Genetic association studies in female reproduction: from candidate-gene approaches to genome-wide mapping

Han Zhao\textsuperscript{1,2,3,4} and Zi-Jiang Chen\textsuperscript{1,2,3,4,5,*}

\textsuperscript{1}Center for Reproductive Medicine, Shandong Provincial Hospital, Shandong University, Shandong, China \textsuperscript{2}National Research Center for Assisted Reproductive Technology and Reproductive Genetics, Jinan, Shandong, China \textsuperscript{3}The Key laboratory for Reproductive Endocrinology of Ministry of Education, Shandong, China \textsuperscript{4}Shandong Provincial Key Laboratory of Reproductive Medicine, Shandong, China \textsuperscript{5}Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

*Correspondence address. E-mail: chenzijiang@hotmail.com

Submitted on December 30, 2012; resubmitted on May 15, 2013; accepted on May 22, 2013

\textbf{ABSTRACT:} Many genetic association studies have been performed to investigate disorders of female reproduction, such as polycystic ovary syndrome, premature ovarian failure and endometriosis. These disorders typically manifest heterogeneously, and their pathogeneses are influenced by polygenic and environmental factors. Researchers evaluating these genetic associations have chosen candidate genes related to hormone action, steroid biosynthesis, inflammatory cytokines and autoimmune factors. Several of these genes have yielded statistically significant associations with female reproductive disorders; however, few associations have been robust and reproducible. Whole-genome association studies generate more reliable and unbiased results and represent a breakthrough in genetic studies of female reproduction. Nevertheless, to date only a very small fraction of the overall heritability has been identified and so further studies are needed.

\textbf{Keywords:} genetic association study / reproduction / genome-wide association study / polycystic ovary syndrome / premature ovary failure / endometriosis

\textbf{Introduction}

The diseases impairing female fecundity include polycystic ovary syndrome (PCOS), premature ovarian failure (POF) and endometriosis. These disorders affect millions of reproductive-aged women every year. Certain chronic complications of these disorders (e.g. obesity, type 2 diabetes (T2D), cardiovascular diseases and endometrial cancer) can jeopardize a woman’s health during her entire lifespan (Wild \textit{et al.}, 2010). A comprehensive understanding of the pathology and pathogenesis of these diseases is necessary to provide better health care. Genetic contributors to female reproductive disorders are well accepted, but it is unclear whether genetics are at the root of the etiology. In this review, we outline the definition and design of genetic association studies and discuss both candidate gene association studies and more recent genome-wide association studies (GWASs). Because well-documented reviews of candidate-gene association studies have been published previously, the current review will focus on GWASs of female reproduction-related traits and disorders.

\textbf{Genetic association studies}

Genetic association studies aim to identify candidate genes or genome regions that contribute to a specific trait or disease by identifying a correlation between disease status and genetic variation (Cordell and Clayton, 2005). This strategy is particularly applicable to complex and polygenic diseases. During the past decade, the genetic association study has emerged as a popular and powerful tool for deciphering the link between a genetic locus and a disease (Sagoo \textit{et al.}, 2009).

In genetic association studies, single-nucleotide polymorphisms (SNPs) are the most widely tested markers, and most studies use the case–control design. The detection of a higher frequency of an SNP allele or genotype in a series of individuals affected with a disease indicates an association with that disease (Mitjans and Arias, 2012) (Fig. 1). Initially genetic association studies employed candidate-gene strategies. To date, candidate genes have been confirmed for many different diseases and traits (Risch, 2000). However, candidate-gene studies have limitations. Most of these studies have evaluated small sample populations with limited statistical power, and few have yielded sufficiently robust results that could be replicated in different populations or by different investigators (Simoni \textit{et al.}, 2008).

With the emergence of high-throughput genotyping techniques comprising hundreds of thousands of SNPs on a single-chip, array-based GWASs provide a more comprehensive, unbiased, discovery-driven approach to learning the genetic basis of complex diseases. By taking advantage of linkage disequilibrium (LD) to capture most of genome’s variations, it became possible to apply a representative group of SNPs
GWASs in female reproduction

645

Figure 1 Allele frequencies comparison in genetic association study. By genotyping certain polymorphisms, allele frequencies are compared between cases and controls. A difference existing between the two groups indicates an association between the genetic marker and the disease.

(tag SNPs) to evaluate genome-wide variation. Because GWAS evaluate >10^6 individual SNP genotypes, the standard P value threshold (P < 0.05) will result in numerous false-positive associations. The genome-wide threshold of significance is often P < 10^{-8} (Zheng et al., 2009). By interrogating global SNP differences among individuals with a specific disease (or trait) and controls (or a large population), GWASs have successfully identified susceptibility loci for common diseases, such as T2D (Sladek et al., 2007) and schizophrenia (Shi et al., 2011); and traits, such as adult height (Weedon et al., 2008) and body mass (Wen et al., 2012). Since the first GWAS was reported in 2005 (Klein et al., 2005), 1533 studies have been added to the Catalog of Published GWASs (www.genome.gov/gwastudies/). GWAS have been recognized as a huge success in uncovering susceptibility genes underlying common diseases; however, we have to point out that GWAS have several limitations. The most fundamental criticism is that these diseases might be caused by some very infrequent small mutations or DNA aberrations and the genetic loci identified so far account for only a small fraction (in the order of 5%) of the overall heritability, which means other factors (though currently unknown) may play a more important role in pathogenesis of these common diseases.

Disorders involving female reproduction, such as PCOS, POF and endometriosis, always manifest as complex, polygenic traits (Jiang et al., 2009; Chen et al., 2011). Numerous genetic association studies to investigate female reproduction have focused on candidate genes related to hormone action, steroid biosynthesis, inflammatory cytokines and autoimmune factors. GWASs examining the genetics of female reproductive disorders initially progressed slowly compared with other complex diseases, such as T2D, in which 32 GWASs have been performed. In 2010, the first GWAS on endometriosis was released, and in early 2011, the first PCOS GWAS was published. Both of these studies represented breakthroughs in the genetic study of female reproduction.

Genetic association studies in PCOS

PCOS is the most common cause of female anovulatory infertility, with a prevalence of 6–8% among women of reproductive age (Carmina and Lobo, 1999; Ehrmann et al., 1999; Goodarzi and Azziz, 2006). PCOS was first described by Stein and Leventhal in 1935, but it remains controversial in its diagnosis and unclear in its etiology. Criteria from the National Institutes of Health (which has decided to adopt Rotterdam criteria for the diagnosis of PCOS), the Androgen Excess and PCOS Society and the Rotterdam consensus coexist (Rotterdam, 2004) and other criteria (e.g. Chinese criteria, Japanese criteria) are applied regionally. PCOS is a complex disorder with a highly prevalent, heterogeneous syndrome of clinical and/or biochemical androgen excess, ovulatory dysfunction and polycystic ovaries. Because of its characteristic familial clustering, PCOS is considered to have a genetic basis, and most genetic studies evaluate PCOS using candidate-gene association strategies (Goodarzi and Azziz, 2006).

Candidate-gene studies in PCOS

The candidate-gene approach to genetic association studies of PCOS is based on the biological plausibility of potential genes. Primary candidates include: (i) genes involved in the biosynthesis and activity of androgens; (ii) genes affecting insulin resistance and (iii) genes correlated with inflammatory cytokines. Because this topic has been well-documented previously, here we roughly depict these candidate genes in a functional or pathway-related order.

Hyperandrogenism is a fundamental characteristic and clinical manifestation of PCOS. For this reason, genes involved in steroidogenesis (e.g. CYP11A, CYP17A, CYP21, EPHX1, HSD17B6) have been regarded as essential candidates. CYP11A1 is a rate-limiting enzyme, and polymorphism D15520 in CYP11A1 is associated with PCOS susceptibility in Caucasian and Chinese populations (Pusalkar et al., 2009; Prazakova et al., 2010; Liu et al., 2011; Zhang et al., 2012). CYP21 mutations in women with PCOS also have been identified (Escobar-Morreale et al., 1999; Witchel and Aston, 2000). Sex hormones and hormone receptors play a fundamental role in folliculogenesis. Therefore, FSH, FSHR, LH, LHGR, AR and SHBG are regarded as compelling candidate genes for PCOS. The SNP rs6166 (p.N680S) in exon 10 of FSHR has been associated with PCOS in Dutch and Japanese women but not in Han Chinese population (Tong et al., 2001; Sudo et al., 2002; Valkenburg et al., 2009; Du et al., 2010). Inactivating mutations in LHGR could result in PCOS-like phenotypes (Toledo et al., 1996; Latronico et al., 1998). The CAG repeat polymorphism in the AR gene correlated with PCOS in an Australian study but was not considered a major determining factor in the development of PCOS via meta-analysis (Hickey et al., 2002; Wang et al., 2012). In a well-designed study conducted by Urbanek et al. (1999), ‘strong evidence’ was revealed for linkage between FOLLISTATIN and PCOS. Insulin resistance, a primary feature of PCOS, is considered an important component of PCOS pathogenesis. Genes such as INS, INSR, IRS-1, IRS-1 and IGF are plausible candidates (Siegel et al., 2002; Ertunc et al., 2005; Jin et al., 2006; Seow et al., 2007; Xu et al., 2009). A body mass-related gene, FTO, has interesting functions in cases of PCOS with obesity. A meta-analysis of eight studies showed that the FTO rs9939609 risk allele, could amplify the effect on body mass index (BMI) and weight in PCOS by more than 2-fold (Wojciechowski et al., 2012). Considering that FTO is also a
longstanding susceptibility gene of T2D, which shares similarities with PCOS, this gene deserves more attention to elucidate the delicate relationship between PCOS and obesity. Increasing evidence has indicated that chronic low-grade inflammation plays a role in the pathogenesis of PCOS. *In vitro* experiments have demonstrated that some proinflammatory cytokines could stimulate insulin resistance and enhance hyperandrogenism (González, 2011). To decipher this issue in the context of genetics, researchers have focused on TNF, TNFR, IL6, IL1R and other genes, but the results have been discordant (Deligeorgiou et al., 2009). The major challenges of these studies are the lack of replications to confirm the initial positive association or the publication of follow-up studies that present little or no evidence of an association.

**PCOS GWASs**

To date, nearly 100 genes have been listed as PCOS candidates; however, very few correlate consistently with PCOS. New approaches based on whole-genome analyses of PCOS show promise in the investigation of the genetic basis of this complex disorder. The first GWAS for PCOS (Chen et al., 2011) represented a 'major milestone' (Strauss et al., 2012) and identified three loci conferring risk to PCOS. This GWAS examined a Chinese population of 744 women with PCOS and 895 unaffected controls using an Affymetrix SNP 6.0 array. Three loci (2p16.3, 2p21 and 9q33.3) were associated with PCOS, and multiple SNPs were identified in the genes LHCGR, THADA and DENND1A (Chen et al., 2011). An association between PCOS and the Lhcg gene (2p16.3), a fundamental PCOS candidate, was re-confirmed in this GWAS. In addition, the thyroid adenoma-associated gene, THADA, mapping to chromosome 2p21, was first identified to associate with PCOS. THADA was initially identified as a target gene in benign thyroid tumors (Drieschner et al., 2006). A recent GWAS of T2D reported that THADA is an associated gene that might affect pancreatic beta-cell function (Zeggini et al., 2008). The PCOS GWAS also identified an association with DENND1A, which is located on chromosome 9q33.3 and encodes a DENN (differentially expressed in normal and neoplastic cells) domain. The DENND1A gene product functions as a guanine nucleotide exchange factor for the early endosomal small GTPase, Rab35 (Marat and McPherson, 2010).

Remarkably, these findings have been replicated in a series of subsequent studies. A transmission disequilibrium test examining Chinese family trios with PCOS confirmed the association and linkage of the THADA variants identified in the GWAS (Zhao et al., 2012). Goodarzi et al. (2012) (1474 cases versus 1802 controls) and Welt et al. (2012) (1131 cases versus 17,607 controls) confirmed the association between DENND1A and PCOS in USA and Icelandic European populations, respectively (Goodarzi et al., 2012; Welt et al., 2012). At the same time, we must point out that not all subsequent studies replicated all of the initial GWAS identified SNPs. Due to the differences in the LD and minor allele frequencies of SNPs between European and Chinese populations, SNPs associated with Chinese individuals (THADA: rs13429458 and rs12478601; LHCGR: rs13405172) were not well replicated in Europeans. However, Goodarzi et al. (2012) reported that the rs12463934 SNP in THADA was associated with PCOS risk. Particularly, by genotyping 94 tag SNPs Mutharasan et al. (2013) found evidence for 2p16.3 (LHCGR and FSHR) as a PCOS susceptibility locus regardless of ethnicity. The informative SNPs were not the same as those identified in the Chinese population, thus highlighting the potential bias when comparing different ethnic groups. For studies that cannot replicate SNP data reported in GWASs, it is necessary to consider the LD pattern between the different populations.

Considering that PCOS is a complicated endocrine–metabolic disorder, the three loci found in the first GWAS (PCOS GWAS-I) were far from sufficient to explain such a complex disorder. Also, the first GWAS in PCOS focused only on the full blown phenotype, i.e. the women with ovulatory dysfunction, hyperandrogenism and polycystic ovarian morphology. This constitutes a subgroup of the whole picture and therefore this group is more homogeneous. Therefore, a second GWAS (PCOS GWAS-II) including all phenotypes was conducted again in a Chinese population using an enlarged sample size of 10,480 PCOS cases and 10,489 controls (Shi et al., 2012). In PCOS GWAS-II, an additional eight loci were identified, including SNPs in the plausible genes FSHR and INS. In total, 11 loci related to PCOS have been uncovered by GWASs (Table I). Of these loci, the following were within the susceptibility loci related to hormone actions and organ growth and/or were in common with T2D: FSHR and LHCGR are plausible candidates related to folliculogenesis and ovulation; INS, HMGAA2 and THADA were identified as candidate genes for T2D in recent GWASs (Zeggini et al., 2008; Voight et al., 2010); 12q13.2 also increases the risk of type 1 diabetes (Burton et al., 2007; Todd et al., 2007; Hakonarson et al., 2008; Barrett et al., 2009; Cooper et al., 2009; Plagnol et al., 2011); and YAP and HMGAA2 (‘pygmy’ gene) may correlate with the enlarged size of the polycystic ovary. These genes function in different classes according to pathway analysis: INS and ERBB3 are implicated in the processes of female gamete generation; LHCGR and FSHR both encode hormone ligand-binding receptors; FBPI and INS are related through the insulin signaling pathway; YAP1 and ERBB3 are involved in the ERBB signaling pathway (Joshi-Tope et al., 2005; Matthews et al., 2009); LHCGR and ERBB3 are involved in a calcium signaling pathway (Ogata et al., 1999) and RASB3 and ERBB3 play roles in endocytosis. Recently, a third GWAS on PCOS was conducted in a Korean population, but it identified the gene GYS2 that is specifically associated with BMI rather than PCOS per se (Hwang et al., 2012).

Susceptibility loci provide clues to PCOS etiology. Deciphering the functional mechanisms and determining the clinical relevance of these findings will perhaps be our greatest challenge in the ‘post-GWAS’ era. Researches to explore quantitative traits or phenotypes are ongoing, likely facilitating the diagnosis and classification of PCOS. A recent genotype–phenotype correlation study demonstrated that THADA and DENND1A carry risk alleles that are associated with endocrine and metabolic disturbances in PCOS patients of Han Chinese descent (Cui et al., 2013). Though vast progress has been achieved in identifying loci associated with PCOS, only a small portion of familial clustering has been explained. So far, efforts to identify PCOS susceptibility loci have focused on Asian populations; GWAS of other ethnic groups are required.

**Genetic association studies in POF**

POF is defined as premature cessation of menstruation with elevated levels of serum gonadotrophins before the age of 40 years (Coulam, 1982). POF is described by a series of disorders, including hypoestrogenemia, high FSH concentrations and early menopause, which lead to
infertility. The disease is often accompanied by other manifestations, such as hot flashes, sleep disturbances, mood symptoms, vaginal dryness and osteoporosis or decreased bone density (Taffe and Dennerstein, 2002; Miro et al., 2004). Due to the long and variable clinical course of ovarian dysfunction, the revised term ‘primary ovarian insufficiency’, has been proposed as a more scientifically accurate definition for POF (Welt, 2008; Nelson, 2009).

Candidate-gene studies in POF

Non-genetic contributors to POF include gonadotoxic chemotherapy, radiation treatment and autoimmune oophoritis, but most cases are idiopathic and genetic. Chromosomal abnormalities involving the X chromosome or autosomes account for 12–14% of cases (Devi and Benn, 1999; Janse et al., 2010; Jiao et al., 2011). Substantial causal relationships with POF have been established for seven genes (POF1-POF7) on OMIM (Online Mendelian Inheritance in Man, http://omim.org/). Premutation of the gene FMR1 (fragile X mental retardation 1, Xq27.3) is responsible for 3.3–6.7% of sporadic POF and 13% of familial cases (Conway et al., 2002; Bodega et al., 2004; Gersak et al., 2004; Persani et al., 2010). Harris et al. (2002) screened FOXL2 among 70 women with POF from Slovenia and New Zealand and identified the following two novel heterozygous mutations: (i) a 30-bp deletion that removed 10 of 14 alanes in a Slovenian patient and (ii) the nonsense mutation p.Y258N (tyrosine to asparagine) in a patient from New Zealand (Harris et al., 2002). However, few other studies have found FOXL2 mutations in POF cases without BPES (De Baere et al., 2001; Bodega et al., 2004; Gersak et al., 2004; Laissue et al., 2009).

Both BMP15 and GDF9 belong to the TGF-β superfamily, are expressed specifically in germ cells, and play crucial roles in folliculogenesis. Among populations of POF cases, mutations in BMP15 and GDF9 have variable prevalences of 1.5–12% (Di Pasquale et al., 2004; Laissue et al., 2006; Kovanci et al., 2007; Zhao et al., 2007; Lakhal et al., 2009). NOBOX and FIGLA are oocyte-specific transcription factors, and deletion of either of these genes could accelerate post-natal oocyte loss. Mutations in NOBOX seem to occur more frequently in the Caucasian POF population. The NOBOX missense variant, p.R355H, first identified in 1 of 96 Caucasian POF subjects, could disrupt the binding of the NOBOX homeodomain to DNA (Qin et al., 2007). Bouilly et al. (2011) subsequently demonstrated that loss-of-function NOBOX mutations accounted for 6.2% of POF cases in a Caucasian cohort of 178 participants (Bouilly et al., 2011). However, two other Asiatic POF studies

<table>
<thead>
<tr>
<th>References</th>
<th>Ethnicity</th>
<th>Sample size (cases; control)</th>
<th>Replication sample size</th>
<th>Region</th>
<th>Mapped genes</th>
<th>SNP</th>
<th>P</th>
<th>OR</th>
<th>Replicated studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. (2011)</td>
<td>Han Chinese</td>
<td>744; 895</td>
<td>I: 2840; 5012 II: 498; 780</td>
<td>2p21</td>
<td>THADA</td>
<td>rs13429458</td>
<td>$2 \times 10^{-23}$</td>
<td>1.49</td>
<td>Goodarzi et al. (2012), Zhao et al. (2012), Shi et al. (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs12478601</td>
<td>$3 \times 10^{-23}$</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>THADA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LHCGR; STON1-GTF2A1L</td>
<td>rs13405728</td>
<td>$8 \times 10^{-21}$</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DENND1A</td>
<td>rs2479106</td>
<td>$8 \times 10^{-19}$</td>
<td>1.34</td>
<td>Mutharasan et al. (2013), Shi et al. (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs10818854</td>
<td>$9 \times 10^{-18}$</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>Shi et al. (2012)</td>
<td>Han Chinese</td>
<td>2254; 3001</td>
<td>8226; 7578</td>
<td>12q13.2</td>
<td>SUOX</td>
<td>rs705702</td>
<td>$9 \times 10^{-26}$</td>
<td>1.27</td>
<td>Li et al. (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs1894116</td>
<td>$1 \times 10^{-22}$</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs2277046</td>
<td>$2 \times 10^{-21}$</td>
<td>1.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs3802457</td>
<td>$5 \times 10^{-14}$</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs2268361</td>
<td>$1 \times 10^{-12}$</td>
<td>1.15</td>
<td>Mutharasan et al. (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2p16.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hwang et al. (2012)</td>
<td>Korean</td>
<td>774; 967</td>
<td>482 children, 468 gestational diabetes mellitus women; 1242 controls</td>
<td></td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GWASs in female reproduction
failed to find causative mutations (Zhao et al., 2005; Qin et al., 2009), implying that mutations in NOBOX may be different among ethnicities. The FIGLA gene has been evaluated for POF associations in a Chinese population (Zhao et al., 2008). These researchers identified the following two causative deletions: (i) c. 15–36 del (p.G66X66), resulting in a frameshift that leads to haploinsufficiency and (ii) c. 419–421 delAAG (p.140 delN) that disrupts FIGLA binding to the TCF3 helix-loop-helix domain. NRSA1 (also named steroidogenic factor 1, SF1) is a key transcriptional regulator of genes involved in the hypothalamic–pituitary–steroidogenic axis (Luo et al., 1994). Several mutations in NRSA1 were detected in members of four families with both 46,XY DSD (disorders of sex development) and 46,XX POF females, as well as in 2 of 25 isolated POF women, but in none of 700 controls (Lourenco et al., 2009).

Besides these causal POF genes, dozens of other candidate genes involved in X chromosomes, hormonal actions and oogenesis expression (e.g., FSHR, LHCGR, CYP19, PTEN and SOX11) have been evaluated in the context of POF. However, each single gene seems to be responsible for less than perhaps 1–6% of POF. In aggregate, only a small portion of POF cases can be explained genetically. And most of the causal genes were chosen as candidates because they are acknowledged to be associated with folliculogenesis or other known-related biological pathways. Other genes which are not so obviously related to POF and the genes without any functional annotation were neglected. Hence, the unbiased, discovery-driven approach-GWAS on POF emerged.

POF GWAS and array-comparative genomic hybridization

Several GWASs have identified various loci potentially linked to POF. Qin et al. (2012) conducted the largest-scale GWAS, investigating 791 POF patients and 1695 controls. These researchers found a plausible association locus at 8q22.3 (rs3847153, \( P = 8.85 \times 10^{-8} \)), but it harbored no candidate genes. Other studies have involved small numbers of cases (Knauff et al., 2009; Pyun et al., 2012), thereby limiting the statistical power to detect a reliable association, and few have been replicated (Table II). Several array-comparative genomic hybridization (array-CGH) analyses have also been performed on POF. Aboura et al. (2009) first reported significantly associated copy number variations (CNVs) comprising genes DNAH5, NAPA, DUSP22, NUPR1 and AKT1, which are known reproductive candidate genes. An array-CGH study by McGuire et al. (2011) (89 POF cases) identified 17 novel microduplications and 7 novel microdeletions (McGuire et al., 2011). These authors emphasized that most of the novel CNVs were derived from autosomes rather than the X chromosome.

Genetic association studies in endometriosis

Endometriosis is an estrogen-dependent gynecologic disorder affecting 5–10% of women of reproductive age (Giudice and Kao, 2004) and 20–50% of women with infertility (Gao et al., 2006). Endometriosis is defined as the growth of endometrial glands and stroma outside the uterine cavity that often results in dyspareunia, dysmenorrhea, pelvic pain and infertility. Immunologic disorders, genetic predispositions, hormonal alterations and environmental factors have been implicated in endometriosis disease risk (Olive and Schwartz, 1993; Brinton et al., 1997). Familial aggregation has been shown in several studies (Simpson et al., 1980; Lamb et al., 1986; Moen and Magnus, 1993) and recurrence risks increase 5–7% in the first-degree relatives; twin studies also showed an increased concordance of endometriosis in monozygotic twins compared with dizygotic twins and the general population (Hadfield et al., 1997). A large study of 3096 Australian female twins suggested that ∼51% of the variation in endometriosis risk was genetic (Trelfa et al., 1999).

Candidate-gene studies in endometriosis

Similar to other complex diseases, endometriosis may be influenced by multiple genes. Numerous candidate genes have been investigated, but no strong candidates have emerged (Rahmioglu et al., 2012). These genes can be generally categorized into several groups: (i) steroids and hormone receptors; (ii) cytokines/inflammation factors; (iii) adhesion molecules and matrix enzymes and (iv) apoptosis- and oncogenesis-related genes. The majority of polymorphisms investigated thus far are not associated with endometriosis, either because of small sample sizes and inconsistent reproducibility, or because of ethnic limitations. For example, null deletions in GSTM1 (encoding phase II detoxification enzyme in the estrogen gene) were related to an increased risk of endometriosis in French, Russian, Indian, Chinese and Taiwanese women (Baranova et al., 1997; Baranova et al., 1999; Ivashchenko et al., 2003; Peng et al., 2003; Hsieh et al., 2004; Babu et al., 2005); however, no links were found in Korean, Japanese or Australian cases of endometriosis (Baxter et al., 2001; Morizane et al., 2004; Hur et al., 2005). Previous candidate-gene association studies have provided less robust information to explain the heritability of endometriosis.

Endometriosis GWAS

Uno et al. (2010) reported the first GWAS on endometriosis, which evaluated 1907 Japanese individuals with endometriosis and 5292 controls (Uno et al., 2010). This study identified a significant association between endometriosis and rs10965235 [\( P = 5.57 \times 10^{-12} \), odds ratio (OR) = 1.44], which is located in the CDKN2BAS (cyclin-dependent kinase inhibitor 2B antisense RNA) gene on chromosome 9p21. These researchers also found an approximate susceptibility locus on chromosome 1p36 containing WNT4 (\( P = 1.66 \times 10^{-10} \), OR = 1.20). A subsequent GWAS (Adachi et al., 2010) on endometriosis was also performed in a Japanese population, but this study was smaller (696 endometriosis patients and 825 controls), and meta-analysis did not detect credible associated loci. A third GWAS (Painter et al., 2011) examined 3194 individuals with endometriosis and 7060 controls from Australia and the UK. This study identified an associated signal (rs12700667) on chromosome 7p15.2 in an intergenic region near the genes NFE2L3 and HOXA10. The authors replicated the finding in an independent cohort from the USA of 2392 endometriosis cases and 2271 controls. A meta-analysis of the Japanese, Australian and UK GWAS data sets (Nyholt et al., 2012) re-confirmed rs12700667 and rs7521902, which had been identified in previous GWASs. These researchers also detected five new signals (rs13394619, rs10859871, rs4141819, rs7739264 and rs1537377) indicating associations with endometriosis risk. Most of these signals were associated with consistent risks for endometriosis across European and Japanese populations. The latest GWAS including 2019 surgically confirmed endometriosis cases and 14 471 controls identified three SNPs achieved genome-wide significance (rs2235529, rs1519761 and rs6757804). Notably, rs2235529 resides near WNT4, which is
Genetic association studies in menarche and menopause traits

There is a strong evidence for a downward secular trend in age at menarche (Euling et al., 2008). Early menarche is known to be associated with increased risks of breast, ovarian, and endometrial cancers; hypertension; T2D and cardiovascular disease (Golub et al., 2008). GWASs have identified more than 40 associated genetic loci for menarche and 50 for menopause (He et al., 2009; Liu et al., 2009; Ong et al., 2009; Elks et al., 2010b; Chen et al., 2012; Hai et al., 2012; Stolk et al., 2012; Perry et al., 2013) (Supplementary data, Table S1). Most importantly, common variants in LIN28B have been associated with age at menarche in several independent GWASs, although the identified variants in the LIN28B regions differed between studies. LIN28B also influences the onset of breast development in girls, the timing of pubic hair development and voice breaking in boys, and the timing of the pubertal growth spurts in boys and girls (Widén et al., 2010). LIN28B is the human homolog of lin-28 in Caenorhabditis elegans, which controls the nematode’s rate of progression from larval stages to adult cuticle formation. Thus, conserved mechanisms may be involved in human developmental timing (Guo et al., 2006). In humans, LIN28B may regulate growth and timing of menarche by sequence pre-processing of the let-7 family of microRNAs, which are responsible for cell pluripotency and cancer growth (Viswanathan et al., 2008). In vitro experiments and transgenic mouse models also support the relationships between LIN28B and breast, ovarian, and endometrial cancers; hypertension; T2D and cardiovascular disease (Golub et al., 2008). GWASs have previously suggested to be associated with endometriosis (Albertsen et al., 2013). Replication studies are few that only two have published so far. Using 305 endometriosis patients, Pagliardini et al. (2013) (Supplementary data, Table S1). Most importantly, common variants in LIN28B have been associated with age at menarche in several independent GWASs, although the identified variants in the LIN28B regions differed between studies. LIN28B also influences the onset of breast development in girls, the timing of pubic hair development and voice breaking in boys, and the timing of the pubertal growth spurts in boys and girls (Widén et al., 2010). LIN28B is the human homolog of lin-28 in Caenorhabditis elegans, which controls the nematode’s rate of progression from larval stages to adult cuticle formation. Thus, conserved mechanisms may be involved in human developmental timing (Guo et al., 2006). In humans, LIN28B may regulate growth and timing of menarche by sequence pre-processing of the let-7 family of microRNAs, which are responsible for cell pluripotency and cancer growth (Viswanathan et al., 2008). In vitro experiments and transgenic mouse models also support the relationships between LIN28B and breast, ovarian, and endometrial cancers; hypertension; T2D and cardiovascular disease (Golub et al., 2008). GWASs have previously suggested to be associated with endometriosis (Albertsen et al., 2013). Replication studies are few that only two have published so far. Using 305 endometriosis patients, Pagliardini et al. (2013) have recently performed a replication study to identify the four previously identified SNPs, rs12700667, rs7798431, rs1250248 and rs7521902. However, except a borderline association observed for rs1250248 and endometriosis ($P = 0.049$), they failed to replicate the other three GWAS-identified endometriosis-associated SNPs (Sundqvist et al., 2013) (Table III). Apparently, the few validated loci are inadequate to explain heritability of endometriosis, and more work awaits to be done.
identified in many of the studies resides in an intergenic region at 9q31.2 (Perry et al., 2009; Elks et al., 2010a).

Concluding remarks

Female reproductive disorders are typically heterogeneous and complex and originate from genetic and environmental factors. Candidate gene association studies have limitations such as small sample sizes and limited statistical power, and few positive results have been reproduced using this strategy. GWASs have changed the landscape of genetic research of PCOS and endometriosis and provide exciting insights into the genetic architecture of PCOS. GWASs of POF based on relatively small sample sizes can identify plausible loci, whereas candidate-gene studies and array-CGH remain valuable strategies to unravel causal variants for POF.

Future association studies need global collaboration to expand sample sizes, identify more susceptibility loci and ultimately to discover missing heritabilities. Based on these results, disease subtypes must be specified so that subtype-related markers can be explored. Besides an understanding of the genetic mechanisms, association studies are the key to establishing risk prediction models, developing less invasive methods of diagnosis and creating more effective and targeted treatments.

Supplementary data

Supplementary data are available at http://molehr.oxfordjournals.org/.

Acknowledgements

We would like to thank Drs Margrit Urbanek, Chris Barratt and Helen Beard for structural and editorial critiques of the manuscript while in preparation.

Authors’ roles

H.Z. drafted the article and revised it critically for important intellectual content. Z-J. C. conceived the study, participated in its design and coordination, helped revising the manuscript and supervised the study. All authors read and approved the final manuscript to be published.

Funding

Research carried out in the authors’ laboratory was supported by the National Basic Research Program of China (973 program) (2012CB944700, 2011CB944502), the National Natural Science Foundation of China (81000238) and the Scientific Research Foundation of Shandong Province of Outstanding Young Scientist (2012BSED27089). Science research foundation item of no-earnings health vocation (201002013) and National Key Technology Research and Development Program (2011BAI11B00).

Conflict of interest

None declared.

References


GWASs in female reproduction


Perry JR, Corre T, Esko T, Chasman DI, Fischer K, Franceschini N, He C, Kutanli Z, Mangino M, Rose LM et al. A genome-wide association study...


