm.

Self-reported ethnicity	Sample size (n)	Genetic clusters and percentage (%) of individuals assigned to the clusters based on MCLUST clustering algorithm								Entropy
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	
Sub-Saharan African	55	1.8	0	3.6	21.8	0	0	0	72.7	0.33
Caribbean	48	2.1	0	0	12.5	0	0	0	85.4	0.21
Surinam Creole	44	0	0	0	15.9	0	40.9	4.5	38.6	0.51
Central-East Asian	30	3.3	3.3	0	3.3	0	13.3	76.7	0	0.35
Surinam Hindustani	83	0	0	1.3	7.2	0	91.6	0	0	0.14
European	1000	90	0.7	6.2	2.6	0	0	0.1	0.4	0.18
Turkey	132	0	93.9	5.3	0.8	0	0	0	0	0.11
Morocco	55	1.8	0	0	3.6	94.5	0	0	0	0.11
Middle East	18	5.6	44.4	5.6	33.3	5.6	5.6	0	0	0.60
South American	14	0	0	0	100	0	0	0	0	0
Admixed	20	5.0	0	10.0	60.0	0	0	0	25.0	0.45

Percentages indicate the proportion of individuals from each self-reported ethnic group present in the genetic clusters. Each row sums up to 100%. Entropy is a measure of genetic diversity. The value of entropy ranges from 0, which indicates no diversity or genetically similar, to the log(K), which is the maximum value of diversity. *n* number

Figure 2 | Individual genetic ancestry estimated by ADMIXTURE using K = 10 ancestral components of the HGDP-CEPH reference populations (**panel A**) and the PCOS patients (**panel B** and **panel C**). Each vertical bar represents an individual and the proportion of each individual to the 10 ancestral components is shown in colors.

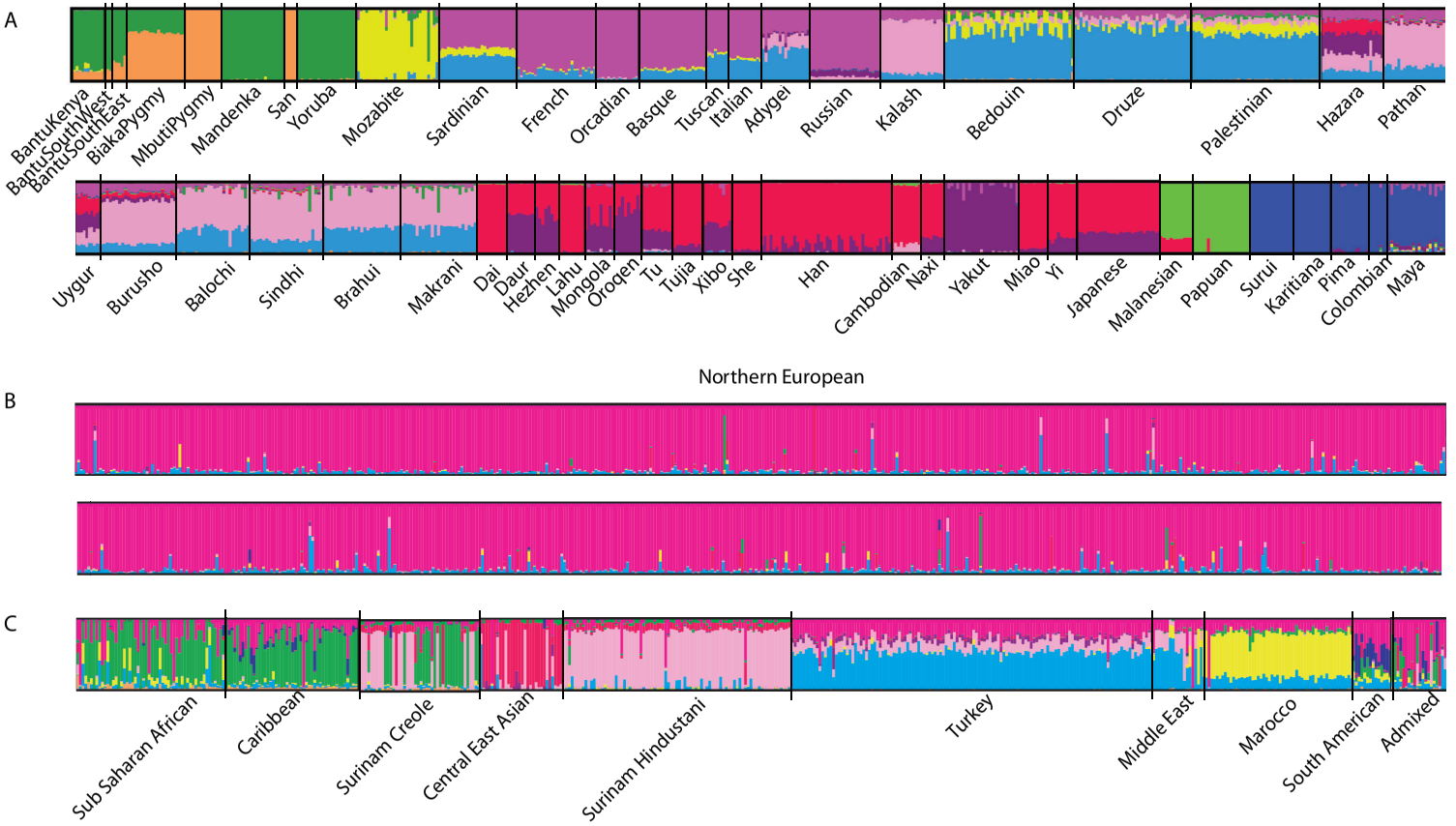


Table 2 | Comparison of self-reported ethnic origin and assigned clusters based on ADMIXTURE clustering algorithm with k = 10.

Self-reported ethnicity	Sample size (n)	Genetic clusters and percentage (%) of individuals assigned to the clusters based on ADMIXTURE clustering algorithm								Entropy
		2	3	4	5	6	7	9	10	
Sub-Saharan African	55	7.2	1.8	0	72.7	0	0	0	18.2	0.35
Caribbean	48	0	0	0	85.4	0	0	0	14.6	0.18
Surinam Creole	44	0	0	40.9	50.0	0	6.8	0	2.2	0.43
Central-East Asian	30	3.3	0	16.7	0	0	76.7	0	3.3	0.32
Surinam Hindustani	83	0	0	94.0	0	0	0	0	6.0	0.10
European	1000	0.07	0	0.1	0.03	0	0.2	0	98.7	0.02
Turkey	132	97.0	0	0	0	0.8	0	0	2.3	0.07
Morocco	55	0	98.2	0	0	0	0	0	1.8	0.04
Middle East	18	72.2	5.5	5.5	0	0	0	0	16.7	0.37
South American	14	0	0	0	7.1	0	0	28.6	64.3	0.36
Admixed	20	5.0	0	5.0	20.0	0	5.0	5.0	60.0	0.53

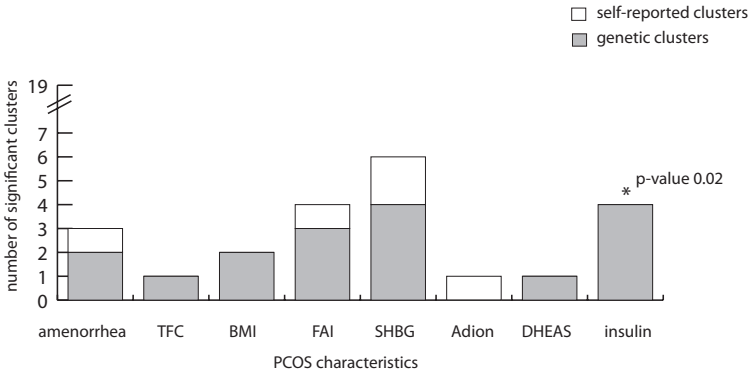
Individuals are assigned to the cluster with the highest proportion of genetic ancestry. None of the individuals were assigned to genetic cluster 1 and genetic cluster 8. Percentages indicate the proportion of individuals from each self-reported ethnic group present in the genetic clusters. Each row sums up to 100%. Entropy is a measure of genetic diversity. The value of entropy ranges from 0, which indicates no diversity or genetically similar, to the log(K), which is the maximum value of diversity. *n*, number

Table 3 | Step-wise forward regression model to determine which self-reported clusters or genetic clusters are responsible for variability in phenotypic characteristics.

	Predictors	**beta**	**s.e.**	***P* value**
Amenorrhea	Age	-0.087	0.013	< 0.0001
	Genetic cluster 4	0.960	0.295	0.001
	Genetic cluster 2	0.745	0.222	0.001
	Morocco	1.384	0.416	0.001
PCOM	-	-	-	-
TFC	Age	-0.008	0.002	< 0.0001
	Genetic cluster 8	0.10	0.115	< 0.0001
BMI	Genetic cluster 8	-0.004	0.001	< 0.0001
	Genetic cluster 5	-0.003	0.001	< 0.0001
Waist	Age	0.001	< 0.000	0.003
LH	Age	-0.005	0.002	0.004
Testosterone	Age	-0.007	0.002	< 0.0001
FAI	Age	-0.01	0.002	< 0.0001
	Genetic cluster 2	0.22	0.03	< 0.0001
	Genetic cluster 6	0.22	0.04	< 0.0001
	Genetic cluster 8	0.19	0.04	< 0.0001
	Central East Asian	0.20	0.07	0.002
SHBG	Age	0.006	0.001	< 0.0001
	Genetic cluster 2	-0.17	0.02	< 0.0001
	Hindustani	-0.18	0.03	< 0.0001
	Genetic cluster 5	-0.19	0.04	< 0.0001
	Genetic cluster 8	-0.13	0.03	< 0.0001
	Central East Asian	-0.20	0.05	< 0.0001
	Genetic cluster 4	-0.08	0.03	0.003
Adion	Age	-0.006	0.001	< 0.0001
	Turkey	0.08	0.018	< 0.0001
DHEAS	Age	-0.03	0.003	< 0.0001
	Genetic cluster 8	-0.17	0.06	0.004
Insulin	Genetic cluster 6	0.33	0.04	< 0.0001
	Genetic cluster 8	0.17	0.03	< 0.0001
	Genetic cluster 5	0.15	0.05	0.002
	Genetic cluster 2	0.10	0.03	0.002

PCOM, polycystic Ovarian Morphology; *TFC,* Total Follicle Count; *BMI,* Body mass Index; *LH,* Luteinizing Hormone; *FAI,* Free Androgen Index; *SHBG,* Sex-Hormone Binding; *Adion,* Androstenedione; *DHEAS,* Dehydroepiandrosterone-sulphate.

Figure 3 | Distribution of genetic versus self-reported significant clusters resulted from the regression analysis per PCOS characteristics.



The number of significant clusters resulted from the regression analysis with a maximum of 19 (11 self-reported ethnic clusters and 8 genetic ancestry clusters) is plotted on the Y-axis. The PCOS characteristics are listed on the X-axis. *TFC*, total follicle count; *BMI*, Body Mass Index; *FAI*, Free Androgen Index; *SHBG*, Sex Hormone Binding Globulin; *Adion*, Androstenedione; *DHEAS*, Dehydroepiandrosterone. * Hypergeometric probability.

Discussion

Usually, ethnicity is assessed by self-reported data and used as a proxy for genetic ancestry. Probably, because self-reported ethnicity is easily accessible via questionnaires or interviews, or medical records. However, disparity regarding the degree to which genetic variation correlates with self-reported ethnicity has been reported [14,23-25]. In general, self-reported ethnic background emphasizes the geographic region of origin of an individual [24]. Moreover, ethnicity also takes into consideration cultural tradition, language, diet, common history and religion among other non-biological factors [26,27], while genetic variation and genetic ancestry solely refer to the biological parts of a person's origin and can be used to describe the bio-geographic ancestry. Moreover, it is not uncommon for individuals to change ethnicities over their own life-course or during that of their direct or distant biological ancestors [28], which can cause discrepancy between a person's bio-geographic origin and his/her ethnicity. Therefore, the use of self-reported ethnicity in biomedical research in principle can lead to misclassifications, biases, and thereby hindering reproducibility between studies [15,29,30]. In addition, errors incurred by using self-reported ethnicity rather than genetic ancestry might produce significant false-positive associations in large studies because of slight differences in genetic ancestry between cases and controls [23,31].

In this study we therefore estimated the ancestral composition based on genome-wide genetic data of a multi-ethnic population of patients diagnosed with PCOS, related this genetic ancestral composition to phenotypic characteristics of PCOS, and compared that to

the impact of self-reported ethnicity. As may be expected, we observed that self-reported ethnicity and genetic ancestry are different in some patients with PCOS. This discrepancy is mostly pronounced in the admixed population, the self-reported Surinam Creoles and the patients from self-reported Middle East, as indicated by differences in entropy scores. On the contrary, the self-reported ethnicity does seem a relative good proxy for genetic ancestry in self-reported European patients, self-reported Surinam Hindustani, self-reported Turkish patients and self-reported Moroccan patients. Over 90% of the patients in each of these self-reported ethnicities end up together in a genetic ancestry cluster. Hence, for some self-reported ethnicities the disparity between self-reported ethnicity and genetic ancestry is larger than for others.

Hitherto, phenotypic differences among patients with PCOS have been only described based on self-reported ethnic assignations, while in this study we test the effect of genetic ancestry on PCOS characteristics and compare these data with the effect of ethnicity. As stated, self-reported ethnicity reflects the sociocultural practices shared among population subgroups, which may or may not include a common genetic origin [32]. If we would have merely found self-reported clusters in the final regression models, one would state that ethnic-specific environmental (sociocultural) factors, determine the variability in PCOS characteristics. In general, we observed that genetic ancestry explains a larger proportion of variability in the phenotypic characteristics of PCOS compared with self-reported ethnicity insinuating that genetic ancestry is a better predictor for phenotypic variability in PCOS. This was in particular the case for insulin levels, which variability was solely determined by genetic ancestry cluster. This highlights the importance of taking genetic factors into account when assessing variability in the PCOS characteristics. Whether this observation supports genetics of PCOS is difficult to disentangle, since it is expected that environmental factors which are relevant for the PCOS phenotype also covariate with genetic ancestry. Diet as well as the level of physical activity differ across ethnic populations and influence the characteristics of PCOS [33]. Moreover, it has been suggested that differences in dyslipidemia in PCOS patients of diverse ethnic and geographical backgrounds are not only explained by variations in body weight, but also genetic and environmental factors, such as diet and activity level, likely contribute to these differences [34]. However, studies combining environmental factors and genetic ancestry instead of self-reported ethnicity are lacking [3]. Therefore, challenges for future studies are incorporation of these environmental and sociocultural factors as well as the genetic factors to explain ethnic variances.

Obviously, also in the light of upcoming GWASs, the results of our study are of interest. The potential confounding effect of genetic population substructure on the interpretation of disease association studies has been well established [23,35]. Our results indicate that this is also an issue in PCOS, stressing the importance of using genetic ancestry to correct for population admixture in genetic studies in patients with PCOS. Although we were able to include a very large study population, the majority of these patients were of European descent providing power-limitations on non-Europeans. Increasing the number of non-European

individuals would increase power and might even identify larger and additional phenotypic differences.

In conclusion, this study indicates that genetic ancestry inferred from genome-wide data is a better predictor for variability of the PCOS phenotype compared to self-reported ethnicity. By using genetic ancestry instead of self-reported ethnicity misclassification based on family records or in admixed individuals can be avoided. Therefore, genetic ancestry inferred from genome-wide data should be recommended for the exact determination of bio-geographic origin. Moreover, as genetic ancestry explained a larger proportion of phenotypic variability associated with PCOS than self-reported ethnicity, future studies should also focus on genetic ancestry verification of PCOS patients for research questions, treatment as well as preventive strategies in these women.

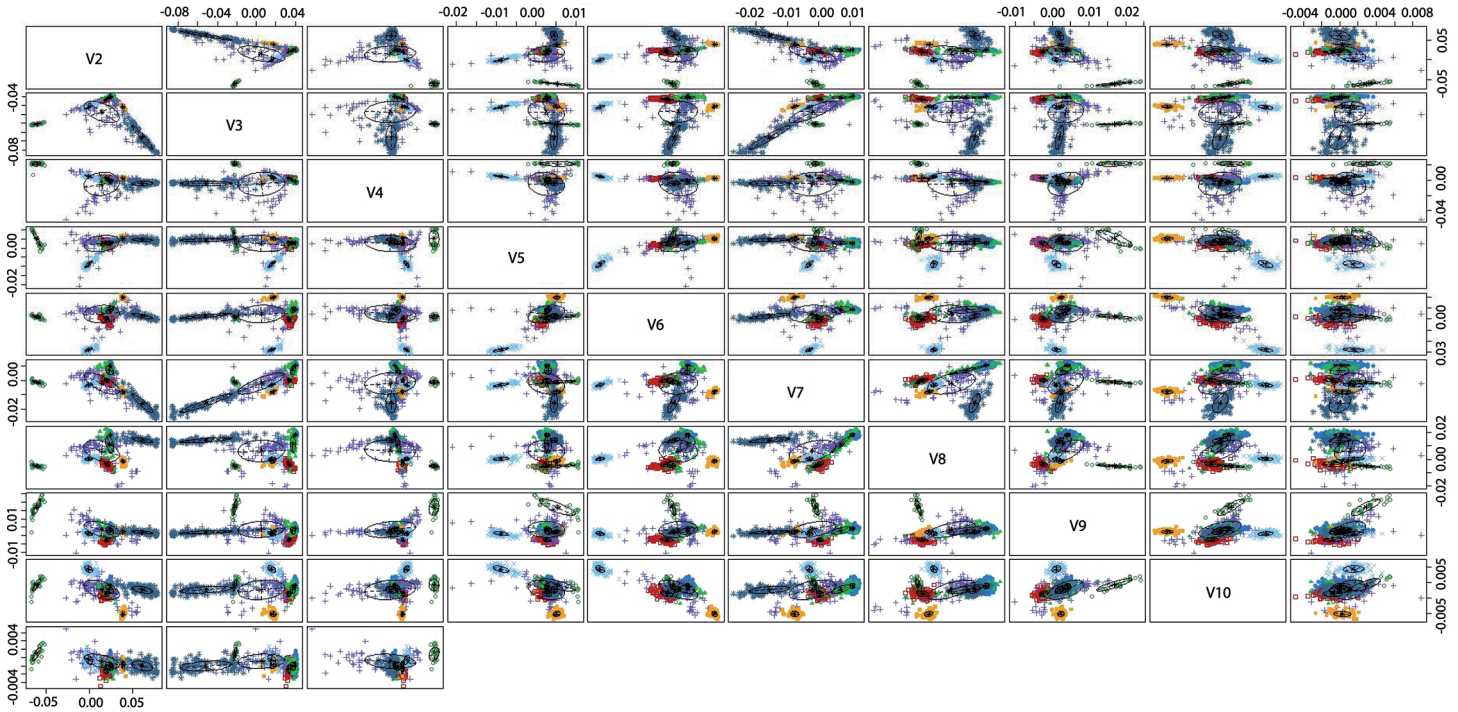
References

1. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
2. ESHRE ASRM Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod.* 2004; Jan; 19: 41-7.
3. Fauser BC, Tarlatzis BC, Rebar RW, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril*. 2012; Jan; 97: 28-38 e25.
4. Zhao Y, Qiao J. Ethnic differences in the phenotypic expression of polycystic ovary syndrome. *Steroids*. 2013; Aug; 78: 755-60.
5. Wang S, Alvero R. Racial and ethnic differences in physiology and clinical symptoms of polycystic ovary syndrome. *Semin Reprod Med*. 2013; Sep; 31: 365-9.
6. Kauffman RP, Baker VM, Dimarino P, et al. Polycystic ovarian syndrome and insulin resistance in white and Mexican American women: a comparison of two distinct populations. *Am J Obstet Gynecol*. 2002; Nov; 187: 1362-9.
7. Glintborg D, Mumm H, Hougaard D, et al. Ethnic differences in Rotterdam criteria and metabolic risk factors in a multiethnic group of women with PCOS studied in Denmark. *Clin Endocrinol (Oxf)*. 2010; Dec; 73: 732-8.
8. Lo JC, Feigenbaum SL, Yang J, et al. Epidemiology and adverse cardiovascular risk profile of diagnosed polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006; Apr; 91: 1357-63.
9. Guo M, Chen ZJ, Eijkemans MJ, et al. Comparison of the phenotype of Chinese versus Dutch Caucasian women presenting with polycystic ovary syndrome and oligo/amenorrhoea. *Hum Reprod*. 2012; May; 27: 1481-8.
10. Welt CK, Arason G, Gudmundsson JA, et al. Defining constant versus variable phenotypic features of women with polycystic ovary syndrome using different ethnic groups and populations. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4361-8.
11. Legro RS, Myers ER, Barnhart HX, et al. The Pregnancy in Polycystic Ovary Syndrome study: baseline characteristics of the randomized cohort including racial effects. *Fertil Steril*. 2006; Oct; 86: 914-33.
12. Carmina E, Legro RS, Stamets K, et al. Difference in body weight between American and Italian women with polycystic ovary syndrome: influence of the diet. *Hum Reprod*. 2003; Nov; 18: 2289-93.
13. Valkenburg O, Lao O, Schipper I, et al. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *J Clin Endocrinol Metab*. 2011; Jul; 96: E1181-7.
14. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, et al. Examining population stratification via individual ancestry estimates versus self-reported race. *Cancer Epidemiol Biomarkers Prev*. 2005; Jun; 14: 1545-51.
15. Kittles RA, Weiss KM. Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet*. 2003; 4: 33-67.
16. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; Sep; 81: 559-75.
17. Rosenberg NA. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet*. 2006; Nov; 70: 841-7.
18. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; Sep; 19: 1655-64.
19. R Core Team. R: A language and environment for statistical computing. Vienna Austria: R Foundation for Statistical Computing; 2013.
20. Fraley C, Raftery A. MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering. University of Washington (revised 2009)2006.

21. Cover TM, Thomas JA. Elements of Information Theory: John Wiley & Sons, Inc.; 1991.
22. Sokal RR, F.J. R. Biometry: The Principles and Practices of Statistics in Biological Research: W. H. Freeman; 2011.
23. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science*. 2002; Dec 20; 298: 2381-5.
24. Risch N, Burchard E, Ziv E, et al. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol*. 2002; Jul 1; 3: comment2007.
25. Bradby H. Describing ethnicity in health research. *Ethn Health*. 2003; Feb; 8: 5-13.
26. Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*. 2003; Mar 20; 348: 1170-5.
27. Race E, Genetics Working G. The use of racial, ethnic, and ancestral categories in human genetics research. *Am J Hum Genet*. 2005; Oct; 77: 519-32.
28. Hitlin S, Scott Brown J, Elder GH, Jr. Racial self-categorization in adolescence: multiracial development and social pathways. *Child Dev*. 2006; Sep-Oct; 77: 1298-308.
29. Caulfield T, Fullerton SM, Ali-Khan SE, et al. Race and ancestry in biomedical research: exploring the challenges. *Genome Med*. 2009; 1: 8.
30. Collins FS. What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era. *Nat Genet*. 2004; Nov; 36: S13-5.
31. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol*. 2001; Nov; 60: 227-37.
32. Lee YL, Teitelbaum S, Wolff MS, et al. Comparing genetic ancestry and self-reported race/ethnicity in a multiethnic population in New York City. *J Genet*. 2010; Dec; 89: 417-23.
33. Moran LJ, Ko H, Misso M, et al. Dietary composition in the treatment of polycystic ovary syndrome: a systematic review to inform evidence-based guidelines. *Hum Reprod Update*. 2013; Sep-Oct; 19: 432.
34. Essah PA, Nestler JE, Carmina E. Differences in dyslipidemia between American and Italian women with polycystic ovary syndrome. *J Endocrinol Invest*. 2008; Jan; 31: 35-41.
35. Knowler WC, Williams RC, Pettitt DJ, et al. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet*. 1988; Oct; 43: 520-6.

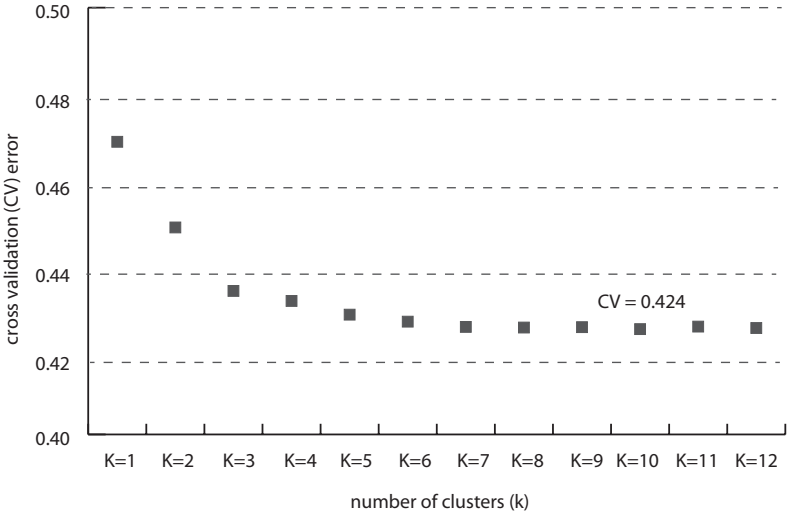
Supplementary files

Supplementary Figure 1 | Pairwise scatter plots of the first 10 dimensions obtained from a classical MDS using an identical by state (IBS) distance matrix between pairs of individuals, and suggested clusters using MCLUST software.



Each dot is a PCOS patient. The coordinates of the first 10 dimensions computed by classical MDS were used to identify clusters of individuals with MCLUST by fitting multivariate normal distributions. The color coding represents the identified clusters based on genetic similarity over all 10 dimensions. The ellipses superimposed on the plot correspond to the covariances of the components.

Supplementary Figure 2 | Cross-validation error (CV) estimates of the cluster analysis using ADMIXTURE software.



The cross-validation error at K = 10 was with 0.424 the smallest, indicating that this is the most sensible model.

Supplementary Table 1 | Self-reported ethnicity and country of birth of all included PCOS patients.

Ethnic origin	Country of Birth
Sub Saharan African	Angola (n=2), Cabo Verde (n=9), Congo (n=2), Eritrea (n=1), Ghana (n=2), Guinea (n=1), Mali (n=1), Mozambique (n=1), Nigeria (n=2), Sao Tome (n=1), Sierra Leone (n=3), Sudan (n=2), Somalia (n=5), South Africa (n=2), the Netherlands (n=21)
Caribbean	Antilles (n=3), Aruba (n=8), Curacao (n= 34), St Maarten (n=1), Dominicans' Rep (n=1), St Kitts (n=1)
Surinam Creole	Surinam (n=44)
Central East Asia	Afghanistan (n=2), China (n=3), Filipinas (n=5), Hong Kong (n=2), Indonesia (n=7), Malaysia (n=1), Singapore (n=1), Sri Lanka (n=2), Vietnam (n=2), South Korea (n=1), the Netherlands (n=4)
Surinam Hindustani	Afghanistan (n=1), India (n=3), Mauritius (n=1), Pakistan (n=11), Suriname (n=45), Sri Lanka (n=1), the Netherlands (n=21)
European	Afghanistan (n=1), Armenia (n=2), Belgium (n=2), Bosnia (n=4), Bulgaria (n=1), Brazil* (n=1), Canada(n=3), Germany (n=5), England (n=3), France (n=4), Georgia (n=1), Greece (n=2), Hungary (n=2), Kosovo (n=1), Lithuania (n=1), the Netherlands (n=937), Norway (n=1), Ukraine (n=1), Poland (n=7), Portugal (n=4), Romania (n=3), Russia (n=2), Serbia (n=2), Slovakia (n=1), Spain (n=2), Yugoslavia (n=1), Swiss (n=3), USA (n=2)
Turkey	Turkey (n=90), Germany (n=2), Netherlands (n=40)
Morocco	Morocco (n=51), Nederland (n=4)
Middle East	Yemen (n=1), Azerbaijan (n=1), Iraq (n=7), Iran (n=3), Israel(n=2), Lebanon (n=1), Syria (n=1), Emirates (n=1), Pakistan (n=1)
South America	Argentina (n=1), Brazil (n=3), Chili (n=1), Colombia (n=2), Costa Rica (n=1), Ecuador (n=1), Mexico (n=2), Peru (n=1), Venezuela (n=2)
Admixed	America (n=1), Colombia (n=1), the Netherlands (n=14), Surinam (n=3)

Chapter 8

General Discussion

More than 105 million women aged 15-49 years have been diagnosed with polycystic ovary syndrome (PCOS) worldwide [1]. As stated, the estimated prevalence of PCOS varies between varies between 5-15%, thereby being the most common endocrine disorder in women of reproductive age [2,3]. Moreover, the economic burden of the disorder is considerable; the healthcare costs are estimated to be over $ 4 billion in the United States annually [1]. The introduction of the additional phenotypes as defined by the Rotterdam criteria [4,5] has increased the phenotypic heterogeneity as well as the prevalence of the syndrome remarkably [6]. The broad diversity of the phenotype emphasizes the need for individual health-risk estimation for a patient with PCOS and her family members in order to provide appropriate clinical care. As stated, the phenotypic characteristics determining PCOS result from a complex interaction between genetic and environmental factors. The focus of this thesis was to identify some of these genetic variants and assess how they might determine PCOS susceptibility. Moreover, we aimed to determine which patients, based on specific phenotypical PCOS characteristics, had the highest risk of developing long-term health complications. The current chapter places these results in a broader context, discusses the potential implications of phenotypic and genetic interaction for clinical practice and suggests future directions for research.

Phenotypic and genetic heterogeneity of PCOS

PCOS has been widely acknowledged as having adverse reproductive as well as metabolic implications resulting in a broad spectrum of signs and symptoms including ovulatory dysfunction, hyperandrogenism, polycystic ovarian morphology (PCOM), obesity, insulin resistance, dyslipidemia, type 2 diabetes mellitus and possibly cardiovascular disease. These individual phenotypic symptoms as well as combinations of these characteristics determine the disease, seem to predict variability in treatment outcome and development of long-term health sequelae. Genetic factors are partly responsible for the diversity of phenotypic characteristics accompanying PCOS and might also have a role in predicting treatment outcome and development of long-term health sequelae.

Ovulatory dysfunction

Menstrual cycle irregularities resulting in oligo- or anovulatory subfertility are one of the keyfeatures of PCOS [4,5]. In case PCOS patients would like to conceive, they have to undergo ovulation induction treatment. Most of the existing ovulation induction protocols are not very much patient-tailored. Nevertheless, oligo-ovulatory patients with PCOS have a more favourable response to ovulation induction treatment using clomiphene citrate compared to anovulatory PCOS patients [7]. In addition, elevated androgen levels and BMI in PCOS patients are associated with clomiphene resistance after ovulation induction treatment [8].

Genetic variants mapping to the *Follicle-Stimulating Hormone Receptor (FSHR)* gene and the *Luteinizing Hormone/Choriogonadotropin Receptor* (*LHCR)* gene have been observed to be associated with PCOS in candidate gene studies in patients from Northern European descent [9,10] and the GWASs in Han-Chinese patients [11,12]. This suggest that these variants play a role in the etiology of PCOS regardless of ethnicity. Moreover, PCOS patients carrying the Ser680 allele in the *FSHR* gene were more often clomiphene citrate resistant than non-carriers [13,14]. Hence, apart from BMI and serum androgen concentrations, the genetically determined sensitivity of the FSH receptor seems to be an important factor too. This implies that genetic data might have a role in further optimizing the existing patient-tailored strategies in ovulation induction treatment in PCOS.

Hyperandrogenism

Hyperandrogenism is considered to play a central role in the etiology of PCOS [15]. One of the factors influencing hyperandrogenism is ethnic origin (for review see [16]). Although patients from the Middle East and those of Mediterranean origin had lower testosterone levels compared to PCOS patients from Northern European descent, they were more often hirsute [17]. In contrast, East Asian patients were less hirsute compared to Europeans, despite having similar androgen concentrations [18]. Such observations suggest interacting genetic factors to influence PCOS phenotype heterogeneity.

Although the ovaries are the principal source of androgen excess in patients with PCOS, the adrenal gland also contributes to the observed hyperandrogenism [19]. Apart from ethnic variation, elevated testosterone as well as DHEAS levels are present in patients with PCOS as well as their brothers and sisters, suggesting that hyperandrogenemia is a genetic trait [20-22]. In general, candidate gene studies as well as GWASs hitherto have not resulted in identification of PCOS susceptibility genes that are directly involved in synthesis and action of androgens. However, genetic variants in these genes have been observed to modify the sub-phenotypic characteristics of PCOS. For example, genetic variants mapping to Cytochrome P450-C17 (*CYP17*) and 11-beta hydroxysteroid dehydrogenase (*HSD11B1*) were not associated with PCOS, nor with the quantitative traits characteristic of PCOS [23]. While genetic variants mapping to the *CYP19* gene [24-26], *SULT2A1* gene [27,28], *SHBG* gene (for a review see [29,30]) and the [CAG] Variable Number of Tandem Repeats (VNTR) in the *Androgen Receptor* (*AR*) gene [31] do not constitute discernible risk alleles for PCOS, they do affect androgen levels in patients with PCOS. Moreover, also genetic variants mapping to genes not directly linked to androgen synthesis were found to modulate androgen levels in PCOS patients. *Thyroid Associated Protein* (*THADA)* is one of the PCOS susceptibility loci identified in the first Chinese GWAS [11] and encodes thyroid adenoma-associated protein, which is expressed in pancreas, adrenal medulla, thyroid, adrenal cortex, testis, thymus, small intestine, and stomach. Chromosomal aberrations of the genomic region containing *THADA* have been observed in benign thyroid adenomas, but are also associated with androgen levels in Han-Chinese and Northern-European patients with PCOS [32,33].

The treatment of hyperandrogenism is challenging. Main treatments are aiming to decrease the free androgen levels by inhibition of ovarian steroid production and simultaneously increase SHBG levels by using the oral contraceptive pill (OCP). The use of OCP has been proven to influence the PCOS phenotype [34]. Unfortunately, not all patients have an improvement of their hirsutism to a similar extent. Since genetic variations mapping to genes involved in androgen synthesis and action have been observed to modify the hyperandrogenic phenotype, it is plausible that these variants might also influence response treatment or even provide potential new drug targets. Studies addressing the effects of genetic variants mapping to genes involved in the synthesis and action of sex steroids in terms of response to OCP, observed no association [35,36]. However, these studies are small and probably underpowered to detect the subtle effect of common variants. As far as ovulation induction is concerned, hyperandrogenism in terms of free androgen levels seems to be the most prominent predictor of the outcome of ovulation induction treatment [37]. Since genetic variants mapping to genes which are involved in the regulation of the steroidogenesis or encode the androgen receptor seem important in determining the androgen levels, these genetic variants might add to the improvement of ovulation induction outcome strategies. Literature addressing this issue is lacking and further investigation of the clinical applicability of the modifiers of the hyperandrogenic phenotype is needed.

Finally, although literature is inconsistent in terms of efficiency of treatment with corticosteroids in PCOS [38], in case of elevated adrenal androgens in PCOS treatment focussing on reducing the adrenal component might have benefits [39].

Polycystic ovarian morphology

The most discussed feature of PCOS is the polycystic ovarian morphology (PCOM). PCOM seems not to be associated with metabolic health implications [40]. However, in contrast to hyperandrogenemia, PCOM seems fairly constant across ethnicities [41,42]. Genetic studies focussing on PCOM are lacking. Hence, PCOM constitutes a kind of epiphenomenon rather than a causal factor. Indeed the fact that ageing in patients having PCOS is similarly associated with a loss of follicles, and subsequently the disappearance of PCOM, suggests that similar genetic factors govern these processes in PCOS patients compared to normal healthy ageing women [43].

Obesity

The relation between elevated BMI-levels and PCOS is obvious considering the fact that the majority of patients with PCOS suffer from obesity [44]. However, whether obese patients are predisposed to PCOS or whether they are obese because of their PCOS status, is subject to a continuous debate [45]. The co-occurrence of PCOS and obesity suggests the possibility of shared genetic susceptibility. Our study showed that, when matching for BMI, genetic variants mapping to well-known BMI-increasing loci are not associated with PCOS suggesting that that these BMI-associated risk alleles do not convey pleiotropic effects on PCOS risk. The

most extensively studied BMI-associated SNPs in relation to PCOS map to the *Fat mass and Obesity-associated* (*FTO*) gene. The majority of these studies suggest that BMI-associated alleles mapping to the *FTO* gene do not play a role in PCOS susceptibility, but that they do influence the metabolic profile of PCOS patients of Northern European descent [46-49]. Current evidence points in the direction that obesity is a modifying rather than a causal factor for PCOS. Indeed, it has been shown that the incidence of PCOS amongst different BMI groups was quite similar [50]. Hence, it seems that obesity aggravates the reproductive and metabolic phenotype of PCOS [51,52]. As a consequence it should not become a necessary prerequisite for the diagnosis of PCOS. Moreover, appropriate first line intervention for patients with PCOS during their reproductive age is lifestyle modification [53]. Dietary changes, increased exercise and weight loss have been shown to decrease adipose tissue and improve insulin sensitivity as well as reproductive function in PCOS indicating that overweight or obesity indeed seems to constitute modifiers of the phenotype instead of being causally involved in PCOS [54-56].

Insulin resistance and type 2 diabetes mellitus

PCOS is associated with insulin resistance leading to hyperinsulinemia, independent of BMI-levels [57,58]. Patients with PCOS seem to have a higher risk of developing gestational diabetes [59,60]. Hindustani patients with PCOS have more often insulin resistance than patients from other ethnic backgrounds [16,42]. Hyperandrogenism is a predictor for insulin resistance and type 2 diabetes mellitus as well [61,62]. Patients with PCOS and also their family members are at an increased risk of developing type 2 diabetes mellitus [57,59,60,63-65].

The major defect in insulin action in PCOS seems to be a post-binding defect in the early steps of insulin signal transduction and appears to be secondary to increased inhibitory serine phosphorylation of the insulin receptor (INSR) and insulin receptor substrate 1 (IRS-1) (for review see [66]). Genetic variants mapping to the *INSR* gene were associated with PCOS in Han-Chinese patients using GWAS [12] as well as patients of Northern European descent using candidate gene analysis [67,68]. Moreover, genetic ancestry compared to self-reported ethnicity explains a large part of the variability in insulin levels suggesting that the genetic factors play an important role, as we have shown. Genetic variants mapping to the *HMGA2* and *THADA* loci, both identified in the GWASs in Han-Chinese PCOS patients [11,12], were also observed to be associated with type 2 diabetes in GWAS [69,70]. Genetic variants mapping to *THADA* were also associated with PCOS susceptibility in Northern European patients [71,72].

Considering that type 2 diabetes mellitus seems to strike stronger in mothers of patients with PCOS, it would be of interest to see whether genetic variants mapping to *THADA* and *HMGA2* influence this elevated mortality risk. Moreover, if long-term follow-up studies in patients with PCOS confirm this burden of insulin resistance and type 2 diabetes mellitus, treatment with metformin might be effective in prevention of these long-term adverse health implications. Treatment with metformin has not proven its efficacy in improving outcome of ovulation induction treatment [73], however, it seems effective in decreasing insulin and androgen levels [74].

Albeit not included in the diagnostic criteria for PCOS, insulin resistance seems to play an important role in the pathogenesis of the syndrome from clinical as well as genetic point of view.

Cardiovascular disease

It has been established that risk factors for cardiovascular events, including insulin resistance, dyslipidemia, increased carotid intima-media thickness and subclinical cardiovascular disease (CVD), are more often present in patients with PCOS compared to controls, independently of BMI [52,75-78]. Hyperandrogenism seems an important predictor for an adverse cardiometabolic profile. Although the increased mortality in mothers might be a proxy for an increased mortality risk in their PCOS daughters, there is need for prospective follow-up studies which associate these CVD predictors in PCOS with hard endpoints, such as stroke and coronary heart disease as well as with CVD mortality rates in women with PCOS. Future research should therefore also focus on the role of genetic factors determining the androgen levels and their contribution to CVD morbidity and mortality in PCOS women. Indeed, recently it has been shown that aromatase is a novel endocrine factor predictive of CVD mortality among postmenopausal women. These authors suggest that extremes of aromatase activity determined through genetic factors might underlie CVD [79]. In this light, larger sufficiently powered studies assessing the role of genetic variation in aromatase activity in women with PCOS are needed.

Remarkably, despite the considerable adverse health implications accompanying PCOS, no guideline for screening and treatment of these adverse implications in PCOS exists in the Netherlands. PCOS guidelines from Australia and the United States recommend that patients with PCOS should be routinely screened for type 2 diabetes mellitus and cardiovascular disease at an early age and on a regular basis [80,81]. Recommendations in terms of the exact age at which the first screening should take place or the frequency of screening, has not yet been determined. Early identification of high-risk patients enables appropriate preventive screening and early treatment of adverse complications. Moreover, since the diagnostic process accounts for only a minor part of the total healthcare costs for PCOS, screening and early treatment will most likely reduce these healthcare costs [1,82].

However, until prospective follow-up studies have associated the presence of CVD predictors in PCOS with cardiovascular events, all patients with PCOS should be monitored similarly. Moreover, to provide the appropriate care, a PCOS guideline in the Netherlands is needed.

Perspectives for research in PCOS

Future research at the Erasmus MC University Medical Centre Rotterdam

All patients with PCOS described in this thesis underwent a similar screening at the outpatient clinic of the Erasmus MC University Medical Centre according to a standardized protocol. The inclusion of these patients is an ongoing effort. Of the large majority of these patients, also DNA samples are available. This multi-ethnic, carefully phenotyped cohort of patients diagnosed with PCOS provides a great opportunity to further explore the pathogenesis of PCOS and the mechanisms of its phenotypic heterogeneity.

Currently, we have initiated a worldwide collaboration of research groups focussing on identifying genetic factors in the pathogenesis of PCOS. This PCOS Genetics Consortium consists of research groups from the Netherlands, the United States, the United Kingdom, Austria, Estonia, Germany, Switzerland, Poland and Iceland. Establishing such a consortium has the advantage of increasing the power to detect subtle genetic effects of common and rare variants and, as such, increasing the probability of identifying genetic variants associated with PCOS. It is expected that this will lead to discoveries of new PCOS susceptibility loci. Further genetic research should also focus on genome-wide genetic association of quantitative traits, such as testosterone levels, to provide insight in the driving underlying biological mechanisms of these individual phenotypes.

Moreover, we are embarking on the prospective long-term follow up of patients who have been diagnosed with PCOS at our clinic and who have reached perimenopausal age. During this first follow-up visit, hard endpoints of cardiovascular disease in terms of mortality or cardiovascular events, have probably not yet occurred considering the relatively young age of the patients. Eventually, by monitoring surrogate markers determining cardiovascular health over time, we will be able to associate these markers to cardiovascular events. Because of the multi-ethnic nature of the patient population visiting our outpatient clinic, we are also able to assess ethnic differences in the development of cardiovascular disease.

Advances in genetics

New technologies, such as whole exome sequencing and whole genome sequencing, have advanced research potentials in genetics even further. The development of these next generation sequencing (NGS) methods has resulted in a better gene-coverage and enables genotyping of rare variants [83]. It is expected that these approaches will also find their way into the genetic research in PCOS [30]. Obviously, identifying the SNPs which are associated with a certain disease is only the beginning. DNA is transcribed to mRNA, which is translated to polypeptide chains, which are folded into protein. The term functional genomics is used to describe combining different approaches to understand the relationship between the genome, including genetic variants, and the phenotype. One of these approaches is measuring expression of genes. Alterations in gene-expression levels could influence phenotypic variation and susceptibility to disease. Gene-expression is detected by using microarrays

and the correlation between SNPs and gene-expression can be studied [84]. Subsequently, to study the effect on cellular level, this specific gene could be knocked out in target tissue. Over the past few decades, several animal models, i.e., mice, rats, sheep and rhesus monkeys, have been developed in an attempt to understand the potential contribution of exposure to excess steroids on the development of PCOS [85]. Especially, mouse models have the benefit of knocking in/out genes for addressing functionality of specific genes [85].

In addition, a few studies addressing differences in DNA methylation profiles between patients with PCOS and controls have been published [86-89]. Epigenetic changes refer to the changes in the genome without changes in DNA sequence. These changes are inheritable through mitosis or meiosis and lead to phenotypic variability. Although the published studies so far have included small sample sizes, methods for genome-wide DNA methylation analysis at a single base pair resolution are evolving quickly [90]. Therefore, it will soon be more feasible, also in PCOS, to carry out studies with sufficiently large sample sizes, in different tissue types and in a genome-wide manner [30].

Non-genetic influences on the PCOS phenotype

PCOS has a substantial heritability, indicating that up to 65% of the risk for PCOS is explained by genetic factors [91]. Therefore, the genetic approach on which this thesis focusses, will not completely explain the pathogenesis of PCOS. Non-genetic factors, for example environmental factors such as the use of OCPs, nutrition and exogenously derived compounds as bisphenol A, might influence also the PCOS phenotype and require further investigation. Also, the phenotype of PCOS is significantly affected by whether the patients arise from a referral population or through unselected screening, likely reflecting the degree of patient concern and awareness and access to healthcare [92].

Conclusions and final remarks

It seems as though genetic variations mapping to genes involved in androgen synthesis and action, are modifying the PCOS phenotype instead of constituting PCOS risk alleles, whereas genes involved in ovulatory dysfunction and insulin signalling or related to type 2 diabetes mellitus are associated with PCOS risk. The GWASs in Han-Chinese patients with PCOS have also been successful in identifying PCOS susceptibility genes involved in unexpected and unknown biological pathways, such as calcium signalling and endocytosis. It is expected that, despite large phenotypic differences between patients from various ethnic origins, GWASs across different ethnicities will reveal at least partly similar PCOS susceptibility loci. Both phenotypic characteristics as well as genetic factors seem to have a role in treatment outcome and long-term health sequelae in PCOS. Obviously, the interaction between these phenotypic and genetic factors is complex and integrating these factors in so-called personalized medicine in terms of treatment and risk prediction will be a challenge for future studies.

References

1. Azziz R, Marin C, Hoq L, et al. Health care-related economic burden of the polycystic ovary syndrome during the reproductive life span. *J Clin Endocrinol Metab*. 2005; Aug; 90: 4650-8.
2. Azziz R, Woods KS, Reyna R, et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab*. 2004; Jun; 89: 2745-9.
3. March WA, Moore VM, Willson KJ, et al. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Hum Reprod*. 2010; Feb; 25: 544-51.
4. ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004; Jan; 81: 19-25.
5. ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004; Jan; 19: 41-7.
6. Broekmans FJ, Knauff EA, Valkenburg O, et al. PCOS according to the Rotterdam consensus criteria: Change in prevalence among WHO-II anovulation and association with metabolic factors. *BJOG*. 2006; Oct; 113: 1210-7.
7. Burgers JA, Fong SL, Louwers YV, et al. Oligoovulatory and anovulatory cycles in women with polycystic ovary syndrome (PCOS): what's the difference? *J Clin Endocrinol Metab*. 2010; Dec; 95: E485-9.
8. Imani B, Eijkemans MJ, te Velde ER, et al. Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab*. 1998; Jul; 83: 2361-5.
9. Laven JS, Mulders AG, Suryandari DA, et al. Follicle-stimulating hormone receptor polymorphisms in women with normogonadotropic anovulatory infertility. *Fertil Steril*. 2003; Oct; 80: 986-92.
10. Mutharasan P, Galdones E, Penalver Bernabe B, et al. Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. *J Clin Endocrinol Metab*. 2013; Jan; 98: E185-90.
11. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011; Jan; 43: 55-9.
12. Shi Y, Zhao H, Shi Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012; Sep; 44: 1020-5.
13. Overbeek A, Kuijper EA, Hendriks ML, et al. Clomiphene citrate resistance in relation to follicle-stimulating hormone receptor Ser680Ser-polymorphism in polycystic ovary syndrome. *Hum Reprod*. 2009; Aug; 24: 2007-13.
14. Valkenburg O, Lambalk CB, van Santbrink EJP, et al. FSH receptor polymorphism predicts outcome of ovulation induction in WHO-II anovulatory subfertility. 67th Annual Meeting of the American Society for Reproductive Medicine, ASRM 2011; Orlando, FL, United States2011.
15. Azziz R, Carmina E, Dewailly D, et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab*. 2006; Nov; 91: 4237-45.
16. Zhao Y, Qiao J. Ethnic differences in the phenotypic expression of polycystic ovary syndrome. *Steroids*. 2013; Aug; 78: 755-60.
17. Glintborg D, Mumm H, Hougaard D, et al. Ethnic differences in Rotterdam criteria and metabolic risk factors in a multiethnic group of women with PCOS studied in Denmark. *Clin Endocrinol (Oxf)*. 2010; Dec; 73: 732-8.
18. Carmina E, Koyama T, Chang L, et al. Does ethnicity influence the prevalence of adrenal hyperandrogenism and insulin resistance in polycystic ovary syndrome? *Am J Obstet Gynecol*. 1992; Dec; 167: 1807-12.

19. Kumar A, Woods KS, Bartolucci AA, et al. Prevalence of adrenal androgen excess in patients with the polycystic ovary syndrome (PCOS). *Clin Endocrinol (Oxf)*. 2005; Jun; 62: 644-9.
20. Legro RS, Driscoll D, Strauss JF, 3rd, et al. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *Proc Natl Acad Sci U S A*. 1998; Dec 8; 95: 14956-60.
21. Legro RS, Kunselman AR, Demers L, et al. Elevated dehydroepiandrosterone sulfate levels as the reproductive phenotype in the brothers of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2002; May; 87: 2134-8.
22. Yildiz BO, Goodarzi MO, Guo X, et al. Heritability of dehydroepiandrosterone sulfate in women with polycystic ovary syndrome and their sisters. *Fertil Steril*. 2006; Dec; 86: 1688-93.
23. Chua AK, Azziz R, Goodarzi MO. Association study of CYP17 and HSD11B1 in polycystic ovary syndrome utilizing comprehensive gene coverage. *Mol Hum Reprod*. 2012; Jun; 18: 320-4.
24. Valkenburg O, Uitterlinden AG, Themmen AP, et al. Genetic polymorphisms of the glucocorticoid receptor may affect the phenotype of women with anovulatory polycystic ovary syndrome. *Hum Reprod*. 2011; Oct; 26: 2902-11.
25. Zhang XL, Zhang CW, Xu P, et al. SNP rs2470152 in CYP19 is correlated to aromatase activity in Chinese polycystic ovary syndrome patients. *Mol Med Rep*. 2012; Jan; 5: 245-9.
26. Xita N, Lazaros L, Georgiou I, et al. CYP19 gene: a genetic modifier of polycystic ovary syndrome phenotype. *Fertil Steril*. 2010; Jun; 94: 250-4.
27. Louwers YV, de Jong FH, van Herwaarden NA, et al. Variants in SULT2A1 Affect the DHEA Sulphate to DHEA Ratio in Patients With Polycystic Ovary Syndrome But Not the Hyperandrogenic Phenotype. *J Clin Endocrinol Metab*. 2013; Sep; 98: 3848-55.
28. Goodarzi MO, Antoine HJ, Azziz R. Genes for enzymes regulating dehydroepiandrosterone sulfonation are associated with levels of dehydroepiandrosterone sulfate in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2007; Jul; 92: 2659-64.
29. Fan W, Li S, Chen Q, et al. Association between the (TAAAA)n SHBG polymorphism and PCOS: a systematic review and meta-analysis. *Gynecol Endocrinol*. 2013; Jul; 29: 645-50.
30. Kosova G, Urbanek M. Genetics of the polycystic ovary syndrome. *Mol Cell Endocrinol*. 2013; Jul 5; 373: 29-38.
31. Zhang T, Liang W, Fang M, et al. Association of the CAG repeat polymorphisms in androgen receptor gene with polycystic ovary syndrome: a systemic review and meta-analysis. *Gene*. 2013; Jul 25; 524: 161-7.
32. Cui L, Zhao H, Zhang B, et al. Genotype-phenotype correlations of PCOS susceptibility SNPs identified by GWAS in a large cohort of Han Chinese women. *Hum Reprod*. 2013; Feb; 28: 538-44.
33. Saxena R, Welt CK. Polycystic ovary syndrome is not associated with genetic variants that mark risk of type 2 diabetes. *Acta Diabetol*. 2013; Jun; 50: 451-7.
34. Mulders AG, ten Kate-Booij M, Pal R, et al. Influence of oral contraceptive pills on phenotype expression in women with polycystic ovary syndrome. *Reprod Biomed Online*. 2005; Dec; 11: 690-6.
35. Maier PS, Mattiello SS, Lages L, et al. 17-Hydroxysteroid dehydrogenase type 5 gene polymorphism (-71A/G HSD17B5 SNP) and treatment with oral contraceptive pills in PCOS women without metabolic comorbidities. *Gynecol Endocrinol*. 2012; Aug; 28: 606-10.
36. Maier PS, Spritzer PM. Aromatase gene polymorphism does not influence clinical phenotype and response to oral contraceptive pills in polycystic ovary syndrome women. *Gynecol Obstet Invest*. 2012; 74: 136-42.
37. Imani B, Eijkemans MJ, de Jong FH, et al. Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab*. 2000; Feb; 85: 676-82.
38. Azziz R, Black VY, Knochenhauer ES, et al. Ovulation after glucocorticoid suppression of adrenal androgens in the polycystic ovary syndrome is not predicted by the basal dehydroepiandrosterone sulfate level. *J Clin Endocrinol Metab*. 1999; Mar; 84: 946-50.

39. Trapp CM, Oberfield SE. Recommendations for treatment of nonclassic congenital adrenal hyperplasia (NCCAH): an update. *Steroids*. 2012; Mar 10; 77: 342-6.

40. Legro RS, Chiu P, Kunselman AR, et al. Polycystic ovaries are common in women with hyperandrogenic chronic anovulation but do not predict metabolic or reproductive phenotype. *J Clin Endocrinol Metab*. 2005; May; 90: 2571-9.

41. Guo M, Chen ZJ, Eijkemans MJ, et al. Comparison of the phenotype of Chinese versus Dutch Caucasian women presenting with polycystic ovary syndrome and oligo/amenorrhoea. *Hum Reprod*. 2012; May; 27: 1481-8.

42. Valkenburg O, Lao O, Schipper I, et al. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *J Clin Endocrinol Metab*. 2011; Jul; 96: E1181-7.

43. Brown ZA, Louwers YV, Fong SL, et al. The phenotype of polycystic ovary syndrome ameliorates with aging. *Fertil Steril*. 2011; Nov; 96: 1259-65.

44. Barber TM, McCarthy MI, Wass JA, et al. Obesity and polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2006; Aug; 65: 137-45.

45. Hoeger KM, Oberfield SE. Do women with PCOS have a unique predisposition to obesity? *Fertil Steril*. 2012; Jan; 97: 13-7.

46. Ewens KG, Jones MR, Ankener W, et al. FTO and MC4R gene variants are associated with obesity in polycystic ovary syndrome. *PLoS One*. 2011; 6: e16390.

47. Kowalska I, Adamska A, Malecki MT, et al. Impact of the FTO gene variation on fat oxidation and its potential influence on body weight in women with polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 2012; Jul; 77: 120-5.

48. Tan S, Scherag A, Janssen OE, et al. Large effects on body mass index and insulin resistance of fat mass and obesity associated gene (FTO) variants in patients with polycystic ovary syndrome (PCOS). *BMC Med Genet*. 2010; 11: 12.

49. Wehr E, Schweighofer N, Moller R, et al. Association of FTO gene with hyperandrogenemia and metabolic parameters in women with polycystic ovary syndrome. *Metabolism*. 2010; Apr; 59: 575-80.

50. Yildiz BO, Knochenhauer ES, Azziz R. Impact of obesity on the risk for polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2008; Jan; 93: 162-8.

51. Welt CK, Gudmundsson JA, Arason G, et al. Characterizing discrete subsets of polycystic ovary syndrome as defined by the Rotterdam criteria: the impact of weight on phenotype and metabolic features. *J Clin Endocrinol Metab*. 2006; Dec; 91: 4842-8.

52. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr Rev*. 1997; Dec; 18: 774-800.

53. Bates GW, Legro RS. Longterm management of Polycystic Ovarian Syndrome (PCOS). *Mol Cell Endocrinol*. 2013; Jul 5; 373: 91-7.

54. Hutchison SK, Stepto NK, Harrison CL, et al. Effects of exercise on insulin resistance and body composition in overweight and obese women with and without polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2011; Jan; 96: E48-56.

55. Kiddy DS, Hamilton-Fairley D, Bush A, et al. Improvement in endocrine and ovarian function during dietary treatment of obese women with polycystic ovary syndrome. *Clin Endocrinol (Oxf)*. 1992; Jan; 36: 105-11.

56. Moran LJ, Hutchison SK, Norman RJ, et al. Lifestyle changes in women with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2011: CD007506.

57. Legro RS, Kunselman AR, Dodson WC, et al. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999; Jan; 84: 165-9.

58. Dunaif A, Segal KR, Futterweit W, et al. Profound peripheral insulin resistance, independent of obesity, in polycystic ovary syndrome. *Diabetes*. 1989; Sep; 38: 1165-74.

59. Joham AE, Ranasinha S, Zoungas S, et al. Gestational Diabetes and Type 2 Diabetes in Reproductive-Aged Women with Polycystic Ovary Syndrome. *J Clin Endocrinol Metab*. 2013; Sep 30.
60. Boomsma CM, Fauser BC, Macklon NS. Pregnancy complications in women with polycystic ovary syndrome. *Semin Reprod Med*. 2008; Jan; 26: 72-84.
61. Goverde AJ, van Koert AJ, Eijkemans MJ, et al. Indicators for metabolic disturbances in anovulatory women with polycystic ovary syndrome diagnosed according to the Rotterdam consensus criteria. *Hum Reprod*. 2009; Mar; 24: 710-7.
62. Moran L, Teede H. Metabolic features of the reproductive phenotypes of polycystic ovary syndrome. *Hum Reprod Update*. 2009; Jul-Aug; 15: 477-88.
63. Davies MJ, Marino JL, Willson KJ, et al. Intergenerational associations of chronic disease and polycystic ovary syndrome. *PLoS One*. 2011; 6: e25947.
64. Sir-Petermann T, Angel B, Maliqueo M, et al. Prevalence of Type II diabetes mellitus and insulin resistance in parents of women with polycystic ovary syndrome. *Diabetologia*. 2002; Jul; 45: 959-64.
65. Yildiz BO, Yarali H, Oguz H, et al. Glucose intolerance, insulin resistance, and hyperandrogenemia in first degree relatives of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2003; May; 88: 2031-6.
66. Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr Rev*. 2012; Dec; 33: 981-1030.
67. Goodarzi MO, Louwers YV, Taylor KD, et al. Replication of association of a novel insulin receptor gene polymorphism with polycystic ovary syndrome. *Fertil Steril*. 2011; Apr; 95: 1736-41 e1-11.
68. Siegel S, Futterweit W, Davies TF, et al. A C/T single nucleotide polymorphism at the tyrosine kinase domain of the insulin receptor gene is associated with polycystic ovary syndrome. *Fertil Steril*. 2002; Dec; 78: 1240-3.
69. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*. 2010; Jul; 42: 579-89.
70. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008; May; 40: 638-45.
71. Goodarzi MO, Jones MR, Li X, et al. Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *J Med Genet*. 2012; Feb; 49: 90-5.
72. Louwers YV, Stolk L, Uitterlinden AG, et al. Cross-Ethnic Meta-analysis of Genetic Variants for Polycystic Ovary Syndrome. *J Clin Endocrinol Metab*. 2013; Oct 8.
73. Tang T, Lord JM, Norman RJ, et al. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev*. 2012; 5: CD003053.
74. Nascimento AD, Silva Lara LA, Japur de Sa Rosa-e-Silva AC, et al. Effects of metformin on serum insulin and anti-Mullerian hormone levels and on hyperandrogenism in patients with polycystic ovary syndrome. *Gynecol Endocrinol*. 2013; Mar; 29: 246-9.
75. Carmina E, Orio F, Palomba S, et al. Endothelial dysfunction in PCOS: role of obesity and adipose hormones. *Am J Med*. 2006; Apr; 119: 356 e1-6.
76. Talbott EO, Zborowski JV, Rager JR, et al. Evidence for an association between metabolic cardiovascular syndrome and coronary and aortic calcification among women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2004; Nov; 89: 5454-61.
77. Talbott EO, Zborowski JV, Sutton-Tyrrell K, et al. Cardiovascular risk in women with polycystic ovary syndrome. *Obstet Gynecol Clin North Am*. 2001; Mar; 28: 111-33, vii.
78. Valkenburg O, Steegers-Theunissen RP, Smedts HP, et al. A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. *J Clin Endocrinol Metab*. 2008; Feb; 93: 470-6.
79. Laughlin GA, Ix JH, Cummins K, et al. Extremes of an aromatase index predict increased 25-year risk of cardiovascular mortality in older women. *Clin Endocrinol (Oxf)*. 2012; Sep; 77: 391-8.

80. Evidence-based guideline for the assessment and management of polycystic ovary syndrome. Australia: PCOS Australian Alliance, 2011.

81. Wild RA, Carmina E, Diamanti-Kandarakis E, et al. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab*. 2010; May; 95: 2038-49.

82. Panattoni LE, Vaithianathan R, Ashton T, et al. Predictive risk modelling in health: options for New Zealand and Australia. *Aust Health Rev*. 2011; Feb; 35: 45-51.

83. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010; Jun; 11: 415-25.

84. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet*. 2011; Feb; 27: 72-9.

85. Padmanabhan V, Veiga-Lopez A. Animal models of the polycystic ovary syndrome phenotype. *Steroids*. 2013; Aug; 78: 734-40.

86. Xu N, Azziz R, Goodarzi MO. Epigenetics in polycystic ovary syndrome: a pilot study of global DNA methylation. *Fertil Steril*. 2010; Jul; 94: 781-3 e1.

87. Sang Q, Zhang S, Zou S, et al. Quantitative analysis of follistatin (FST) promoter methylation in peripheral blood of patients with polycystic ovary syndrome. *Reprod Biomed Online*. 2013; Feb; 26: 157-63.

88. Ting W, Yanyan Q, Jian H, et al. The Relationship Between Insulin Resistance and CpG Island Methylation of LMNA Gene in Polycystic Ovary Syndrome. *Cell Biochem Biophys*. 2013; Apr 20.

89. Dasgupta S, Sirisha PV, Neelaveni K, et al. Androgen receptor CAG repeat polymorphism and epigenetic influence among the south Indian women with Polycystic Ovary Syndrome. *PLoS One*. 2010; 5: e12401.

90. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet*. 2010; Mar; 11: 191-203.

91. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006; Jun; 91: 2100-4.

92. Ezeh U, Yildiz BO, Azziz R. Referral bias in defining the phenotype and prevalence of obesity in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2013; Jun; 98: E1088-96.

Chapter 9

Summary
Samenvatting

Summary

The Polycystic Ovary Syndrome (PCOS) is the most common endocrine disorder in women of reproductive age. PCOS has reproductive as well metabolic symptomatology and is considered to be a complex genetic disease. During reproductive age, key features include oligomenorrhea or amenorrhea, hirsutism or hyperandrogenism and polycystic ovarian morphology. Later in life, adverse metabolic implications, such as obesity, insulin resistance, type 2 diabetes and cardiovascular disease, become more prominent. In this thesis, we aimed to identify genetic factors for PCOS susceptibility using a candidate-gene approach and a hypothesis-free genome-wide approach. Moreover, we have identified several high-risk groups for long-term health implications within the broad phenotype of PCOS.

Chapter 1 provides a general introduction describing the diagnostic criteria of PCOS, its phenotypic heterogeneity and complex genetic background.

The variability of the phenotype of PCOS seems to represent variability in the long-term health implications of the syndrome. In **chapter 2** the cardiometabolic profile of 2,278 carefully phenotyped patients with PCOS from two University Hospitals was analyzed. Hyperandrogenic patients had besides their elevated androgen levels, an increased waist circumference and more often obesity and overweight, insulin resistance as well as the metabolic syndrome compared to patients with non-hyperandrogenic PCOS. Overall, no differences were observed between the endocrine profiles of the component hyperandrogenic PCOS phenotypes. Current efforts to monitor patients with PCOS prospectively to obtain these hard endpoint data, such as cardiovascular events and excess mortality later in life, need approximately 25 more years of follow-up. Therefore, in **chapter 3**, we made use of the heritable nature of the syndrome by performing a reverse parent-offspring study and determined all-cause mortality in parents of patients with PCOS using Standardized Mortality Rate analysis. Mothers above age 60 years had a significant excess mortality compared to the general Dutch population. Moreover, diabetic mothers of patients with PCOS had a two times higher mortality risk compared to patients with type 2 diabetes who were not selected through PCOS. No excess mortality among fathers was observed. The severe excess mortality justifies active screening of the mothers of PCOS patients to ensure that good preventive and therapeutic measures are taken timely.

Chapter 4 includes the results of three candidate gene analyses which aimed to identify PCOS susceptibility genes. For this approach, candidate genes were chosen based on prior knowledge of the function of the gene and its potential biological mechanisms underlying the disease. Since PCOS is associated with insulin resistance, genes involved in the insulin signalling pathway seem obvious PCOS candidate genes. In **chapter 4.1**, a thorough evaluation of nearly 300 genetic variants in 39 genes which are involved in the insulin signalling pathway was performed. At the discovery stage, 4 SNPs mapping to the insulin receptor gene (*INSR*) showed significant association with PCOS. One SNP, rs2252673, was associated with PCOS in the replication cohort, also after adjusting for BMI. This implies that the insulin receptor is involved in the etiology of PCOS.

Not only the ovary, but also the adrenal gland contributes to the androgen excess in PCOS, as reflected by elevated DHEAS levels. **Chapter 4.2** concerns the possible mechanisms behind the heritability of DHEAS levels in PCOS patients and studies the genetic variations of the key enzymes regulating the level of DHEA sulfation. Although the candidate-gene study was appropriately powered, genetic variants mapping to the *SULT2A1*, *PAPSS2* and *STS* genes were not associated with PCOS predisposition or with the clinical signs of PCOS. SNP rs2910397 mapping to *SULT2A1* was associated with a decrease in the DHEAS to DHEA ratio. Because PCOS is a complex genetic disorder, interaction of several genetic variants in genes encoding enzymes in the steroidogenesis might contribute to the variation in adrenal androgen levels in PCOS.

The third candidate-gene study studied BMI-associated risk alleles in association with PCOS. The relation between obesity and PCOS clearly exists. Obesity is a highly heritable trait and genome-wide association studies (GWASs) investigating BMI-loci have been very successful. **Chapter 4.3** describes a combined analysis of over 4,000 BMI-matched individuals from the United Kingdom and the Netherlands. We observed no association with PCOS of BMI risk alleles mapping to *BDNF*, *FAIM2*, *ETV5*, *FTO*, *GNPDA2*, *KCTD15*, *MC4R*, *MTCH2*, *NEGR1*, *SEC16B*, *SH2B1*, and *TMEM18*. This suggests that there is no systematic effect of BMI-associated alleles on PCOS risk independent of BMI, suggesting that these alleles do not have a pleiotropic effect on PCOS susceptibility. Moreover, adjusting for BMI in PCOS case-control GWAS studies should be an effective strategy for removing confounding effects of BMI on the association of other genetic variants and PCOS.

Obviously, candidate-gene selection is limited by incomplete knowledge of the pathophysiology of PCOS. By performing a hypothesis-free GWAS, one can overcome this limitation. Hitherto, two GWASs in Han-Chinese patients and one in Korean patients with PCOS have been published. **Chapter 5** reports the preliminary results of a GWAS in patients with PCOS of Northern European ancestry. In GWASs > 1,000,000 common genetic variants (SNPs) are tested for an association with the disease without a prior knowledge about the potential effect of the variants or genes on the disease. Although this Northern European GWAS in patients with PCOS has revealed a number of new potentially interesting PCOS-associated loci, no genome-wide significant SNPs were detected. The strongest PCOS-associated SNP was rs11681377 on chromosome 2. However, this result should be interpreted with care. It is crucial that these variants are replicated in large independent samples of patients with PCOS and controls. The lack of power seems to be underlying the negative results of this GWAS. To overcome this issue, meta-analyses in a large consortium are crucial.

Since it has been well established that ethnic background adds to phenotypic diversities in PCOS, it seems plausible that genetic variants associated with PCOS act differently in various ethnic populations. **Chapter 6** describes a cross-ethnic meta-analysis of susceptibility loci for Han-Chinese PCOS patients in patients of Northern European descent. We studied the effects of genetic variants identified in Han-Chinese GWASs mapping to the *LHCGR*, *THADA*, *DENND1A*, *FSHR, c9orf3*, *YAP1*, *RAB5B/SUOX*, *HMGA2*, *TOX3*, *INSR*, *SUMO1P1* loci in patients of Northern

European ancestry and assessed the cross-ethnic effect of the SNPs. Overall, we observed for 12 out of 17 tested genetic variants mapping to the Chinese PCOS loci similar effect size and identical direction in PCOS patients of Northern European ancestry. Two variants even reached genome-wide significance level. This indicates a common genetic risk profile for PCOS across populations. Hence, it is expected that upcoming large GWASs in PCOS patients from Northern European ancestry will partly identify similar loci as the GWAS in Chinese PCOS patients.

Usually in medical studies, including those on PCOS, ethnicity is assessed by self-reported data, either via interviews or obtained from medical records. **Chapter 7** concludes that self-reported ethnicity is not a perfect proxy for genetic ancestry in patients with PCOS. To avoid false-positive associations because of misclassification, it is therefore advised to use genetic ancestry to correct for population admixture in genetic studies in patients with PCOS. Moreover, overall, a larger proportion of variation in phenotypic characteristics of PCOS is explained by genetic ancestry. This seems especially the case with insulin resistance.

Finally, **Chapter 8** places the results described in this thesis in wider context and discusses to which extent genetic studies in PCOS have contributed to unraveling its underlying mechanisms. Moreover it discusses potential implications for clinical practice and suggests future directions for research.

Samenvatting

Het Polycysteus Ovarium Syndroom (PCOS) is de meest voorkomende endocriene (hormonale) aandoening bij vrouwen in de reproductieve (vruchtbare) levensfase. PCOS is een complex genetische aandoening. Bij complex genetische aandoeningen zijn genetische variaties betrokken bij het ontstaan van het ziektebeeld. Tijdens de reproductieve levensfase wordt PCOS gekenmerkt door anovulatie (het uitblijven van de eisprong), hyperandrogenisme (verhoogde waarden van mannelijk hormoon in het bloed) of hirsutisme (overbeharing) en polycysteuze ovaria (de aanwezigheid van veel kleine eiblaasjes in de eierstokken). Dit leidt bij het merendeel van de patiënten tot vruchtbaarheidsproblemen. Later in hun leven hebben zij een grotere kans op het ontwikkelen van insuline resistentie (ongevoeligheid voor insuline), type 2 diabetes mellitus (ouderdomssuikerziekte), en mogelijk ook hart- en vaatziekten.
Het doel van dit proefschrift was het identificeren van genetische variaties die een rol spelen bij het ontstaan van PCOS. Ook worden voorspellers op het ontwikkelen van de lange termijn complicaties van PCOS beschreven.

Hoofdstuk 1 beschrijft de criteria op basis waarvan de diagnose PCOS gesteld wordt. Ook zijn mogelijke onderliggende biologische mechanismen van het ziektebeeld kort besproken. Verder is uitgelegd wat genetische variaties zijn en hoe deze het ontstaan van ziekten kunnen beïnvloeden. De meest voorkomende genetische variatie is het zogenaamde Single Nucleotide Polymorfisme (SNP), een enkelvoudige base-verandering in het DNA. Wanneer bepaalde SNPs vaker voorkomen bij patiënten dan bij individuen uit de algemene populatie, lijken deze geassocieerd met het ziektebeeld.

De aanwezigheid van specifieke symptomen van PCOS in de reproductieve levensfase lijken geassocieerd te zijn met een grotere kans op het ontstaan van complicaties op latere leeftijd. In **hoofdstuk 2** is geanalyseerd of de aanwezigheid van hyperandrogenisme gerelateerd is aan een ongunstig metabool profiel. In totaal zijn 2278 patiënten uit twee academische ziekenhuizen geïncludeerd in deze studie. PCOS patiënten met hyperandrogenisme hadden inderdaad een meer ongunstig metabole uitkomst. Naast de hogere androgeen waarden hadden zij een grotere middelomtrek, een hogere Body Mass Index (BMI) en hogere insuline waarden. Bovendien vonden we – in vergelijking met PCOS patienten zonder hyperandrogenisme – vaker overgewicht, insulineresistentie en het metabool syndroom (i.e. een combinatie van insulineresistentie, overgewicht, hoge bloeddruk, en verhoogd cholesterol). PCOS patiënten met hyperandrogenisme lijken dus een hoger risico te hebben op het ontstaan van lange termijn complicaties dan PCOS patiënten met normale androgeen waarden. Momenteel worden de patiënten prospectief vervolgd om te kunnen vaststellen of deze complicaties, zoals type 2 diabetes mellitus, hart- en vaatziekten en misschien zelfs mortaliteit, ook daadwerkelijk vaker optreden. Echter, om hier met zekerheid uitspraken over te kunnen doen, is nog jaren van vervolgonderzoek nodig. Om toch nu al meer inzicht te verkrijgen in het optreden van deze lange termijn complicaties, is in **hoofdstuk 3** gebruik gemaakt van de erfelijke aard van het ziektebeeld. Middels een zogenoemde 'reversed parent-offspring'

analyse hebben we de sterfte van ouders van PCOS patiënten geïnventariseerd. Moeders die ouder dan 60 jaar waren, lieten een significant grotere sterfte zien in vergelijking met de sterfte in de algemene Nederlandse bevolking. Bovendien hadden de moeders die leden aan type 2 diabetes mellitus een twee keer zo hoog sterfterisico dan vrouwen met type 2 diabetes uit de algemene – dus niet via PCOS geselecteerde - populatie. Deze ernstige oversterfte rechtvaardigt screening van de moeders van PCOS patiënten zodat de juiste preventieve en therapeutische maatregelen tijdig genomen kunnen worden.

Hoofdstuk 4 beschrijft een drietal kandidaat-gen studies. Bij kandidaat-gen onderzoek worden genen geselecteerd die mogelijk een rol kunnen spelen bij het ontstaan van PCOS. Deze selectie wordt gebaseerd op reeds aanwezige kennis over het ziektebeeld. Vervolgens wordt getest of genetische variaties (SNPs) in deze genen vaker voorkomen bij PCOS patiënten dan bij controle-vrouwen uit de algemene populatie. Wanneer dit het geval is, lijken deze variaties dus een rol te spelen bij het ontstaan van PCOS.

Aangezien patiënten met PCOS vaker ongevoelig zijn voor insuline, zijn genen betrokken in de insuline signaaltransductie route voor de hand liggende kandidaat-genen voor PCOS. Bij signaaltransductie wordt een signaal (zoals verhoging van de insuline spiegels in het bloed na het eten van een maaltijd) herkend door de cel en dit heeft specifieke veranderingen in die cel tot gevolg (zoals stimulatie van de glucose opname in bijvoorbeeld de spieren). Wanneer de insuline signaaltransductie niet naar behoren functioneert, zou dat insuline resistentie en hogere insuline waarden in het bloed, zoals bij PCOS het geval is, kunnen verklaren. **Hoofdstuk 4.1** laat zien dat een SNP in de insuline receptor inderdaad vaker voorkomt bij patiënten met PCOS dan bij de controlegroep. Dit was niet het geval voor andere bekende genetische variaties die betrokken zijn bij de insuline signaaltransductie. De aangetoonde associatie is een sterke aanwijzing dat de insuline receptor betrokken is bij de etiologie van PCOS.

Bij PCOS is er niet alleen sprake van een verhoogde androgeen productie vanuit het ovarium, maar er is ook een toegenomen androgeen productie vanuit de bijnier. PCOS patiënten en hun familieleden hebben verhoogde DHEAS waarden in het bloed. DHEAS wordt exclusief door de bijnier geproduceerd en kan worden omgezet in DHEA. DHEA is een voorloper van de actieve mannelijke hormonen in het bloed, zoals testosteron. In **hoofdstuk 4.2** is getest of variaties in genen die belangrijk zijn voor de omzetting van DHEAS naar DHEA en andersom (namelijk *SULT2A1*, *PAPSS2* en *STS*), vaker voorkomen bij PCOS patiënten dan bij de controlegroep. Ook is de associatie van deze genetische variaties met de variabiliteit van de androgeen waarden bij PCOS patiënten bestudeerd. Er werd geen associatie gevonden tussen de genetische variaties in *SULT2A1*, *PAPSS2* en *STS* en de aanwezigheid van PCOS. SNP rs2910397 in *SULT2A1* was geassocieerd met een lagere DHEAS/DHEA ratio, maar beïnvloedde het androgene fenotype van PCOS niet. Aangezien PCOS een complex genetische aandoening is, lijkt het waarschijnlijk dat een interactie tussen verscheidene genetische varianten in genen die belangrijk zijn bij de productie van androgenen kan bijdragen aan de variatie in androgeen waarden in PCOS.

De derde kandidaat-gen studie richt zich op genetische variaties die geassocieerd zijn met BMI. Grote genetische studies in de algemene bevolking hebben veel BMI-geassocieerde genetische variaties geïdentificeerd. Omdat het merendeel van de patiënten met PCOS overgewicht heeft, is het aannemelijk dat deze BMI-geassocieerde genetische variaties vaker voorkomen bij patiënten met PCOS. Eerdere studies hebben aangetoond dat dit het geval lijkt. Echter, na correctie voor het verschil in BMI tussen de PCOS patiënten en controlegroep, lijkt deze associatie te verdwijnen. **Hoofdstuk 4.3** beschrijft een gecombineerde analyse van ruim 4.000 PCOS patiënten en controles uit Nederland en het Verenigd Koninkrijk. De patiënten en controlegroep hadden een vergelijkbaar BMI om te voorkomen dat het verschil in BMI de resultaten zou beïnvloeden. De bestudeerde BMI-geassocieerde genetische variaties kwamen net zo vaak voor in beide groepen. Dit suggereert dat deze genetische variaties niet geassocieerd zijn met PCOS wanneer er rekening wordt gehouden met BMI.

Vanzelfsprekend wordt kandidaat-gen selectie beperkt door de onvolledige kennis van de onderliggende biologische mechanismes van PCOS. Bij een hypothese-vrije genoomwijde associatie studie (GWAS) is dit niet het geval. Tot nu toe zijn er twee GWASs in PCOS patiënten van Chinese afkomst verschenen en één GWAS in Koreaanse patiënten. **Hoofdstuk 5** rapporteert de voorlopige resultaten van de GWAS studie in PCOS patiënten van Europese afkomst. In een GWAS worden meer dan 1.000.000 veelvoorkomende genetische variaties in het humane DNA getest in associatie met het ziektebeeld, zonder dat er op voorhand een hypothese is over mogelijke effecten van deze genetische variaties of genen op het ziektebeeld. Hoewel er geen genoomwijd significant geassocieerde SNPs werden gevonden, waren er wel een aantal SNPs suggestief voor associatie met PCOS. SNP rs11681377 op chromosoom 2 was het sterkst geassocieerd met PCOS. Echter, alvorens harde conclusies te kunnen trekken uit deze voorlopige resultaten is het van wezenlijk belang om deze SNPs in een groter cohort te repliceren. Zeer waarschijnlijk zijn grotere aantallen patiënten nodig om de kleine effecten van SNPs te kunnen signaleren in deze GWAS setting. Daarom is samenwerking binnen een groot PCOS consortium essentieel.

De etnische achtergrond van patiënten is van invloed op de ernst van de PCOS symptomen. PCOS patiënten van mediterrane afkomst hebben vaker hirsutisme dan Europese patiënten. Patiënten van Hindoestaanse afkomst hebben frequenter insuline resistentie dan patiënten van andere etnische achtergronden. Het lijkt daarom plausibel dat genetische variaties die met PCOS geassocieerd zijn, een verschillend effect kunnen hebben op het fenotype van patiënten van verschillende etnische afkomst. In **hoofdstuk 6** worden de effecten van de genetische varianten die geïdentificeerd zijn in de GWAS in Chinese patiënten, bestudeerd in een groep patiënten met PCOS van Europese afkomst. Deze gecombineerde analyse laat zien dat 12 van de 17 geteste genetische varianten een effect van dezelfde grootte en in dezelfde richting hebben als in de Chinese studie. Dit duidt op een gemeenschappelijk genetisch risicoprofiel voor PCOS patiënten van verschillende etnische achtergrond. Het is dan ook de verwachting dat de grote PCOS GWAS in patiënten van Europese afkomst voor een deel dezelfde gebieden zal identificeren als de GWASs in Chinese PCOS patiënten hebben gedaan.

Voor de bepaling van de etnische achtergrond van patiënten worden in medische studies meestal anamnestisch verkregen gegevens of gegevens uit het medisch dossier gebruikt. **Hoofdstuk 7** laat zien dat de zelf-gerapporteerde etnische achtergrond geen perfecte maat is voor de genetische etnische afkomst van de patiënten met PCOS. Misclassificatie op basis van zelf-gerapporteerde etniciteit kan tot vals-positieve associaties leiden. Om deze vals-positieve associaties te voorkomen, zou het gebruik van de genetisch etnische afkomst de voorkeur hebben. Bovendien bepaalt deze genetische afkomst een groter deel van de variatie in de uiterlijke kenmerken van PCOS. Dit lijkt met name bij insulineresistentie het geval.

Tenslotte plaatst **hoofdstuk 8** de beschreven resultaten uit dit proefschrift in breder perspectief en bediscussieert het hoe genetische studies hebben bijgedragen aan het ontrafelen van onderliggende biologische mechanismen van PCOS. Ook worden mogelijke implicaties voor de klinische praktijk besproken en richtingen voor vervolgonderzoek gesuggereerd.

Addendum

Authors and affiliations

R. Azziz	Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, USA
	Department of Medicine and Obstetrics and Gynecology, David Geffen School of Medicine, University of California, Los Angeles, USA
T.M. Barber	Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom
M. van den Berg	Division of Reproductive Medicine, Department of Obstetrics and Gynecology, University Medical Center Groningen, Groningen, The Netherlands
H. J. G. Bilo	Diabetes Center, Isala clinics, Zwolle, The Netherlands
	Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands
D. Boomsma	Department of Biological Psychology, VU University, Amsterdam, The Netherlands
	Centre for Medical Systems Biology, Leiden, The Netherlands
Y.I. Chen	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, USA
J. Cui	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, USA
N.M.P. Daan	Department of Reproductive Medicine and Gynaecology, Division Woman and Baby, University Medical Centre Utrecht, The Netherlands
M.C.J. Eijkemans	Department of Reproductive Medicine and Gynaecology, Division Woman and Baby, University Medical Centre Utrecht, The Netherlands
	Julius Centre for Health Sciences and Primary care, University Medical Centre Utrecht, The Netherlands
B.C.J.M. Fauser	Department of Reproductive Medicine and Gynaecology, Division Woman and Baby, University Medical Centre Utrecht, The Netherlands
S. Franks	Institute of Reproductive and Developmental Biology, Imperial College London, Hammersmith Hospital, London, United Kingdom

M.O. Goodarzi	Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, USA
	Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California
	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California
	Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, California
C.J. Groves	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
	Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom
X. Guo	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California
B.M. Herrera	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
N.A.A. van Herwaarden	Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, ErasmusMC University Medical Centre, Rotterdam, The Netherlands
M.R. Jones	Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California
F.H. de Jong	Department of Internal Medicine, Erasmus MC University Medical Centre, Rotterdam, the Netherlands
M. Kayser	Department of Forensic Molecular Biology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
N. Kleefstra	Diabetes Center, Isala clinics, Zwolle, The Netherlands
	Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands
M.P.H. Koster	Department of Reproductive Medicine and Gynaecology, Division Woman and Baby, University Medical Centre Utrecht, The Netherlands
S. Kwon	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California
C.B. Lambalk	Department of Obstetrics, Gynecology and Reproductive Medicine, VU University Medical Centre, Amsterdam, The Netherlands

O. Lao	Department of Forensic Molecular Biology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
J.S.E. Laven	Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, ErasmusMC University Medical Centre, Rotterdam, The Netherlands
E.W.G. Lentjes	Department of Clinical Chemistry and Haematology, University Medical Centre Utrecht, The Netherlands
M.I. McCarthy	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
	Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom
N.W. Rayner	Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
	Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom
Y.B. de Rijke	Departments of Clinical Chemistry, Endocrine Laboratory, ErasmusMC University Medical Centre, Rotterdam, The Netherlands
M.E. Roest-Schalken	Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, ErasmusMC University Medical Centre, Rotterdam, The Netherlands
J. Roeters van Lennep	Department of Internal Medicine, ErasmusMC University Medical Center, Rotterdam, The Netherlands
E.J.G. Sijbrands	Department of Internal Medicine, ErasmusMC University Medical Center, Rotterdam, The Netherlands
L. Stolk	Department of Internal Medicine, Erasmus MC University Medical Centre, Rotterdam, the Netherlands
K.D. Taylor	Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California
A.G. Uitterlinden	Department of Internal Medicine, Erasmus MC University Medical Centre, Rotterdam, the Netherlands
	Department of Epidemiology, Erasmus MC University Medical Centre, Rotterdam, the Netherlands
O. Valkenburg	Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, ErasmusMC University Medical Centre, Rotterdam, The Netherlands

List of abbreviations

11β-HSD	11β-hydroxysteroid dehydrogenase
3β-HSD	3β-hydroxysteroid dehydrogenase
Adion	Androstenedione
AE	Androgen Excess
AES	Androgen Excess and PCOS Society
AMH	Anti-Müllerian Hormone
ASRM	American Society of Reproductive Medicine
BMI	Body Mass Index
BP	Blood Pressure
Chr	Chromosome
CI	Confidence Interval
CVD	Cardiovascular Disease
CVRF	Cardiovascular Risk Factors
DHEA	Dehydroepiandrosterone
DHEAS	Dehydroepiandrosterone-sulfate
DM2	Type 2 Diabetes Mellitus
DNA	DeoxyriboNucleic Acid
E2	Estradiol
ESHRE	European Society of Reproductive Medicine
FAI	Free Androgen Index
FG	Ferriman Gallway Score
FSH	Follicle-stimulating Hormone
GnRH	Gonadotropin-Releasing Hormone
GRS	Genetic Risk Score
GSK-3b	glycogen synthase kinase 3b
GWAS	Genome-Wide Association Study
gws	genome-wide significant
h2	Heritability
HA	Hyperandrogenemia
HDL	High Density Lipoprotein
HOMA	Homeostasis Model Assessment
I^2	Heterogeneity
IQR	Interquartile Range
IR	Insulin Resistance
LC-MS/MS	Liquid Chromatography–Tandem Mass Spectrometry
LD	Linkage Disequilibrium
LDL	Low Density Lipoprotein
LH	Luteinizing Hormone
MAF	Minor Allele Frequency
MBS	Metabolic Syndrome
n	number
NA	not available
NGS	Next Generation Sequencing
NIH	National Institutes of Health
NTR	Netherlands Twin Registry
OCP	Oral Contraceptive Pill
OD	Ovulatory Dysfunction
OR	Odds Ratio
P	P-value
PCOM	Polycystic Ovarian Morphology
PCOS	Polycystic Ovary Syndrome
PYRS	Person Years
QQ-plot	Quantile Quantile plot
RIA	Radio-Immuno Assay
RNA	RiboNucleic Acid
RR	Relative Risk
SHBG	Sex-Hormone Binding Globulin
SMR	Standardized Mortality Ratio
SNP	Single Nucleotide Polymorphism
T	testosterone
TC	Total Cholesterol
TG	triglycerides
UK	United Kingdom
US	United States of America
VNTR	Variable Number of Tandem Repeats
WC	Waist Circumference
WHR	Waist-Hip Ratio
ZODIAC	Zwolle Outpatient Diabetes Project Integrating Available Care

List of gene name abbreviations

ACTR2	Actin-Related protein 2 homolog
AR	Androgen Receptor
BDNF	brain-derived neurotrophic factor
C14orf177	chromosome 14 open reading frame 177
C14orf64	chromosome 14 open reading frame 64
C15orf54	chromosome 15 open reading frame 54
C9orf3	chromosome 9 open reading frame 3
CAPZA3	capping protein (actin filament) muscle Z-line, alpha 3
CHRM2	cholinergic receptor, muscarinic 2
CYP17	cytochrome P450, family 17
CYP19	cytochrome P450, family 19
CYP3A7	cytochrome P450, family 3, subfamily A, polypeptide 7
DEK	DEK oncogene
DENND1A	DENN domain-containing protein 1A
DPP10	dipeptidyl-peptidase 10
DRD2	dopamine receptor D2
EML6	echinoderm microtubule associated protein like 6
ETV5	ets variant 5
FAIM2	Fas apoptotic inhibitory molecule 2
FBN3	fibrillin
FSHR	follicle-stimulating hormone receptor
FTO	fat mass and obesity associated
GATA4	GATA binding protein 4
GNPDA2	glucosamine-6-phosphate deaminase 2
GTF2A1L	general transcription factor IIA, 1-like
HMGA2	high mobility group AT-hook 2
HSD11B1	11-beta hydroxysteroid dehydrogenase
ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
INSR	Insulin receptor
IRS2	insulin receptor substrate 2
IYD	iodotyrosine deiodinase
KCTD15	potassium channel tetramerization domain containing 15
LHCGR	luteinizing hormone / choriogonadotropin receptor
MC4R	melanocortin 4 receptor
MKX	mohawk homeobox
MTCH2	mitochondrial carrier 2
NEGR1	neuronal growth regulator 1
OLFML2B	olfactomedin-like 2B
PAPSS2	3-phosphoadenosine 5-phospho sulfate synthase isoform 2
PPP1R14C	protein phosphatase 1, regulatory (inhibitor) subunit 14C
RAB1A	Ras-related in Brain type 1A
RAB5B	RAB5B, member RAS oncogene family
RBFOX1	RNA binding protein, fox-1 homolog (C. elegans) 1
RPL7P6	ribosomal protein L7 pseudogene 6
SEC16B	SEC16 homolog B (S. cerevisiae)
SH2B1	SH2B adaptor protein 1
SHBG	Sex Hormone Binding Globulin
STS	steroid sulfatase
SULT2A1	sulfotransferase 2A1
SUMO1P1	SUMO1 pseudogene 1
SUOX	sulfite oxidase
THADA	thyroid adenoma associated

THBS1	thrombospondin 1
TMEM18	transmembrane protein 18
TMPRSS5	transmembrane protease, serine 5
TOX3	TOX high mobility group box family member 3
YAP1	Yes-associated protein 1
ZNF217	zinc finger protein 217

Bibliography

List of manuscripts related to this thesis

N. M. P. Daan, **Y. V. Louwers**, M. P. H. Koster, M. J. C Eijkemans, B. C. J. M. Fauser, J. S. E. Laven. The cardiometabolic profile of non-hyperandrogenic compared to hyperandrogenic patients with PCOS. *Manuscript submitted.*

Y. V. Louwers, M. E. Roest-Schalken, N. Kleefstra, J. Roeters van Lennep, M. van den Berg, B. C. J. M. Fauser, H. J. G. Bilo, E. J. G. Sijbrands, J. S. E. Laven. Severe Excess Mortality in Mothers of Patients with Polycystic Ovary Syndrome. *Manuscript submitted.*

M. O. Goodarzi, **Y. V. Louwers**, K. D. Taylor, M. R. Jones, J. Cui, S. Kwon, Y. D. Chen, X. Guo, L. Stolk, A. G. Uitterlinden, J. S.E. Laven, R. Azziz. Replication of association of a novel insulin receptor gene polymorphism with polycystic ovary syndrome. *Fertility & Sterility*. 2011; Apr; 95: 1736-41.

Y. V. Louwers, F. H. de Jong, N. A. van Herwaarden, L. Stolk, B. C. Fauser, A. G. Uitterlinden, J. S. E. Laven. Variants in *SULT2A1* Affect the DHEA Sulphate to DHEA Ratio in Patients With Polycystic Ovary Syndrome But Not the Hyperandrogenic Phenotype. *Journal of Clinical Endocrinology and Metabolism*. 2013; Sep; 98: 3848-55.

Y. V. Louwers, N. W. Rayner, B.M. Herrera, L. Stolk, C. J. Groves, T. M. Barber, A. G. Uitterlinden, S. Franks, J. S.E. Laven* and M. I. McCarthy* *(contributed equally)*. BMI-associated alleles do not constitute risk alleles for the Polycystic Ovary Syndrome independently of BMI: a case-control study. *Plos One.* 2014, accepted.

Y. V. Louwers, L. Stolk, O. Valkenburg, G. Willemsen, J. J. Hottenga, C. B. Lambalk, D. Boomsma, B. C. M. J. Fauser, A. G. Uitterlinden, J. S. E. Laven. Genome-Wide Association Study in Patients with PCOS from Northern European Descent. *Preliminary results*

Y. V. Louwers, L. Stolk, A. G. Uitterlinden, J. S. E. Laven. Cross-Ethnic Meta-analysis of Genetic Variants for Polycystic Ovary Syndrome. *Journal of Clinical Endocrinology & Metabolism.* 2013; Dec; 98(12): E2006-12.

Y. V. Louwers, O. Lao, B. C. J. M. Fauser, M. Kayser, J. S. E. Laven. The impact of self-reported ethnic background versus genetic ancestry in patients with Polycystic Ovary Syndrome (PCOS). *Manuscript submitted.*

Other manuscripts

J. A. Burgers, S. L. Fong, **Y. V. Louwers**, O. Valkenburg, F. H. de Jong, B. C. Fauser, J. S. E. Laven. Oligoovulatory and anovulatory cycles in women with polycystic ovary syndrome (PCOS): what's the difference? *Journal of Clinical Endocrinology and Metabolism*. 2010; Dec; 95: E485-9.

O. Valkenburg, O. Lao, I. Schipper, **Y. V. Louwers**, A. G. Uitterlinden, M. Kayser, J. S. Laven. Genetic ancestry affects the phenotype of normogonadotropic anovulatory (WHOII) subfertility. *Journal of Clinical Endocrinology and Metabolism*. 2011; Jul; 96: E1181-7.

Z. A. Brown, **Y. V. Louwers**, S. L. Fong, O. Valkenburg, E. Birnie, F. H. de Jong, B. C. Fauser, J. S. E. Laven. The phenotype of polycystic ovary syndrome ameliorates with aging. *Fertility & Sterility* 2011; Nov; 96: 1259-65.

L. Stolk, J. R. Perry, D. I. Chasman, C. He, M. Mangino, P. Sulem, M. Barbalic, L. Broer, E. M. Byrne, F. Ernst, T. Esko, N. Franceschini, D. F. Gudbjartsson, J. J. Hottenga, P. Kraft, P. F. McArdle, E. Porcu, S. Y. Shin, A. V. Smith, S. van Wingerden, G. Zhai, W. V. Zhuang, E. Albrecht, B. Z. Alizadeh, T. Aspelund, S. Bandinelli, L. B. Lauc, J. S. Beckmann, M. Boban, E. Boerwinkle, F. J. Broekmans, A. Burri, H. Campbell, S. J. Chanock, C. Chen, M. C. Cornelis, T. Corre, A. D. Coviello, P. d'Adamo, G. Davies, U. de Faire, E. J. de Geus, I. J. Deary, G. V. Dedoussis, P. Deloukas, S. Ebrahim, G. Eiriksdottir, V. Emilsson, J. G. Eriksson, B. C. Fauser, L. Ferreli, L. Ferrucci, K. Fischer, A. R. Folsom, M. E. Garcia, P. Gasparini, C. Gieger, N. Glazer, D. E. Grobbee, P. Hall, T. Haller, S. E. Hankinson, M. Hass, C. Hayward, A. C. Heath, A. Hofman, E. Ingelsson, A. C. Janssens, A. D. Johnson, D. Karasik, S. L. Kardia, J. Keyzer, D. P. Kiel, I. Kolcic, Z. Kutalik, J. Lahti, S. Lai, T. Laisk, J. S. Laven, D. A. Lawlor, J. Liu, L. M. Lopez, **Y. V. Louwers**, P. K. Magnusson, M. Marongiu, N. G. Martin, I. M. Klaric, C. Masciullo, B. McKnight, S. E. Medland, D. Melzer, V. Mooser, P. Navarro, A. B. Newman, D. R. Nyholt, N. C. Onland-Moret, A. Palotie, G. Pare, A. N. Parker, N. L. Pedersen, P. H. Peeters, G. Pistis, A. S. Plump, O. Polasek, V. J. Pop, B. M. Psaty, K. Raikkonen, E. Rehnberg, J. I. Rotter, I. Rudan, C. Sala, A. Salumets, A. Scuteri, A. Singleton, J. A. Smith, H. Snieder, N. Soranzo, S. N. Stacey, J. M. Starr, M. G. Stathopoulou, K. Stirrups, R. P. Stolk, U. Styrkarsdottir, Y. V. Sun, A. Tenesa, B. Thorand, D. Toniolo, L. Tryggvadottir, K. Tsui, S. Ulivi, R. M. van Dam, Y. T. van der Schouw, C. H. van Gils, P. van Nierop, J. M. Vink, P. M. Visscher, M. Voorhuis, G. Waeber, H. Wallaschofski, H. E. Wichmann, E. Widen, C. J. Wijnands-van Gent, G. Willemsen, J. F. Wilson, B. H. Wolffenbuttel, A. F. Wright, L. M. Yerges-Armstrong, T. Zemunik, L. Zgaga, M. C. Zillikens, M. Zygmunt, T. L. Study, A. M. Arnold, D. I. Boomsma, J. E. Buring, L. Crisponi, E. W. Demerath, V. Gudnason, T. B. Harris, F. B. Hu, D. J. Hunter, L. J. Launer, A. Metspalu, G. W. Montgomery, B. A. Oostra, P. M. Ridker, S. Sanna, D. Schlessinger, T. D. Spector, K. Stefansson, E. A. Streeten, U. Thorsteinsdottir, M. Uda, A. G. Uitterlinden, C. M. van Duijn, H. Volzke, A. Murray, J. M. Murabito, J. A. Visser, K. L. Lunetta. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nature Genetics.* 2012; Mar; 44: 260-8.

E. L. van Houten, J. S. E. Laven, **Y. V. Louwers**, A. McLuskey, A. P. Themmen, J. A. Visser. Bone morphogenetic proteins and the polycystic ovary syndrome. *Journal of Ovarian Reserve*. 2013 Apr 30; 6(1): 32

Y. H. Krul-Poel, C. Snackey, **Y. V. Louwers**, P. Lips, C. B. Lambalk, J. S. E Laven, S. Simsek. The role of vitamin D in metabolic disturbances in polycystic ovary syndrome (PCOS): a systematic review. *European Journal of Endocrinology*. 2013; Oct 23; 169(6): 853-65.

M. G. Mes-Krowinkel, **Y. V. Louwers**, A. G. M. G. J. Mulders, F. H. de Jong, B. C. J. M. Fauser, J. S. E. Laven. Influence of oral contraceptives anthropomorphometric, endocrine and metabolic profiles in anovulatory polycystic ovary syndrome patients. *Manuscript submitted.*

Y. V. Louwers*, W. van Dorp*(contributed equally), T. J. Visser, H. Hooijkaas, H. Tournaye, K. Poppe, J. S. E. Laven. Do patients with PCOS have more thyroid disease? *Manuscript submitted.*

C. M. van der Kruk, **Y. V. Louwers**, F. H. de Jong, E. van den Akker, S. J. C. M. M Neggers, J. S. E. Laven Evaluation of the value of dexamethasone suppression and the Synacthen stimulation tests as a screening tool in women with PCOS. *Manuscript submitted.*

PhD Portfolio Summary

Summary of PhD training and teaching activities

Name PhD student: Yvonne Louwers	PhD period: 2009 - 2013
Erasmus MC Department: Subdivision of Reproductive Medicine, Department of Obstetrics and Gynaecology	Promotor(s): Prof. dr. J.S.E. Laven and Prof. dr. A. G. Uitterlinden
Research School: NIHES, MolMed	Supervisor: Dr. L. Stolk

PHD TRAINING

	Year	ECTS
Research skills		
– Basic methods and reasoning in Biostatistics (Boerhaave, Leiden)	2009	1.4
– Endnote course	2010	0.1
– Regression analysis (Boerhaave, Leiden)	2010	1.4
– Biomedical English Writing Course (David Alexander)	2011	4.0
– Basic data analysis on gene expression arrays (MolMed)	2011	0.7
In-depth courses (e.g. Research school, Medical Training)		
– Genomics in Molecular Medicine (NIHES)	2009	0.7
– Genome Wide association Analysis (NIHES)	2009	1.4
– SNPs and Human Disease (NIHES and MolMed)	2009	1.4
– Principles of Genetic Epidemiology (NIHES)	2010	1.4
– Advances in Genomic Research (NIHES)	2010	1.4
– Adobe Photoshop and Adobe Illustrator course (MolMed)	2011	0.3
– Linux course (NIHES)	2011	0.3
Oral presentations at (inter)national conferences		
– Prof Wladimiroff Symposium, Rotterdam	2010	1.0
– Vereniging voor Fertiliteit Studies spring-meeting, Leuven (*award*)	2010	1.0
– Dutch Society of Reproductive Medicine (DSRM), Amsterdam	2010	1.0
– 3e PCOS Consensus workshop: Women's health aspects of PCOS	2010	1.0
– Prof Wladimiroff Symposium, Rotterdam (*award*)	2011	1.0
– Vereniging voor Fertiliteit Studies (VFS) spring-meeting, Gent	2011	1.0
– Refereeravond Rotterdams Cluster Obstetrie en Gynaecologie	2011	1.0
– American Society for Reproductive Medicine (ASRM), Orlando, Florida USA	2011	1.0
– European Society of Reproductive Medicine (ESHRE), Istanbul, Turkey	2012	1.0

– Prof Wladimiroff Symposium, Rotterdam	2013	1.0
– Endocrine Society for Reproductive Medicine (ENDO), San Francisco USA	2013	1.0
– Vereniging voor Fertiliteitsartsen (VVF), Utrecht	2013	1.0
– Vereniging voor Fertiliteit Studies, Maastricht (*award*)	2013	1.0
Poster presentations at (inter)national conferences		
– European Society of Reproductive Medicine (ESHRE), Rome, Italy	2010	1.0
– American Society of Reproductive Medicine (ASRM), Denver, USA	2010	1.0
– European Society of Reproductive Medicine (ESHRE), Stockholm, Sweden	2011	1.0
– Society of Gynecological Investigation (SGI), San Diego, USA	2012	1.0
– Endocrine Society for Reproductive Medicine (ENDO), Houston, USA	2012	1.0
– Endocrine Society for Reproductive Medicine (ENDO), San Francisco, USA	2013	1.0
– European Society of Reproductive Medicine (ESHRE), London, United Kingdom	2013	1.0
Seminars and workshops		
– Postgraduate Program ASRM: PCOS, origins and destiny, Atlanta, USA	2009	0.2
– Serona Symposia International Foundation: Challenges in ART, Barcelona	2010	0.7
– Erasmus PhD day	2010	0.2
– Masterclass Kinderendocrinologie: PCOS, Amsterdam	2011	0.4
– Annual meeting Androgen Excess and PCOS society, Houston, Texas	2012	0.2
– Österreichischen Gesellschaft für Endokrinologie und Stoffwechsel, Graz	2012	0.4
– Erasmus PhD day	2013	0.2
– AAV Wetenschapsmiddag	2013	0.2
– Weekly research meeting of the department Reproductive Medicine	2009-2013	2.0
– Weekly research meeting of the department Genetic Laboratory	2010-2013	2.0
– Two-weekly multidisciplinary gynecology- endocrinology meeting	2009-2013	2.0

TEACHING ACTIVITIES	Year	ECTS
Lecturing		
– Curriculum Bachelor Medicine, Faculty Medicine and Health Sciences, Erasmus University Medical Center Rotterdam	2009-2013	5.0
– Curriculum Master Medicine, Faculty Medicine and Health Sciences, Erasmus University Medical Centre Rotterdam	2012	1.0
Supervising Master's theses		
– Supervising elective research programs (21 weeks) of medical students: Janneke Burgers, Zoe Brown, Marieke Roest-Schalken, Anne van Zessen, Kayo Hirano, Miranda Krowinkel, Samantha Verburg, Cindy van der Kruk, Karin Vroegindeweij, Jill de Wit	2009-2013	5.0
– Supervising Master Thesis of medical student: N.A.A. van Herwaarden	2011-2012	2.0
– Substitute supervisor of exchange student: Olivera Duric	2013	0.2
Other skills		
– Peer reviewing of articles for scientific journals	2011-2013	2.0
– Coordinator of the PCOS Genetics Consortium	2013	5.0

About the Author

Yvonne Vera Louwers werd op 3 mei 1983 geboren te Rotterdam. Ze groeide op in Delft met haar ouders en jongere zus Karin. Delft – bekend om het Delfts Blauw, Johannes Vermeer, de historische binnenstad én de gezellige cafés – is nog altijd de plaats waar ze met veel plezier woont.

In 2001 behaalde Yvonne het gymnasiumdiploma aan het Stanislascollege. Vanwege een numerus fixus voor de studie Geneeskunde besloot ze zich aan te melden voor de decentrale selectie van de Erasmus Universiteit in Rotterdam om zo meer kans te maken om te worden toegelaten. Dat lukte en in september van hetzelfde jaar kon ze met de studie beginnen. In 2005 studeerde ze af met als doctoraalscriptie: 'Samen zorgen, een exploratief onderzoek naar de samenwerking tussen verschillende hulpverleningsinstanties bij kindermishandeling'. Op 18 januari 2008 behaalde ze cum laude haar artsexamen.

Tijdens de studie maakte ze ruim 2 jaar deel uit van een studententeam, dat werkzaam was op de Palliatieve Zorg Unit van de Daniel den Hoedkliniek. Om zich zo goed mogelijk voor te bereiden op een keuzecoschap in een ontwikkelingsland, besloot ze zich ook op te geven voor de Tropencursus van de stichting STOLA. Deze cursus werd afgerond met het essay: 'Cultuur versus reproductieve gezondheid, vijanden of bondgenoten'. In 2007 vertrok ze naar Malawi voor een gecombineerd gynaecologie/chirurgie coschap onder begeleiding van dr. M. Schutgens. Tijdens een rondreis per jeep werden verschillende ziekenhuizen in het noorden van het land bezocht. Veel bijzondere ervaringen rijker keerde ze naar Nederland terug.

Inmiddels was duidelijk geworden dat de actieve sportcarrière niet langer meer op dezelfde manier zou kunnen worden voortgezet. Na jaren op het hoogste niveau te hebben gespeeld in de korfbalcompetitie met diverse kampioenschappen en interlands, besloot ze te stoppen. Een leven zonder sport is voor haar echter ondenkbaar: ze geeft nog training en is nu een enthousiast racefietser.

Na haar artsexamen in 2008 heeft Yvonne een jaar als ANIOS gewerkt op de afdeling gynaecologie van het Albert Schweitzer Ziekenhuis te Dordrecht. Er bestond voor haar geen twijfel dat ze zich graag verder wilde ontwikkelen in de gynaecologie. Het is ook een bewuste keuze geweest om zich eerst te bekwamen in het doen van wetenschappelijk onderzoek.

Onder begeleiding van prof. dr. J.S.E. Laven en prof. dr. A.G. Uitterlinden begon ze in 2009 als promovendus op de subafdeling Voortplantingsgeneeskunde van het Erasmus MC met een onderzoek gericht op de genetische achtergrond van het polycysteus ovarium syndroom. Het resultaat vindt u beschreven in dit boek. Ze kijkt er nu erg naar uit om op 1 maart 2014 binnen het cluster Rotterdam aan de specialisatie obstetrie en gynaecologie beginnen.

Irene Louwers - Schmidt

About the author

Yvonne Vera Louwers was born on May 3rd, 1983 in Rotterdam and she grew up in Delft. In 2001 she started medical school at the Erasmus University in Rotterdam. For 6 weeks she travelled through Malawi with a Dutch surgeon (Dr. M. Schutgens) for a project which was part of the Clinical Officers Training Program. After her graduation, she worked as a resident at the department of Obstetrics and Gynaecology in the Albert Schweitzer Ziekenhuis in Dordrecht (dr. S. Kooi). In 2009 she started her PhD project at the subdivision Reproductive Medicine at the Department of Obstetrics and Gynaecology at the Erasmus MC University Hospital under supervision of professor J.S.E Laven, professor A.G Uitterlinden and dr. L. Stolk. The studies performed during her PhD are described in this thesis. After the defence of her thesis she will start her clinical training in Obstetrics and Gyneacology at the Reinier de Graaf Groep (dr. H. Bremer and dr. C. Hilders) in Delft.

Dankwoord

HET IS AF! En dus tijd om – met schitterend uitzicht op de Volcán Arenal in Costa Rica, wat een unieke plek – het dankwoord te schrijven. Wat een goed gevoel! Terugkijkend heb ik de afgelopen jaren als uitdagend, leuk, met ups en downs, maar vooral als heel leerzaam ervaren. Graag wil ik de mensen die direct of indirect betrokken zijn geweest bij de totstandkoming van dit proefschrift bedanken. In de eerste plaats gaat uiteraard mijn dank uit naar de patiënten die het COLA spreekuur bezochten en toestemming hebben gegeven voor het gebruik van hun gegevens voor wetenschappelijk onderzoek. Zonder hen was dit proefschrift er nooit geweest.

Geachte professor Laven, beste Joop, de afgelopen jaren hebben we elkaar bijna iedere dag gezien en gesproken. Je enthousiasme en passie voor de wetenschap werken aanstekelijk. Ik bewonder je grote kennis over zoveel aspecten van ons vak en je vermogen om deze kennis meteen te kunnen relateren aan resultaten van onderzoek. Onze overleg momenten leiden vaak tot nieuwe ideeën voor projecten, er is nog genoeg te doen! Ik denk met heel veel plezier terug aan onze discussies over de betekenis van de bevindingen. Tijdens congressen heb je me altijd betrokken bij besprekingen met buitenlandse collega's. Ik heb daar heel veel van geleerd. Ik ben er trots op dat het gelukt is om meer vorm te geven aan het genetics consortium. Ik wil je voor alle steun de afgelopen jaren heel hartelijk bedanken. Ik had me geen betere promotor kunnen wensen!

Geachte professor Uitterlinden, beste André, vanaf het begin bent u betrokken geweest bij mijn promotietraject. Uw kritische blik, grote kennis over genetische epidemiologie en al wat komt kijken bij internationale samenwerking, hebben mij veel geleerd. Na de maandelijkse "PCOS-overleggen" had ik altijd het gevoel dat we een stap vooruit hadden gemaakt. De prettige samenwerking en uw motiverende en stimulerende interesse zijn heel waardevol voor mij geweest. Ik ben dan ook heel dankbaar dat ook u mijn promotor bent.

Beste dr. Stolk, beste Lisette, geen hééééél lange mails meer, het is klaar! Toen ik begon met dit traject wist ik eigenlijk nog niet eens precies wat een SNPs was, laat staan dat ik enig idee had wat de identificatie van SNPs in relatie tot een ziektebeeld betekende. Je hebt me enthousiast meegenomen in de wondere wereld van SNP en GWAS analyses. Bedankt dat je altijd tijd maakte als ik – al dan niet in combinatie met plink, putty, R, metal, BCSNPmax of GRIMP – vastliep. Wat was het leuk om David en jullie op komst zijnde kleine in de 3D sneak preview te zien! Ik ben heel blij dat je mijn copromotor bent!

Beste professor De Jong, professor Fauser en professor Sijbrands, bedankt dat u plaats wilt nemen in de leescommissie. Beste professor De Jong, wat heb ik ongelofelijk veel van u geleerd tijdens de gyn-endo besprekingen en het schrijven van onze gezamenlijke manuscripten. Uw

kennis over (steroid)hormonen is ongekend. Ook wil ik u, als secretaris van de leescommissie, bedanken voor het kritisch lezen van mijn proefschrift. Ik waardeer dat enorm! Beste professor Fauser, dank voor de altijd waardevolle commentaren op mijn artikelen. Ik heb daar veel aan gehad en heb dat zeer op prijs gesteld. Beste professor Sijbrands, ik denk met veel plezier terug aan de uurtjes dat we achter uw "oude" laptop hebben gezeten om analyses te runnen. Ik ben u heel dankbaar dat u hiervoor steeds tijd heeft vrijgemaakt. Het is een manuscript met een belangrijke boodschap geworden.

I also like to thank the other members of my committee. Dear professor Franks, thank you for coming to the Netherlands to be part of my committee, I very much appreciate that. Dear professor Kayser, thank you for learning me so much on genetic ancestry. I'm very happy that you agreed on being part of my commitee. Professor Steegers-Theunissen, bedankt voor uw interesse in en meedenken over mijn projecten, ik heb dat altijd zeer gewaardeerd! Fijn dat u plaats wilt nemen in mijn promotiecommissie.

Beste co-auteurs, bedankt voor de fijne samenwerking en alle hulp en waardevolle feedback op de artikelen in dit proefschrift. Dear co-authors, thank you so much for contributing to the manuscripts in this thesis and other manuscripts. I very much appreciate your input and collaboration.

Beste (ex-)collega's van de afdeling voortplantingsgeneeskunde, Jits, Evert, Susanne, Hikke, Esther, Godfried, Pauline, Berthe, Jenny, Hjalmar, Elisabeth, Nicole, Sofie, Jolanda, Lizka, Geranne, Cindy, Hanneke en Hester, dank voor de belangstelling in mijn projecten, de input tijdens researchbesprekingen, en de gezelligheid op de poli of bij het koffie automaat. Ik heb met veel plezier met jullie samengewerkt en ga jullie zeker missen! Ramon, bedankt voor je enthousiasme, het meedenken, je luisterend oor en het maken van een geintje op z'n tijd. Dat gezamenlijke artikel gaat er zeker komen! Remko, Beate, Simone, Kris en anderen van de backoffice en alle medewerkers van de prikkamer, dank voor al jullie geregel rondom het COLA spreekuur en de gezelligheid op de poli. Beste Jolanda, een speciaal woord van dank voor jou: zeker als secretaresse van Joop, maar voor zoveel meer (praktische) zaken was je onmisbaar tijdens mijn promotietraject! Bedankt!

Beste collega's van het genetica lab, en in het bijzonder Joyce, Fernando, Marjolein, Michael, Ramazan, Marijn, Robert, Annemieke, Jeroen en Mila, ik kon altijd met vragen of problemen bij jullie terecht en jullie waren altijd bereid tot meedenken en meehelpen. Heel erg bedankt!

Dear members of the genetic work discussion group, thank you for your feedback during the work discussions. These meetings have been very helpful!

Dear Oscar, it's amazing how you can make very complex statistical formulas sound easy. I want to thank you for all your help with the genetic ancestry paper.

Dear members of the PCOS genetics consortium, I'm very excited that the consortium is really coming together! Thank you so much for your collaboration. I feel privileged to be part of this consortium and I can't wait for the first results.

En natuurlijk verdienen mijn collega's uit Utrecht een plek in dit dankwoord. Lieve Femi, Marlies en Nadine, wat hebben we heel wat interessante en uitdagende klusjes moeten klaren – DNA samples uitzoeken, multi-etnische GWAS analyses, correctiefactoren voor verschillende assays – maar wat is het fijn dat in leuk gezelschap te doen. Bedankt! Het zou leuk zijn als onze wegen zich nog eens kruisen (misschien als collega- AIOS?)!

Beste collega's uit het Albert Schweitzer Ziekenhuis in Dordrecht, ik kijk met heel veel plezier terug op de tijd die ik bij jullie werkte als ANIOS. Wat een fijne gezellige plek om te werken. Ik heb heel veel van jullie geleerd, bedankt daarvoor! Beste Dr. Kooi, beste Sjarlot, bedankt voor het vertrouwen en de steun.

Beste Dr. Bremer, beste Dr. Hilders en andere nieuwe collega's uit het Reinier De Graaf ziekenhuis in Delft, per 1 maart start ik bij jullie als AIOS. Ik heb ongelofelijk veel zin in deze nieuwe uitdaging!

Lieve collega AIOSsen, ik kijk alweer uit naar het volgende assistentenweekend! En de leuke gezellige borrels daarvoor natuurlijk!

Lieve Sharon, Lindy, Annelinde, Mariëlle, Olivier en Wendy, wat was Hs508 een geweldige werkplek: vieren wanneer er iets te vieren viel (natuurlijk met Koekela), meeleven, meedenken, een bemoedigend woord als het even tegen zat, hard werken, maar zeker ook veel lachen! Zonder jullie was het promotietraject veel minder leuk geweest. Dank dank dank! Ik hoop (en weet eigenlijk wel zeker!) dat het blijft lukken om elkaar regelmatig te zien en bij te praten.

En natuurlijk wil ik andere alle (oud)-mede-onderzoekers – Nicole, Melek, Evelyne, Kim, Emilie, Bas, Marit, Jennifer, Leonie, Ruben, Sam, Hein, Robbert, Annelien, Nienke, Manon, Babette, John, Marieke, Zoe, Averil, Nina, Pauline, Charlotte, Irene, Caroline, Jinke, Sylvia, Claudia, Jashvant, Babs, Marijana, Sabine, Anke, Amber, Chantal, Sevilay, Maria, Matthijs, Sarah en Fatima… de lijst wordt steeds langer – heel erg bedanken voor jullie interesse en natuurlijk ook de gezellige koffies, lunches en borrels. Succes met jullie onderzoek en carrière. Lieve Christine, die gezellige etentjes (Istanbul, Rotterdam, Delft, Eindhoven de gekste, Ede?) houden we erin! Jeroen, mede-kantoortuin-genoot, hoe moet dat straks met het gat in mijn ontwikkeling? Enne, vanaf nu valt de plant onder jouw verantwoordelijkheid!

Lieve Wendy, onze promotietrajecten zijn heel erg gelijk opgegaan. En hoe leuk is het dat we nu vlak na elkaar promoveren en met de opleiding starten. Ik wil je bedanken voor je steun, betrokkenheid en de gezelligheid! Begonnen als collega's, geëindigd als vrienden, beter kan niet. Ik ben heel blij dat je straks als paranimf naast me staat.

Om op het werk goed te kunnen presteren, is ontspanning met familie en vrienden natuurlijk belangrijk. Indirect hebben zij dan ook zeker bijgedragen aan dit proefschrift. Lieve Mirjam, Jennifer, Denise, Petra, Ellen, Vanessa en Robbert, wat ooit begonnen is als samen korfballen, is uitgegroeid tot vriendschap. En ook al zien we elkaar niet meer iedere week (en eigenlijk gewoon veel te weinig!!), ik hecht heel veel waarde aan onze vriendschap. Lieve Debby, ik heb je gemist bij het fietsen afgelopen jaar, maar gelukkig zijn daar etentjes en drankjes drinken voor in de plaats gekomen. Misschien minder sportief, maar net zo gezellig! Lieve Karin, wanneer gaat er gekookt worden in je nieuwe keuken? Lieve Irene, jammer dat je er niet bij kunt zijn vandaag, maar we kletsen snel bij. Leo, Nathalie en Michael, er zit weer een soort van regelmaat in onze dates! Ik zou het heel leuk vinden als we dat vol kunnen houden.

Lieve Jo en Ray, de weekendjes weg, etentjes, tochten met de racefiets of mountainbike en natuurlijk de geweldige reizen samen zijn een feestje! Ik ben heel blij met jullie! Wanneer gaan we weer?

Lieve familie Verheezen – Van der Kruk, vanaf het eerste moment dat ik bij jullie over de vloer kwam heb ik me helemaal thuis gevoeld. Ik kan me geen fijnere schoonfamilie wensen! Lieve Laris, wat leuk om er een schoon-nicht (bestaat dat?) bij te hebben!! Heel fijn dat je lekker in Delft blijft wonen! Lieve Nika en Senne, komen jullie snel weer logeren?

Ik ken maar weinig families die waterpistool-gevechten houden en ieder jaar dezelfde sinterklaas foto maken… wat is het boffen dat ik nu juist zo'n familie getroffen heb! Lieve Dick en Anja, Marc en Nienke, Leanne en Jeroen, lieve tante Hilly en Leo: met elkaar meeleven is als vanzelfsprekend, en dat is heel bijzonder en waardevol. Lieve Frans en Marian, we fietsen snel weer langs!

Lieve Karin, zus en paranimf, ik denk nog vaak terug aan onze Malawi reis, wat was dat geweldig om samen meegemaakt te hebben. Werk, studie, korfbal en nog veel meer, het is ongelofelijk hoe jij alle ballen in de lucht weet te houden en steeds nieuwe uitdagingen ziet. Ik ben hartstikke trots op je en ben heel blij dat je als paranimf naast me staat. Lieve Marcel, hoe jij al je bezigheden weet te combineren zonder iets half te doen (zouden jullie daarom zo goed bij elkaar passen?), vind ik ongelofelijk knap. En je staat altijd klaar wanneer dat nodig is, je bent een top zwager!

Lieve pap en mam, dank voor jullie onvoorwaardelijke steun, meeleven en meedenken. Ik kan altijd bij jullie terecht en zonder jullie was ik niet geweest wie ik nu ben. Hou van jullie!

Lieve Amanda, soms weet je meteen dat iets goed zit. Wat vullen we elkaar goed aan en wat hebben we het leuk! Op naar nog heel veel meer mooie jaren samen! x

Yvonne
december 2013