1. ABSTRACT

Family and twin studies have shown that heritability accounts for endometriosis development to an extent similar to other complex genetic diseases. Both linkage analysis and association studies have been performed to identify genetic determinants for the disease. Results from the linkage scan of 1,176 families collected thanks to a joint effort between an Australian and a UK group supported significant linkage to a novel susceptibility locus on chromosome 10q26. Although gene variants with effects on the disease predisposition have been proposed to exist and several candidates have been put forward, their effects have not been or are yet to be confirmed. The main categories of candidate genes studied have been those involved in detoxification processes, sex steroid biosynthesis and action, immune system regulation. Genetic studies on endometriosis face numerous challenges as the disease has several manifestations and different forms. Moreover, strong gene-environmental interactions might definitively influence approaches to identify genetic variants involved. Genome-wide association studies that survey most of the genome for causal genetic variants provide the potential for future progress.

2. INTRODUCTION

Advances in genetics and molecular biology, such as the completion of the human genome sequence (1, 2), the deposition of millions of single nucleotide polymorphisms (SNPs) into public databases (3) and the rapid improvements in SNP genotyping technology, have accompanied the development of studies in which SNPs across the genome are genotyped to survey the most common genetic variation for a role in disease or to identify the heritable quantitative traits that are risk factors for disease. Many common diseases do actually cluster in families in patterns that demonstrate that genetics plays a role in determining susceptibility. For example, the identical twin of a patient with type 1 diabetes will also get type 1 diabetes 30–50% of the time; dizygotic twins who share a common environment but only 50% of their genes are less concordant (4, 5). A sibling of a patient affected by type 1 diabetes has a 15 times higher risk to get diabetes than an unrelated individual (6).

The increased risk to relatives (\(\lambda\)) is one measure of the influence of genetics; another measure of the contribution of inherited factors is termed heritability (\(h^2\)).
Genetics of endometriosis: current status and prospects

which indicates the fraction of the population variation that can be explained by genetic factors working together in an additive fashion. Heritability can be evaluated either from family studies or twin studies, and is often between 30 and 50% for common diseases such as diabetes (4), or quantitative traits such as body mass index or blood pressure (7, 8). Thus, multiple genetic factors and nongenetic factors combine to influence the risk of common diseases and quantitative traits. Because multiple genetic and nongenetic factors interact to affect phenotype, these diseases and traits are termed complex genetic traits in contrast to phenotypes that are controlled by single genes called monogenic or Mendelian traits. The propensity of genetic background to modify the phenotypic expression of most if not all Mendelian traits suggests that few if any traits are truly monogenic and that instead most are genetically complex (5). Genes that contribute to complex traits pose special challenges that make gene discovery more difficult, including locus heterogeneity, epistasis, low penetrance, variable expressivity and pleiotropy, and limited statistical power.

Endometriosis, a common, benign, oestrogen-dependent, chronic gynaecological disorder associated with pelvic pain and infertility has a prevalence that approaches 1% for women in the general population (11). There is also evidence that endometriosis clusters in some families. In 1980, Simpson and coworkers (13) reported for the first time that women with an affected sibling or parent were more likely to have the disease and a more severe form. Female siblings of probands with endometriosis had a 5.9% incidence of the disease while only 1% of the patients’ husband’s first-degree relatives were affected. Similar findings were observed in studies conducted in UK and Norway (14, 15). Subsequently, evidence of the genetic basis of endometriosis has emerged from the analysis of large, clinical databases in Australia, Iceland, and Utah. Stefansson et al. have compared 750 Icelandic women with endometriosis to matched controls (16). The former were descendent from a smaller number of ancestors with risk ratios for sisters and cousins of 5.20 and 1.56, respectively. The mean kinship coefficient (KC) for the affected women was significantly higher than that for 1000 sets of matched controls, and this remained significant even when the contribution from first-degree relatives was excluded. Higher recurrent risks have been reported among Mormons in the US state of Utah (17). Of 419 women with endometriosis, 326 had at least one sister affected and 11.2% of the probands’ mothers has a surgical diagnosis. Affected relatives were more often of maternal (10.3%) than paternal (5.6%) lineage (17). However, such studies can only provide a suggestion for the genetic influence because clustering can be due to both genetic factors and a shared family environment.

3. GENETIC EPIDEMIOLOGY OF ENDOMETRIOSIS

Twin studies have found that approximately 51% of the variance of the latent liability to the disease may be attributable to genetic influences. Utilizing the Australian National Health and Medical Research Council Twin Register, Treloar et al. studied 3298 monozygotic (MZ) and dizygotic (DZ) twin pairs and of the twins surveyed, 215 self-reported that they were affected giving a prevalence rate of 0.07 among responders (12). Higher correlations were found for women with surgically confirmed disease which provides support for the search for genes that affect endometriosis in humans.

There is also evidence that endometriosis clusters in some families. In 1980, Simpson and coworkers (13) reported for the first time that women with an affected sibling or parent were more likely to have the disease and a more severe form. Female siblings of probands with endometriosis had a 5.9% incidence of the disease while only 1% of the patients’ husband’s first-degree relatives were affected. Similar findings were observed in studies conducted in UK and Norway (14, 15). Subsequently, evidence of the genetic basis of endometriosis has emerged from the analysis of large, clinical databases in Australia, Iceland, and Utah. Stefansson et al. have compared 750 Icelandic women with endometriosis to matched controls (16). The former were descendent from a smaller number of ancestors with risk ratios for sisters and cousins of 5.20 and 1.56, respectively. The mean kinship coefficient (KC) for the affected women was significantly higher than that for 1000 sets of matched controls, and this remained significant even when the contribution from first-degree relatives was excluded. Higher recurrent risks have been reported among Mormons in the US state of Utah (17). Of 419 women with endometriosis, 326 had at least one sister affected and 11.2% of the probands’ mothers has a surgical diagnosis. Affected relatives were more often of maternal (10.3%) than paternal (5.6%) lineage (17). However, such studies can only provide a suggestion for the genetic influence because clustering can be due to both genetic factors and a shared family environment.

4. GENETIC STUDY DESIGNS IN IDENTIFYING VARIANTS UNDERLYING COMPLEX TRAITS AND COMMON DISEASES

In Mendelian traits and diseases, the first step in gene discovery involves mapping the gene precisely and unambiguously to a small genetic interval. Given the strong relation between genotype and phenotype, single recombinants are sufficient to define minimal intervals of less than 1 cM. Consequently, the gene involved in the affected individuals is usually identified through the analysis of the coding sequence variants in a small number of candidate genes. These concepts do not apply to genetically complex traits. For most common diseases and complex traits, the underlying genetic variations remains unknown (5). In theory, the relevant genetic variation could be rare with allele frequencies under 1% in the population, as is true for most single-gene disorders, or more common with allele frequencies above 1% in the population. The frequency of the alleles for complex traits is important to understand, because it will guide approaches to find the causal genetic variants. Theoretical and empirical considerations suggest that for common diseases and complex traits, some of the causal genetic variants may be common (18-20). There are at least three arguments in favour of a role for common variation in complex traits (5):

- by definition, causal alleles for monogenic disorders are highly penetrant and often lead to severe phenotypes. Accordingly, these alleles often cause severe changes in protein function, and the spectrum of disease alleles usually includes not only missense mutations but also nonsense mutations, severe splicing mutations and insertion or
Genetics of endometriosis: current status and prospects

deletion mutations, which can induce frameshifts. Clearly, these mutations are often subject to negative selection. By contrast, the alleles that underlie complex traits have more subtle effects on disease risk and might be more likely to include non-coding regulatory variants with a modest impact on expression. In addition, given the modest effects of these alleles on disease risk and the late-onset of many common diseases, the causal alleles are far less likely to be subject to strong negative selection and might therefore comprise different types of variants to those that underlie Mendelian disorders. Thus, for complex traits, the impact of selective pressure is diluted for the variants;

• most single-gene diseases are rare, whereas most polygenic diseases are common; based on the demographic history of the human population, it can be predicted that for common diseases and quantitative traits, some of the causal genetic variation should have a high frequency in the population (20);
• empirical evidence suggests that common variants do contribute to the risk of common diseases (21).

These concepts are valid also for endometriosis and in trying to identify genes implicated in the etiology, pathophysiology, and progression of the disease, similarly to what has been done for other pathologies, two major approaches have been used to map genetic variants that influence the disease risk: linkage analysis and the candidate gene approach.

4.1. Linkage analysis

Linkage analysis is the method traditionally used to identify disease genes, and has been very successful for mapping genes that underlie monogenic 'Mendelian' diseases (2). For linkage analysis to succeed, markers that flank the disease gene must segregate with the disease in families. Variants that cause monogenic disorders are often rare so each segregating disease allele will be found in the same 10-20 cM chromosomal background within each family. Furthermore, because Mendelian diseases are caused by highly penetrant variants, markers within 10-20 cM of the disease-causing alleles will co-segregate with disease status. This approach has also been used for many common diseases and quantitative traits and in specific cases, genomic regions that show significant linkage to the disease have been identified, leading to the discovery of gene variants contributing susceptibility to several complex diseases (22-25). However, for most common diseases, the genes discovered usually explain only a small fraction of the overall hereditability of the disease. Various factors may explain this incomplete success (2):

• the inability of the standard set of microsatellite markers to extract complete information;
• limit heritability of most complex traits;
• the inaccurate definition of the disease phenotypes;
• and limited powered study designs.

However, even if statistically significant evidence of linkage is obtained, extensive candidate gene studies are still required to progress from a broad region of linkage — usually exceeding 10 cM (~10 million bases) — to the causal gene or genes within this region (2). So, results from small-scale sib-pair investigations should be interpreted with caution.

4.2. Candidate gene approach

Candidate-gene studies have been the only practical alternative to linkage analysis. In these hypothesis-based studies, genes are selected for further study, either by their location in a region of linkage, or on the basis of other evidence that they might affect disease risk (2). Association studies using common allelic variants are cheaper and simpler than the complete resequencing of candidate genes, and have been proposed as a tool for identifying the common variants that underlie complex traits (2, 5, 26). These studies compare the frequency of alleles or genotypes of a particular variant between disease cases and controls. In any case, candidate-gene studies rely on having predicted the identity of the correct gene or genes, usually on the basis of biological hypotheses or the location of the candidate within a previously determined region of linkage. Even if these hypotheses are broad, they will, at best, identify only a fraction of genetic risk factors, even for diseases in which the pathophysiology is relatively well understood. When the fundamental physiological defects of a disease are unknown, the candidate-gene approach will clearly be inadequate to fully explain the genetic basis of the disease. Moreover, they are often misinterpreted and therefore appear to be poorly reproducible. Based on the results of 55 meta-analyses of genetic associations, a recent report has established that only 16% of them were subsequently replicated with formal statistical significance (26). A different magnitude of effect is often seen in large versus smaller studies or differences in first versus subsequent results. Typically, large studies suggest weak associations or no associations at all while smaller studies or first research usually propose strong associations. Thus, a common phenomenon in association studies is the “winner’s curse” in which the first report overestimates the genetic effect size (26). The possible causes of this inconsistency are false-positive reports, false-negative studies that incorrectly fail to replicate a valid association, or true heterogeneity between studies. Assessment of the strength of heterogeneity or bias is mandatory in these studies. Genuine heterogeneity could be due to variation in frequency of alleles, variable effects of linkage disequilibrium for other genetic markers, variable disease expression or differential disease susceptibility across the studied populations. In the presence of large biological and environmental variability, genetic effects can differ in different populations or even among generations within a population. Thus, in case of genuine genetic heterogeneity, since most associations refer to small odds ratios, single studies with hundreds participants are greatly underpowered. Several thousands of patients are necessary to address these genetic risk factors and more than 10000 individuals need to be studied for adequate powered analysis. Among sources of biases, publication bias could be prominent. Since investigators, reviewers and editors tend to submit or accept manuscripts for publication based in the direction of strength of the study findings results of smaller and negative studies are difficult to locate. Misclassification bias from errors in case-control
assignment or genotyping errors should also be considered. Confounding by ethnicity or other factors can also cause bias. Finally, claims of significance should be interpreted cautiously since results of statistical test might indicate chance alone (26).

5. WHAT HAS BEEN FOUND SO FAR IN ENDOMETRIOSIS?

5.1. Challenges to genetic studies in endometriosis

In considering to perform a genetic study on endometriosis, some aspects of the disease need to be beard in mind:

- Endometriosis is a complex pathology with multiple manifestations and different forms (27). According to some experts, different histogeneses are responsible for the different forms of the disease and more than one disease entity exists. Indeed, while most authors agree that the genesis of peritoneal lesions has to be attributed to the implantation of endometrial tissue regurgitated through the fallopian tubes during menstruation, endometriotic ovarian cysts and the endometriosis of rectovaginal septum might have different pathogenesis. If this is the case, the phenotypic definitions should be extremely precise to limit genetic heterogeneity.

- Selection of proper controls is particularly critical. For these kind of studies, it would be advantageous to find a control group that is representative of the source population of cases in order to eliminate potential confounding factors such as referral or health care seeking pattern and socioeconomic status (28). Controls should have had the same opportunity to develop the disease of interest, and they would have had the same opportunity as cases to have been included in the study. Unfortunately, definite diagnosis of endometriosis relies on surgery and this may introduce important selection biases for both cases and controls.

- Epidemiological analyses have identified numerous factors that are thought to be associated with endometriosis such as age, smoking, reproductive characteristics, body mass index and dioxin exposure (10). The possibility of strong gene-environmental interactions should definitively influence the approaches that are used to identify genetic variants involved and the study size.

5.2. The International Endogene Linkage Study

Two independent groups, the Australian Genes Behind Endometriosis Study and the United Kingdom–based, International Oxford Endometriosis Gene (OXEGENE) Study have recruited >1,000 families, mainly affected sister pair families, with the aim to identify genomic regions likely to harbour endometriosis susceptibility loci (29). In 2005, they have reported results from this linkage scan in 1,176 families (931 from the Australian group and 245 from the UK group), each with at least two members with surgically diagnosed disease. In total, 4,985 individuals were genotyped, including 2,709 women with endometriosis. The combined resource was aimed to have a 80% power to detect loci of modest effect, which is consistent with current expectations for most complex diseases. Analysis of the combined set of families identified significant linkage to a novel susceptibility locus on chromosome 10q26 [Maximum LOD Score=3.09], p=0.047] and suggestive linkage on chromosome 20p13 (MLS=2.09). Minor peaks (with MLS > 1.0) were also found on chromosomes 2, 6, 7, 8, 12, 14, 15, and 17. This is actually the first report of linkage to a major locus for endometriosis and these findings might favour the discovery of novel positional genetic variants that influence the risk of the disease. In this context, it has to be observed that chromosome 10q had already been implicated in the disease development toward a more severe and invasive form since in mice harbouring an oncogenic allele of K-ras resulting in the development of benign lesions reminiscent of endometriosis, a conditional deletion of PTEN which is located on 10q23.3, caused the progression toward the ovarian tumor (30). Although outside the reported interval, PTEN gene falls within the 99.9% CI. On the other hand, it has also to be considered that:

- although this study had sufficient power to detect linkage to loci with modest effect sizes, only one peak achieved significance. This would suggest substantial genetic heterogeneity and/or that more than one disease entity exists (27);
- stratification by subphenotypes such as more severe disease, pelvic pain, fertility status did not result in more significant effects. Except for families with two or more affected members reporting pelvic pain, no subphenotype stratum contributed more to the chromosome 10 peak, and no particular subanalysis contributed more to the chromosome 20 peak. Replication of these findings and/or candidate gene studies to progress from these regions of linkage to the causal genes would represent the next step for this approach.

5.3. Association studies

Using the candidate gene study approach described above, many gene variants have been investigated for putative associations with endometriosis. The main categories of candidate genes studied have been those that are involved in detoxification processes, sex steroid biosynthesis and their receptors, immune system-regulation. An extended list of candidate genes that have been investigated in endometriosis is given in Table 1.

As mentioned above and similarly to what had happened for other complex trait diseases, many initially positive findings have not been replicated. Guo has performed meta-analyses of the association studies performed so far evaluating glutathione S-transferase M1/T1 gene polymorphisms (31). No evidence of association between the GSTM1 null genotype and endometriosis has been found while the risk associated with the GSTT1 null genotype was 29% higher than the other genotypes (pooled OR=1.29, 95% CI=1.01-1.65). A meta-analysis of 12 association studies on 5 genes (CYP17, CYP19, AR, PR and ER) has been performed for genetic polymorphisms involving steroid receptors and steroidogenesis enzymes and reported positive findings were not supported by the data (32). Furthermore, in general, most of the reported associations for variants of
immunoregulatory genes or genes involved in growth, angiogenesis and detoxification processes have no strong evidence of replication. Lack of consistency is likely explained by:

- publication bias (31, 32);
- small sample sizes;
- many reported associations are often found in subgroups and usually without an a priori hypothesis about which polymorphism is advantageous and what the biological basis of this might be. Such results are therefore probably chance findings, and consequently most genetic associations fail to be replicated in independent studies (31, 32);
- failure to control for known risk factors. Because many genetic and non-genetic factors affect endometriosis, it is crucial to gather data on and statistically control for the influence of as many of these factors as possible;
- the modest effects of the causal variants on disease risk and the consequent false-negative studies. In several cases, the causal variant was associated with a 10–50% increased risk of disease, meaning that sample sizes in the thousands are required to achieve even a nominally significant p value < 0.05. Because most association studies had used samples of hundreds of individuals, the lack of consistency is not surprising.

6. THE FUTURE OF COMPLEX DISEASE GENETICS

There are several possible reasons why most of the studies described above have been inconclusive: the probable involvement of numerous genes with modest effects and the implementation of small-scale studies, are features that will produce many chance findings. Bearing these factors in mind, we hope will help future studies to provide more conclusive results.

In the future, the most comprehensive approach towards understanding complex disease would be complete genome resequencing in a large population of cases and controls. This approach would not be limited by the choice of candidate genes, it would cover the complete spectrum of coding and non-coding variants. Unfortunately, over the past few years, sequencing technologies have remained fundamentally unchanged, and it is the automation and refinement of existing methods that has led to a reduction in cost (2). Current costs are still higher than would be required for affordable whole-genome sequencing. Therefore, waiting for whole-genome sequencing to become a reality, the genome-wide association approach is acquiring great promise.

6.1. The genome-wide association approach

A genome-wide association approach is defined as an association study that surveys most of the genome for causal genetic variants. These studies will be greatly facilitated by the recognition of that many common variants are strongly correlated (in linkage disequilibrium), and hence redundant (5), meaning that a few hundred thousand SNPs will suffice to survey the approximately ten million variants with frequency 5% or greater. Because no assumptions are made about the genomic location of the causal variants, this approach could exploit the strengths of association studies without having to guess the identity of the causal genes. Thus, genome-wide association studies have the advantage that they do not depend on biologically plausible candidate genes or knowledge of specific variants (2). However, several objectives need to be met before genome-wide association studies become truly practical. First, a set of SNPs must be chosen that comprehensively captures the common variation across the genome. Methods for selecting such SNPs, and for using them efficiently for tests of association, are being developed. At present, costs are very high and not affordable in most laboratories. Because of this high cost, there is pressure to limit the sample size, with a consequent reduction in power. However, because variants that contribute to complex traits are likely to have modest effects, large sample sizes are crucial and since large number of hypotheses are tested, p-values must be corrected for multiple-hypothesis testing. On the basis of initial successes in candidate-gene association studies that represent only a tiny fraction of the genome, more comprehensive genome-wide association studies should greatly advance our understanding of the genetic basis of common diseases and complex traits.

7. REFERENCES

Genetics of endometriosis: current status and prospects

65. P. Vigano, M. Infantino, D. Lattuada, R. Laulett, E. Ponti, S. Somigliana, M. Vignali and A.M. Di Blasio:
Genetics of endometriosis: current status and prospects


**Key Words**: Endometriosis, Genetics, Linkage Studies, Association Studies, Review

**Send correspondence to**: Dr. P. Vigano’, Molecular Biology Laboratory, Istituto Auxologico Italiano, Via Zucchi 18, Cusano Milanino, Milano, Italy, Tel: 3902619112576, Fax: 3902619113033, E-mail: paola.vigano@unimi.it

http://www.bioscience.org/current/vol12.htm