Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome

Article *in* Nature Genetics · January 2011

DOI: 10.1038/ng.732 · Source: PubMed

34 authors, including:

Han Zhao
Tongji University
92 PUBLICATIONS **2,092** CITATIONS
SEE PROFILE

Lin He
Dongguan University of Technology
821 PUBLICATIONS **24,304** CITATIONS
SEE PROFILE

Yuhua Shi
Shandong University
56 PUBLICATIONS **1,659** CITATIONS
SEE PROFILE

Zhiqiang Li
Shanghai Jiao Tong University
166 PUBLICATIONS **4,364** CITATIONS
SEE PROFILE

nature genetics

Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3

Zi-Jiang Chen[1,2], Han Zhao[1,2,25], Lin He[3–5,25], Yuhua Shi[1,2,25], Yingying Qin[1,2], Yongyong Shi[4], Zhiqiang Li[4], Li You[1,2], Junli Zhao[6], Jiayin Liu[7], Xiaoyan Liang[8], Xiaoming Zhao[9], Junzhao Zhao[10], Yingpu Sun[11], Bo Zhang[12], Hong Jiang[13], Dongni Zhao[14], Yuehong Bian[1,2], Xuan Gao[1,2], Ling Geng[1,2], Yiran Li[15,16], Dongyi Zhu[17], Xiuqin Sun[18], Jin-e Xu[19], Cuifang Hao[20], Chun-e Ren[16], Yajie Zhang[21], Shiling Chen[22], Wei Zhang[23], Aijun Yang[24], Junhao Yan[1,2], Yuan Li[1,2], Jinlong Ma[1,2] & Yueran Zhao[1,2]

Polycystic ovary syndrome (PCOS) is a common metabolic disorder in women. To identify causative genes, we conducted a genome-wide association study (GWAS) of PCOS in Han Chinese. The discovery set included 744 PCOS cases and 895 controls; subsequent replications involved two independent cohorts (2,840 PCOS cases and 5,012 controls from northern Han Chinese; 498 cases and 780 controls from southern and central Han Chinese). We identified strong evidence of associations between PCOS and three loci: 2p16.3 (rs13405728; combined *P*-value by meta-analysis $P_{meta} = 7.55 \times 10^{-21}$, odds ratio (OR) 0.71); 2p21 (rs13429458, $P_{meta} = 1.73 \times 10^{-23}$, OR 0.67); and 9q33.3 (rs2479106, $P_{meta} = 8.12 \times 10^{-19}$, OR 1.34). These findings provide new insight into the pathogenesis of PCOS. Follow-up studies of the candidate genes in these regions are recommended.

PCOS is characterized by the presence of two or more of the following features: chronic oligo-ovulation or anovulation, androgen excess and polycystic ovaries[1]. As the most common cause of anovulatory infertility, PCOS affects 6%–8% of childbearing-aged women[2,3]. PCOS is also associated with endocrine-metabolic derangements leading to a broad range of adverse sequelae that include dyslipidemia, atherosclerosis, insulin resistance and type 2 diabetes (T2D)[4–6]. Among individuals with PCOS diagnosed as having impaired glucose tolerance (IGT) and diabetes mellitus, approximately 20% were recognized at a younger age to have PCOS[7–9].

The pathogenesis of PCOS is not fully understood. Heritable tendencies have long been recognized, but complex interactions exist between genetic and environmental factors. Association studies have been conducted on at least 70 candidate genes, principally related to reproductive hormones, insulin resistance and chronic inflammation—for example, follicle stimulating hormone receptor (*FSHR*), insulin receptor (*INSR*) and interleukin-6 (*IL6*)[10–13]. However, few genes influencing susceptibility to PCOS have been determined. Array-based genome-wide SNP association analyses should provide a more comprehensive, unbiased approach to detect susceptibility genes for diseases like PCOS[14].

We conducted a two-stage GWAS to identify genetic markers for PCOS in a Han Chinese population. The initial discovery set for GWAS consisted of 744 PCOS cases and 895 controls. The second stage of the replication study included two independent cohorts: 2,840 PCOS cases and 5,012 controls from northern Han Chinese (Replication 1) and 498 cases and 780 controls from southern and central Han Chinese (Replication 2)[15]. All women defined as having PCOS were diagnosed according to the Revised 2003 Consensus on Diagnostic Criteria and Long-term Health Risks Related to Polycystic Ovary Syndrome[1]. Any two of the following three criteria were required to be present: oligo- and/or anovulation, clinical and/or biochemical signs of hyperandrogenism, and ovarian morphology showing characteristic polycystic features on ultrasound. Other causes of oligomenorrhea or hyperandrogenism (for example, nonclassical

[1]Center for Reproductive Medicine, Shandong Provincial Hospital, Shandong University, Jinan, Shandong, China. [2]Shandong Key Laboratory of Reproductive Medicine, Jinan, Shandong, China. [3]Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [4]Bio-X Center, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China. [5]Institute for Nutritional Sciences, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [6]Affiliated Hospital of Ningxia Medical University, Ningxia, China. [7]The First Affiliated Hospital with Nanjing Medical University, Jiangsu, China. [8]The Sixth Affiliated Hospital of Sun Yat-sen University, Guangdong, China. [9]Renji Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China. [10]The First Affiliated Hospital of Wenzhou Medical College, Zhejiang, China. [11]The First Affiliated Hospital of Zhengzhou University, Henan, China. [12]Maternal and Child Health Hospital in Guangxi, Guangxi, China. [13]105th Hospital of People's Liberation Army, Anhui, China. [14]Shengjing Hospital of China Medical University, Liaoning, China. [15]Qingdao Women & Children Medical Healthcare Center, Shandong, China. [16]Affiliated Hospital of Weifang Medical College, Shandong, China. [17]Linyi People's Hospital, Shandong, China. [18]Shandong Jingning First People's Hospital, Shandong, China. [19]The Affiliated Hospital of Medical College Qingdao University, Shandong, China. [20]Yantai Yuhuangding Hospital, Shandong, China. [21]Jinan Health Institute of Maternity and Infant, Shandong, China. [22]Nanfang Hospital, Guangdong, China. [23]Obstetrics & Gynecology Hospital of Fudan University, Shanghai, China. [24]Affiliated Hospital of Jining Medical College, Shandong, China. [25]These authors contributed equally to this work. Correspondence should be addressed to Z.-J.C. (zjchen59@yahoo.com) or Yongyong Shi (shiyongyong@gmail.com).

Received 13 July; accepted 3 November; published online 12 December 2010; doi:10.1038/ng.732

Table 1 Characteristics of PCOS case and control subjects

Characteristic	Discovery set (NHC[a])	Replication 1 (NHC)	Replication 2 (S-CHC)
Cases			
Number	744	2,840	498
Age (years)[b]	28.85 ± 3.62	28.28 ± 3.99	27.00 ± 4.00
BMI	24.55 ± 3.99	24.76 ± 4.74	23.28 ± 4.13
T (ng dl^{-1})	81.11 ± 21.03	61.17 ± 27.27	61.56 ± 34.51
HOMA-IR	2.27 ± 1.72	2.64 ± 2.13	3.84 ± 3.56
Controls			
Number	895	5,012	780
Age (years)	30.68 ± 4.68	31.38 ± 7.73	31.31 ± 4.07
BMI	22.68 ± 3.23	22.35 ± 3.63	21.61 ± 2.63

[a]NHC, northern Han Chinese; S-CHC, southern and central Han Chinese; BMI, body mass index; T, testosterone; HOMA-IR, homeostasis model assessment–insulin resistance; values expressed as mean ± s.d. [b]Age of subjects with PCOS was that at diagnosis.

21-hydroxylase deficiency, Cushing's syndrome, hypothyroidism, significant elevations in serum prolactin) were excluded on clinical grounds. Controls were gathered primarily from healthy women who presented with regular menstrual cycles, excluding hyperandrogenism and PCOS. Characteristics of the case and control cohorts are described in **Table 1**, with more details concerning subjects used in the discovery and replication series provided in Online Methods.

In the initial discovery phase, we genotyped 744 affected individuals (cases) and 895 controls using Affymetrix SNP 6.0 chips. Principal component analysis (PCA) was used to evaluate the population structure of the samples in comparison to the HapMap 270 individuals. No obvious population substructure was observed in our samples (**Supplementary Fig. 1**). After quality control filtering (Online Methods), a total of 611,633 SNPs were subjected to statistical analysis. A quantile-quantile (Q-Q) plot was used to assess the number and magnitude of observed associations between genotyped SNPs and PCOS, compared to the statistics expected under the null hypothesis of no association. The Q-Q plot of the $-\log_{10}$ of unadjusted *P*-values (**Supplementary Fig. 2**) showed some evidence for inflation due to population stratification (genomic inflation factor $\lambda = 1.097$; **Supplementary Fig. 3**). To exclude the possibility of a false-positive association caused by population stratification, the observed test statistics were adjusted using the PCA method[16]. In an initial analysis, four distinct regions comprising 29 SNPs (1 in 2p16.3, 21 in 2p21, 1 in 5q14.3 and 6 in 9q33.3) showed strong evidence of association at PCA-adjusted $P < 5 \times 10^{-6}$, including 9 SNPs (7 in 2p21 and 2 in 9q33.3) that met genome wide significant association at PCA-adjusted $P < 5 \times 10^{-8}$ (**Fig. 1**).

To further validate the GWAS findings, we genotyped subjects from two independent replication sets and also carried out a combined analysis for those 29 SNPs mentioned above (**Supplementary Table 1**). In Replication 1, involving northern Chinese, 28 of 29 SNPs still had a significant association ($P < 10^{-3}$) with PCOS (**Supplementary Table 1**); 1 SNP in 5q14.3 was not replicated. In Replication 2, 20 of the SNPs showed nominal significance ($P < 0.05$) for southern and central Han subjects (**Supplementary Table 1**). In the combined study, we performed a meta-analysis including GWAS, Replication 1 and Replication 2 data. Heterogeneity across the three data sets was evaluated using Cochran's *Q* test, with no statistically significant heterogeneity observed among the stages ($P > 0.05$). Therefore, the fixed-effect model was chosen in the meta-analysis, and the Mantel-Haenszel method was used to estimate the pooled odds ratio for all strata. All 28 SNPs within 2p16.3, 2p21 and 9q33.3 showed genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-analysis. The most significant signals of those three regions were observed at SNP rs13405728 (2p16.3, $P_{meta} = 7.55 \times 10^{-21}$, OR 0.71, 95% confidence interval (c.i.) 0.67–0.77), rs13429458 (2p21, $P_{meta} = 1.73 \times 10^{-23}$, OR 0.67, 95% c.i. 0.62–0.72) and rs2479106 (9q33.3, $P_{meta} = 8.12 \times 10^{-19}$, OR 1.34, 95% c.i. 1.26–1.43) (**Table 2**). Those results strongly support the associations between 2p16.3, 2p21, 9q33.3 and PCOS.

To exclude the potential confounding effects of age and body mass index (BMI), we conducted a logistic regression analysis of disease traits using age and BMI as covariates. For the combined results, the adjusted *P*-values of the 28 associated SNPs within the three susceptibility loci still showed genome-wide significance ($P < 5 \times 10^{-8}$; **Supplementary Table 2**); therefore, the associations of 2p16.3, 2p21 and 9q33.3 with PCOS could not be accounted for by age and BMI. However, previous GWAS reports have revealed a series of candidate regions for BMI (**Supplementary Table 3**). We thus carefully reviewed all published GWAS reports involving BMI, but found no region that overlapped with the findings of the current study.

At 2p16.3, the most significant SNP is rs13405728. Several known genes are located nearby (±500 kb of rs13405728; **Fig. 2a**), and a small linkage disequilibrium (LD) block approximately 53.4 kb in size contains both the *GTF2A1L* (TFIIA-alpha and beta-like factor) and *LHCGR* genes (**Supplementary Fig. 4**). *GTF2A1L* is germ cell specific and may be important with respect to the biology of the testis; abnormal expression of *GTF2A1L* might specifically be a cause for human infertility[17]. *LHCCR* 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 preovulatory follicles. Notably, the induction of *LHCGR* during granulosa cell differentiation allows the preovulatory 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[18,19]. In contrast, activating mutations in affected women produce hyperandrogenism, yet no other effect on reproduction[20]. In the current study, we also investigated the characteristic clinical features BMI, testosterone and homeostasis model assessment–insulin resistance (HOMA-IR) in PCOS patients carrying different rs13405728

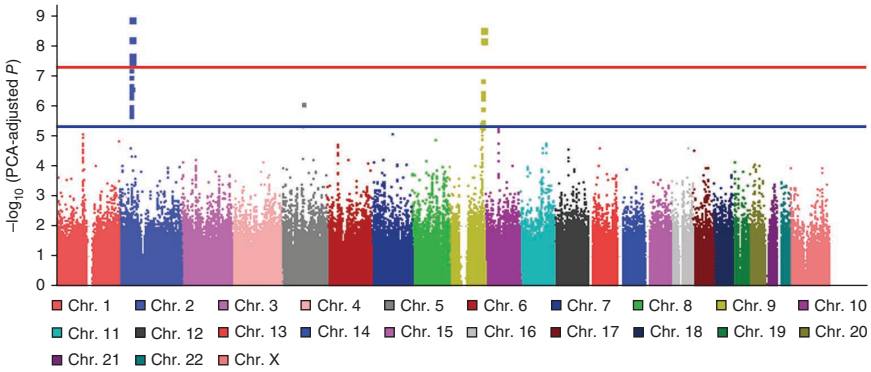

Figure 1 Genome-wide association scan for PCOS. Negative $\log_{10}$ *P*-values adjusted by principal component analysis (PCA) are shown for SNPs that passed quality control. The red line (5×10^{-8}) is the global significance level, whereas the blue line (5×10^{-6}) is the threshold used in our study. Chr., chromosome.

Table 2 GWAS, case-control replication study and meta-analysis results for the most significant SNPs

				GWAS 744 cases, 895 controls				Replication 1 2,840 cases, 5,012 controls				Replication 2 498 cases, 780 controls				Meta-analysis	
				MAF				MAF				MAF				OR 95% c.i.	
SNP	Chr.	Position	Allele[a]	Case	Control	OR	*P*-value[b]	Case	Control	OR	*P*-value	Case	Control	OR	*P*-value		*P*-value
rs13405728	2p16.3	48831663	G/A	0.188	0.274	0.61	2.54×10^{-7}	0.192	0.237	0.76	3.11×10^{-10}	0.181	0.269	0.60	7.93×10^{-7}	0.71 0.67–0.77	7.55×10^{-21}
rs12468394	2p21	43414665	A/C	0.207	0.305	0.60	1.20×10^{-9}	0.230	0.281	0.76	7.42×10^{-11}	0.209	0.274	0.70	0.00042	0.72 0.68–0.78	1.59×10^{-20}
rs13429458	2p21	43492342	C/A	0.132	0.207	0.59	1.05×10^{-7}	0.134	0.186	0.68	5.55×10^{-16}	0.153	0.196	0.74	0.0073	0.67 0.62–0.72	1.73×10^{-23}
rs12478601	2p21	43575012	T/C	0.220	0.314	0.61	5.55×10^{-9}	0.235	0.291	0.75	2.80×10^{-13}	0.228	0.298	0.69	0.00014	0.72 0.67–0.77	3.48×10^{-23}
rs10818854	9q33.3	125486599	A/G	0.135	0.079	1.80	1.20×10^{-6}	0.123	0.087	1.46	8.08×10^{-12}	0.097	0.073	1.36	0.042	1.51 1.37–1.65	9.40×10^{-18}
rs2479106	9q33.3	125565033	G/A	0.294	0.216	1.51	5.09×10^{-7}	0.276	0.223	1.33	4.59×10^{-13}	0.252	0.219	1.21	0.059	1.34 1.26–1.43	8.12×10^{-19}
rs10986105	9q33.3	125589776	C/A	0.134	0.069	2.08	6.13×10^{-9}	0.112	0.083	1.39	8.20×10^{-9}	0.085	0.073	1.17	0.31	1.47 1.33–1.61	6.90×10^{-15}

Chr., chromosome; MAF, minor allele frequency.
[a]Minor allele/major allele. [b]PCA-adjusted *P*-values.

genotypes. We did not find a significant difference in comparison to those PCOS patients lacking the risk allele (**Supplementary Table 4**). Therefore, SNP rs13405728 is associated with susceptibility to PCOS as defined by the Rotterdam criteria. However, this SNP does not account for other clinical phenotypes frequently found in PCOS (for example, increased BMI or HOMA-IR).

Of note, the *FSHR* gene is situated 211 kb downstream of rs13405728. Previous studies have shown that *FSHR* is associated with the ovarian response to FSH, thus being a compelling candidate gene for PCOS[10]. The SNP rs6166 (N680S), located in exon 10 of *FSHR*, has been reported to be associated with PCOS in Dutch and Japanese women[10,21], but not in Chinese[22,23]. We also checked the significance of 65 SNPs located in *FSHR* in our array data; we found that 13 of 65 SNPs had a PCA-adjusted *P*-value between 2×10^{-3} and 4×10^{-4}; therefore, a role for *FSHR* could not be excluded (**Supplementary Table 5**). *FSHR* is separated from the associated LD block that contains *LHCGR* by a strong recombination hotspot and thus lies in a different block than rs13405728 (**Fig. 2a**). However, it is possible that SNPs in *LHCGR* could influence expression of *FSHR* even though the genes are located in different LD blocks.

At 2p21 there are 21 significant SNPs ($2.36 \times 10^{-13} \leq P_{meta} \leq 1.73 \times 10^{-23}$; **Supplementary Table 1**) located within several known genes (*ZFP36L2*, *LOC100129726* and *THADA*; **Fig. 2b**). These SNPs span 304 kb and are located in two different LD blocks: one LD block of approximately 186 kb that includes *ZFP36L2*, *LOC100129726* and the 3′ region of *THADA*, and a second block of approximately 197 kb that is within *THADA* (**Supplementary Fig. 5**). A conditional logistic regression analysis adjusted for stages (see details in Online Methods) was also carried out to test for independent effects of individual SNPs in 2p21. The analysis revealed two independent associations, which can be represented by rs13429458 and rs12478601, both located in *THADA*. The estimated *P*-value of rs12478601 was 1.87×10^{-7}, conditioned on rs13429458. The effect estimates of the remaining SNPs in this region showed no association ($P > 0.05$), conditioned on both

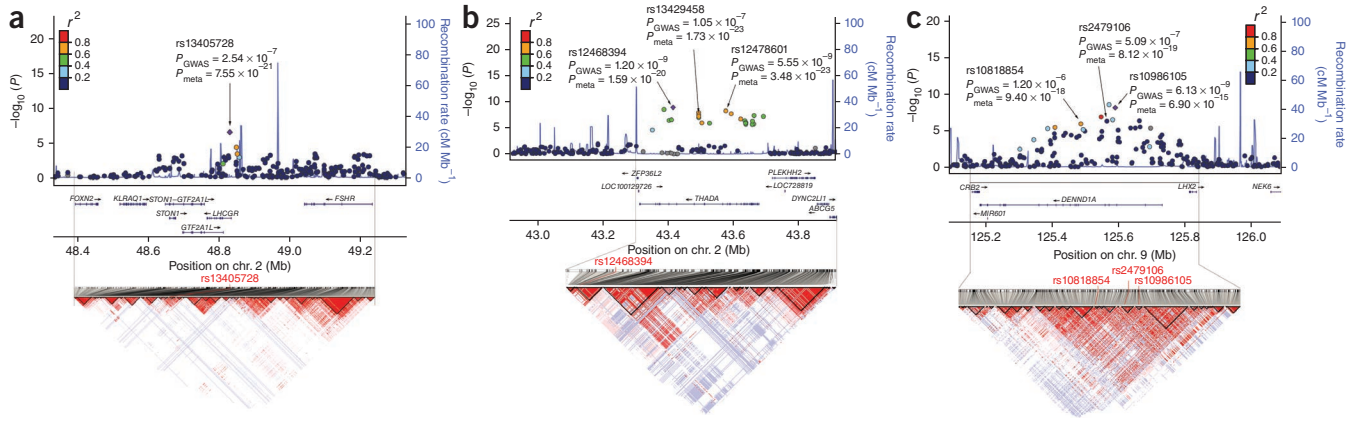

Figure 2 Regional plots of the three newly discovered PCOS loci (2p16.3, 2p21 and 9q33.3). (**a**–**c**) Genotyped SNPs passing quality control measures in GWAS plotted with the *P*-values (as $-\log_{10}$ values) as a function of genomic position (in the University of California Santa Cruz March 2006 human reference sequence, hg18). (**a**) 2p16.3. (**b**) 2p21. (**c**) 9q33.3. P_{GWAS} and P_{meta} represent the *P*-values of GWAS and meta-analysis. In each panel, the index association SNP is represented by a diamond. Estimated recombination rates (taken from HapMap) are plotted to reflect the local LD structure. Color of the remaining SNPs (circles) indicates LD with the index SNP according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pairwise r^2 values from HapMap JPT (Japanese in Tokyo) + CHB; gray, no r^2 value available. Gene annotations were taken from the University of California Santa Cruz genome browser. LD blocks were obtained from the HapMap project (release 22, CHB + JPT).

of these two SNPs (**Supplementary Table 6**). Both rs13429458 and rs12478601 are located in the introns of *THADA*. *THADA* was initially identified in thyroid adenomas by chromosomal rearrangements resulting in disruption of *THADA* and fusion to an intron of peroxisome proliferator-activated receptor gamma (*PPARG*)[24]. Recently, *THADA* SNP rs7578597 was reported to be associated with T2D in a large European GWAS[25], but this failed to be replicated in Chinese subjects with T2D[26]; SNP rs7578597 is a rare variant in Asians. We calculated LD between rs7582497 (a marker for the T2D susceptibility locus) and the PCOS-associated SNPs in CHB (Han Chinese in Beijing) samples (HapMap phase 3; **Supplementary Table 7**): all r^2 values were less than 0.02. However, in CEU (Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection) samples, the LD values are much tighter; for example, $r^2 = 0.976$ between rs7582497 and rs7568365, and $r^2 = 1$ between rs7582497 and rs7567607. This indicates that it would be worthwhile to test whether this region is also associated with PCOS in the European population, potentially revealing risk factors shared between PCOS and T2D. Of note, we did not find subjects with PCOS who carried the *THADA* risk allele to have more severe insulin resistance (**Supplementary Table 4**).

Another region showing a significant association with PCOS is on chromosome 9q33.3. The six significant SNPs ($1.42 \times 10^{-13} \leq P_{meta} \leq 8.12 \times 10^{-19}$; **Supplementary Table 1**) were located within a 42.3-kb region in a LD block within *DENND1A* (**Fig. 2c**). Conditional logistic regression analysis showed two independent associations, which can be represented by rs10818854 and rs2479106, both within *DENND1A*. Conditioned on rs2479106, the estimated *P*-value of rs10818854 was 1.46×10^{-7}. Further conditioned on these two SNPs, the rest of the SNPs in this region showed no association ($P > 0.05$; **Supplementary Table 6**). *DENND1A* encodes a domain differentially expressed in normal and neoplastic cells (DENN) that can bind to endoplasmic reticulum aminopeptidase 1 (ERAP1) as a negative regulator through a death domain–death domain interaction[27]. Serum ERAP1 elevation has been previously reported to be associated with PCOS accompanied by obesity[28]. We speculate that in PCOS patients having the *DENND1A* risk allele, DENND1A might influence the pathogenesis of PCOS through misregulation of ERAP1.

We also focused on PCOS candidate genes with SNPs that had been previously studied among subjects with PCOS in at least one report. There were 51 such genes containing SNPs on the Affymetrix SNP6.0 chip, and we determined that 85 SNPs (out of 755) within 17 genes (out of the 51) had PCA-adjusted *P*-values less than 0.05 (**Supplementary Table 5**). However, none of these 85 SNPs has been previously studied or reported to be directly associated with PCOS.

Our conditional logistic regression analysis twice revealed more than one independent association signal, namely in 2p21 (rs13429458 and rs12478601) and 9q33.3 (rs10818854 and rs2479106). This might indicate two different risk genes for PCOS in each of these regions. However, both rs13429458 and rs12478601 are located in *THADA*, and both rs10818854 and rs2479106 are located in *DENND1A*; thus, these two genes remain the most probable candidate genes in their regions. It is also possible that more than one functional variant exists in these two genes, and the two variants separately contribute to the risk of PCOS. Irrespective, these results provide useful information for follow-up fine-mapping studies in those significant regions, which may reveal the candidate genes and functional variants.

In summary, our GWAS and subsequent replication studies identified three new susceptibility loci (2p16.3, 2p21 and 9q33.3) for polycystic ovary syndrome in Han Chinese.

URLs. R, http://www.r-project.org/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Haploview, http://www.broad.mit.edu/mpg/haploview/; The International HapMap Project, http://www.hapmap.org/; SHEsis, http://analysis.bio-x.cn/myAnalysis.php.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all participants involved in this study. We thank J. Simpson for revising this manuscript and X. Xu, X. Xing, T. Li, M. Guo, L. Cui, Q. Zheng, C. Li, J. Zhang, D. Wu, C. Zhang, X. Yan, W. He, Y. Cui, M. Xia, J. Li, P. Wang, H. Lv, S. Xu and L. Wang for subject recruitment. This study was supported by grants from the National Basic Research Program of China (973 Program-2006CB944004, 2010CB945000, 2007CB947403, 2007CB914703, 2007CB947300, 2010CB529600), the National 863 Project of China grants (2006AA02A407, 2009AA022701), the National Natural Science Foundation of China (30973170) and the Shanghai Municipal Commission of Science and Technology Program (09DJ1400601).

AUTHOR CONTRIBUTIONS

Z.-J.C., L.H. and Yongyong Shi designed the study and revised the manuscript. Z.-J.C. supervised patients' diagnosis, subject recruitment and performance of experiments. Yongyong Shi supervised the experiments and data analysis. H.Z. and Z.L. conducted data analyses and drafted the manuscript. H.Z., Yuhua Shi, Y.Q., L.Y., L.G. and J.Y. recruited subjects. Junli Zhao, J.L., X.L., X.Z., Junzhao Zhao, Y. Sun, B.Z., H.J., D. Zhao, Yiran Li, D. Zhu, X.S., J.-e.X., C.H., C.-e.R., Y. Zhang, S.C., W.Z. and A.Y. coordinated and provided samples from different hospitals. L.Y., Y.B., Yuan Li, J.M. and Y. Zhao performed DNA extraction. X.G. performed endocrine biochemical examination. All authors critically reviewed the article and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and longterm health risks related to polycystic ovary syndrome. *Fertil. Steril.* **81**, 19–25 (2004).
2. Goodarzi, M.O. & Azziz, R. Diagnosis, epidemiology, and genetics of the polycystic ovary syndrome. *Best Pract. Res. Clin. Endocrinol. Metab.* **20**, 193–205 (2006).
3. Ehrmann, D.A., Barnes, R.B., Rosenfield, R.L., Cavaghan, M.K. & Imperial, J. Prevalence of impaired glucose tolerance and diabetes in women with polycystic ovary syndrome. *Diabetes Care* **22**, 141–146 (1999).
4. Carmina, E. Cardiovascular risk and events in polycystic ovary syndrome. *Climacteric* **12** (suppl. 1), 22–25 (2009).
5. Kandaraki, E., Christakou, C. & Diamanti-Kandarakis, E. Metabolic syndrome and polycystic ovary syndrome…; and vice versa. *Arq. Bras. Endocrinol. Metabol.* **53**, 227–237 (2009).
6. Wild, S., Pierpoint, T., Jacobs, H. & McKeigue, P. Long-term consequences of polycystic ovary syndrome: results of a 31 year follow-up study. *Hum. Fertil. (Camb.)* **3**, 101–105 (2000).
7. Espinós-Gómez, J.J., Corcoy, R. & Calaf, J. Prevalence and predictors of abnormal glucose metabolism in Mediterranean women with polycystic ovary syndrome. *Gynecol. Endocrinol.* **25**, 199–204 (2009).
8. Kulshreshtha, B. *et al.* Insulin response to oral glucose in healthy, lean young women and patients with polycystic ovary syndrome. *Gynecol. Endocrinol.* **24**, 637–643 (2008).
9. Shi, Y. *et al.* Analysis of clinical characteristics in large-scale Chinese women with polycystic ovary syndrome. *Neuroendocrinol. Lett.* **28**, 807–810 (2007).
10. Sudo, S. *et al.* Genetic and functional analyses of polymorphisms in the human FSH receptor gene. *Mol. Hum. Reprod.* **8**, 893–899 (2002).
11. Wang, Y., Wu, X., Cao, Y., Yi, L. & Chen, J. A microsatellite polymorphism (tttta)n in the promoter of the CYP11a gene in Chinese women with polycystic ovary syndrome. *Fertil. Steril.* **86**, 223–226 (2006).
12. Chen, Z.J. *et al.* Correlation between single nucleotide polymorphism of insulin receptor gene with polycystic ovary syndrome. *Zhonghua Fu Chan Ke Za Zhi* **39**, 582–585 (2004).

13. Villuendas, G., San Millán, J.L., Sancho, J. & Escobar-Morreale, H.F. The -597 G→A and -174 G→C polymorphisms in the promoter of the IL-6 gene are associated with hyperandrogenism. *J. Clin. Endocrinol. Metab.* **87**, 1134–1141 (2002).
14. Carlson, C.S., Eberle, M.A., Kruglyak, L. & Nickerson, D.A. Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446–452 (2004).
15. Xu, S. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* **85**, 762–774 (2009).
16. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
17. Huang, M. *et al.* Involvement of ALF in human spermatogenesis and male infertility. *Int. J. Mol. Med.* **17**, 599–604 (2006).
18. Latronico, A.C. *et al.* A homozygous microdeletion in helix 7 of the luteinizing hormone receptor associated with familial testicular and ovarian resistance is due to both decreased cell surface expression and impaired effector activation by the cell surface receptor. *Mol. Endocrinol.* **12**, 442–450 (1998).
19. Toledo, S.P.A. *et al.* An inactivating mutation of the luteinizing hormone receptor causes amenorrhea in a 46, XX female. *J. Clin. Endocrinol. Metab.* **81**, 3850–3854 (1996).
20. Latronico, A.C., Lins, T.S., Brito, V.N., Arnhold, I.J. & Mendonca, B.B. The effect of distinct activating mutations of the luteinizing hormone receptor gene on the pituitary-gonadal axis in both sexes. *Clin. Endocrinol.* **53**, 609–613 (2000).
21. Valkenburg, O. *et al.* Genetic polymorphisms of GnRH and gonadotrophic hormone receptors affect the phenotype of polycystic ovary syndrome. *Hum. Reprod.* **24**, 2014–2022 (2009).
22. Tong, Y., Liao, W.X., Roy, A.C. & Ng, S.C. Absence of mutations in the coding regions of follicle-stimulating hormone receptor gene in Singapore Chinese women with premature ovarian failure and polycystic ovary syndrome. *Horm. Metab. Res.* **33**, 221–226 (2001).
23. Du, J. *et al.* Two FSHR variants, haplotypes and meta-analysis in Chinese women with premature ovarian failure and polycystic ovary syndrome. *Mol. Genet. Metab.* **100**, 292–295 (2010).
24. Drieschner, N. *et al.* Evidence for a 3p25 breakpoint hot spot region in thyroid tumors of follicular origin. *Thyroid* **16**, 1091–1096 (2006).
25. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
26. Hu, C. *et al.* PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. *PLoS ONE* **4**, e7643 (2009).
27. Del Villar, K. & Miller, C.A. Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proc. Natl. Acad. Sci. USA* **101**, 4210–4215 (2004).
28. Olszanecka-Glinianowicz, M. *et al.* Is the polycystic ovary syndrome associated with chronic inflammation per se? *Eur. J. Obstet. Gynecol. Reprod. Biol.* **133**, 197–202 (2007).

ONLINE METHODS

Subjects. All Han Chinese participants in this study were recruited from women presenting in reproductive and gynecology clinics in several collaborating hospitals. In all subjects from whom peripheral blood samples were obtained, we recorded such anthropometric variables as age, body height, weight and menstrual cycle, as well as selected endocrine and biochemical parameters. All PCOS cases were diagnosed according to the Revised 2003 Consensus on Diagnostic Criteria and Long-term Health Risks Related to Polycystic Ovary Syndrome[1], requiring presence of any two of the following three criteria: oligo- and/or anovulation, clinical and/or biochemical signs of hyperandrogenism, and polycystic ovary morphology. In our initial discovery set, the 744 PCOS cases were diagnosed strictly and satisfied all three features. Transvaginal ultrasound was used to detect polycystic ovaries, defined as presence of at least one ovary >10 cm^3 or containing at least 12 follicles 2–9 mm in diameter. Ultrasound examination was performed rectally if subjects were virginal. Androgen excess was defined on the basis of hirsutism (Ferriman-Gallwey score ≥ 6) or elevated circulating total testosterone ≥ 60 ng dl^{-1} (refs. 29,30). BMI was calculated using the following formula: [weight (kg) / height^2 (m^2)]. Insulin resistance was defined using the HOMA-IR, applying the following formula: [plasma glucose (mmol l^{-1}) × insulin (μU ml^{-1})] / 22.5 (ref. 31). Other causes of oligomenorrhea or hyperandrogenism (for example, nonclassical 21-hydroxylase deficiency, Cushing's syndrome, hypothyroidism or elevated prolactin) were excluded by clinical evaluation.

Individuals in the GWAS discovery set of 744 Han Chinese PCOS cases and 895 controls were all recruited from Shandong Province, northern China. Replication 1 included 2,840 PCOS cases and 5,012 controls, all from individuals born in 12 provinces of northern China (Shandong, Heilongjiang, Jilin, Liaoning, Inner Mongolia, Hebei, Henan, Tianjin, Beijing, Shaanxi, Gansu and Ningxia). Replication 2 included 498 PCOS cases and 780 controls, all from individuals born in nine provinces of central and southern China (Jiangsu, Anhui, Shanghai, Guangdong, Guangxi, Fujian, Zhejiang, Hubei and Hunan)[15]. The control group was primarily recruited from healthy women, excepting 20% who were infertile women having tubal factor infertility. All controls underwent full clinical evaluations. Endocrine and biochemical parameters were measured to exclude hyperandrogenism, and ultrasound imaging was performed to exclude ovarian morphology indicative of PCOS.

Written informed consent was obtained from all subjects. Further demographic information was collected from both case and control subjects through a structured questionnaire. This study was approved by each participating center's Institutional Ethical Committee and was conducted according to Declaration of Helsinki principles.

DNA extraction. EDTA-anticoagulated venous blood samples were collected from all participants. Genomic DNA was extracted from peripheral blood lymphocytes by standard procedures using Flexi Gene DNA kits (Qiagen) and diluted to working concentrations of 50 ng μl^{-1} for genome-wide genotyping and 15–20 ng μl^{-1} for the validation study.

GWAS genotyping and quality control. Genome-wide genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Quality control filtering of the GWAS data was performed by excluding arrays with a Contrast QC <0.4 from further data analysis. Genotype data were generated using the birdseed algorithm[32]. For sample filtering, arrays with generated genotypes <95% of loci were excluded ($n = 32$). For SNP filtering (after sample filtering), SNPs with call rates <95% in either case or control samples were removed. SNPs in which the minor allele frequency (MAF) was either < 1% or deviated significantly from Hardy-Weinberg equilibrium ($P \leq 0.001$) in controls were also excluded. A total of 611,633 SNPs passed the quality criteria and were used for subsequent analysis.

SNP selection for validation. The following criterion was used for the selection of SNPs for validation: strong significant SNPs from the GWAS with an allele-based PCA-adjusted P-value ≤ 5×10^{-6}. SNPs with ambiguous genotype scatter plots were excluded from the subsequent stage.

Replication genotyping. Genotyping for the replication study was performed using the ligation detection reaction (LDR) method[33,34], with technical support from Shanghai Biowing Applied Biotechnology Company.

Statistical analysis. PCA was conducted using EIGENSTRAT software. The first two eigenvectors generated were selected for plotting, and an EIGENSTRAT procedure was used to generate adjusted statistics for GWAS. Genome-wide association analysis at the single-marker level and Hardy-Weinberg equilibrium analysis were performed using PLINK. In the replication study, allelic association analysis was conducted using SHEsis[35].

In the logistic regression analysis considering age and BMI, we used the following model to fit the data: $Y = b_0 + (b_1 \times \mathrm{ADD}) + (b_2 \times \mathrm{age}) + (b_3 \times \mathrm{BMI}) + e$, where Y represents the phenotype (1 for disease; 0 for normal), ADD represents the additive effects of allele dosage (minor allele) and e represents the random error effect.

Conditional logistic regression was used to test for independent effects of an individual SNP. The basic principle[36] is that when several SNPs clustered together are all significantly associated with a trait, two alternative explanations exist. First, LD between the alleles may account for the association of each SNP. In such a scenario, a logistic regression analysis conditioning on any one of the clustered SNPs will remove evidence of association for the other SNPs. Alternatively, the effects of the SNPs may be statistically independent, residing on distinct haplotypes that are independently inherited. In this case, conditioning on one SNP will not change the effect estimate of the other. In our study, for each SNP among the significantly associated SNPs within 2p21 and 9q33.3, we compared the original estimate to an adjusted estimate obtained by entering other SNPs as covariates. SNPs were added by forward selection, one by one based on significance, using the model $Y = b_0 + (b_a \times \mathrm{SNP_{add}}) + (b_1 \times \mathrm{SNP_1}) + \ldots + (b_n \times \mathrm{SNP}_n) + (b_s \times \mathrm{stage}) + e$, where $\mathrm{SNP_{add}}$ indicates the SNP being added for testing of its independence with $\mathrm{SNP_1}$ through SNP_n. If the estimated P-value of $\mathrm{SNP_{add}}$ was less than 0.05, we considered the effect of $\mathrm{SNP_{add}}$ to be independent from the effects of $\mathrm{SNP_1}$ through SNP_n.

Haploview was used for the genome-wide P-value plot (**Fig. 1**)[37].The Q-Q plot was created using the R qq.plot function[38]. The regional plots were generated using LocusZoom (see URLs). The GWAS and replication data were then combined using meta-analysis. The meta-analysis was conducted using the R 'meta' package. The heterogeneity across the three stages was evaluated using a Q-statistic P-value. The Mantel-Haenszel method was used to calculate the fixed effect estimate[39].

To compare the clinical phenotypes in the PCOS patients with different genotypes, one-way analysis of variance was used to analyze the key variants of BMI, testosterone and HOMA-IR. Calculations were performed using SPSS 16.0 (SPSS Inc.). Data were expressed as the mean ± s.d. Continuous variables were tested for distribution using a histogram in which abnormal values were excluded. The least significant difference (LSD) test was used for *post hoc* analysis. Appropriate transformations were applied (logarithmic, sine or square) as needed to ensure homogeneity of variance. For subgroups with a significant difference in age, covariance analysis was used to control for the age effect in a general linear model. The level of statistical significance was set at a $P < 0.05$ for all statistical analyses.

29. Ferriman, D. & Gallwey, J.D. Clinical assessment of body hair growth in women. *J. Clin. Endocrinol. Metab.* **21**, 1440–1447 (1961).
30. Shi, Y., Gao, X., Sun, X., Zhang, P. & Chen, Z. Clinical and metabolic characteristics of polycystic ovary syndrome without polycystic ovary: a pilot study on Chinese women. *Fertil. Steril.* **90**, 1139–1143 (2008).
31. Radziuk, J. Insulin sensitivity and its measurement: structural commonalities among the methods. *J. Clin. Endocrinol. Metab.* **85**, 4426–4433 (2000).
32. Korn, J.M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
33. Thomas, G. *et al.* Capillary and microelectrophoretic separations of ligase detection reaction products produced from low-abundant point mutations in genomic DNA. *Electrophoresis* **25**, 1668–1677 (2004).
34. Yi, P. *et al.* PCR/LDR/capillary electrophoresis for detection of single-nucleotide differences between fetal and maternal DNA in maternal plasma. *Prenat. Diagn.* **29**, 217–222 (2009).
35. Shi, Y.Y. & He, L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res.* **15**, 97–98 (2005).
36. Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466**, 113–117 (2010).
37. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263 (2005).
38. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
39. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).

doi:10.1038/ng.732