

Genetics and infertility II

Affected sib-pair analysis in endometriosis

Stephen Kennedy^{1,4}, Simon Bennett² and Daniel E. Weeks³

¹Nuffield Department of Obstetrics and Gynaecology, University of Oxford, ²Oxagen Ltd, Oxford, UK and ³Department of Human Genetics, University of Pittsburgh, PA, USA

⁴To whom correspondence should be addressed at: Nuffield Department of Obstetrics & Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. E-mail: skennedy@molbiol.ox.ac.uk

This paper (i) reviews the current clinical and molecular genetic data which strongly suggest that endometriosis has a genetic basis; (ii) outlines the general principles of affected-sib pair analysis; and (iii) describes the Oxford Endometriosis Gene (OXEGENE) Study which aims, using a positional cloning approach, to identify susceptibility genes involved in the development of the disease.

Key words: affected sib-pair analysis/endometriosis/genetics/phenotype-genotype/positional cloning

TABLE OF CONTENTS

Introduction
What is a complex genetic trait?
Positional cloning approach
Affected sib-pair method
Disease phenotype for OXEGENE Study
OXEGENE Study Family Collection
Current status of OXEGENE and other linkage studies
Candidate functional genes
Conclusion
Acknowledgements
References

Introduction

Endometriosis is almost certainly a common disease, although its epidemiological characteristics in the general population are difficult to calculate accurately because of the need for a surgical procedure to define precisely who is and who is not affected. One of the few reliable estimates of incidence was derived from in-patient data collected in 1979 from women aged 15–49 years in Rochester, USA (Houston *et al.*, 1987). The incidence of histologically confirmed disease was 109 per 100 000 woman-years, but the estimated rates varied depending upon case definitions, to a maximum of 247 per 100 000 woman-years for all cases including those that were only suspected clinically. In the 20 years since that study was conducted, the epidemiology of endometriosis has become even more complicated because new

types of disease such as non-pigmented lesions (Jansen and Russell, 1986) and deeply infiltrating endometriosis (DIE) have been recognized (Koninckx *et al.*, 1994b).

Undoubtedly, the controversies surrounding the precise definition of the phenotype have hindered research into the aetiology of endometriosis. It is possible that subtle lesions, classical peritoneal implants, ovarian endometriomas and DIE represent different manifestations of a single disease entity; alternatively, they may be distinct pathologies with separate aetiologies (Koninckx *et al.*, 1999). It has even been suggested that minimal endometriosis is not a disease state at all: it could merely represent a variation of normal physiology resulting from retrograde menstruation, which is itself a common phenomenon (Koninckx *et al.*, 1994b).

Some, if not all, types of endometriotic implants probably result from the transport of menstrual debris into the peritoneal cavity via the Fallopian tubes, and the subsequent attachment of endometrial cells to the peritoneal/ovarian surface. An alternative explanation is that the implants result from in-situ metaplasia. In either model, there is proliferation and differentiation, and subsequent invasion into the underlying tissue although the precise cellular and molecular mechanisms involved remain unknown despite years of hypothesis-driven research.

Irrespective of the mechanisms involved, there is increasing evidence of a germline predisposition to endometriosis as: (i) familial clustering has been reported in humans (Kennedy *et al.*, 1995; Stefansson *et al.*, 2000) and in rhesus monkeys (Hadfield *et al.*, 1997b); (ii) there is concordance in monozygotic twins (Moen 1994; Hadfield *et al.*, 1997a; Treloar *et al.*, 1999); (iii) the age at onset of symptoms is similar in affected, non-twin sisters

Table I. Disease prevalence in the 1st degree relatives of affected women

Authors	Controls	Sisters	Mothers	Sisters or mothers
Simpson <i>et al.</i> (1980)	1st-degree relatives of patients' husbands 1/104 (1)	9/153 (5.8) versus 1/107 (0.9)	10/123 (8.1)- versus	
Coxhead and Thomas (1993)	1st-degree relatives of women with a normal pelvis	–	–	6/64 (9.4) versus 2/128 (1.6)
Moen and Magnus (1993) ^a	1st-degree relatives of women with a normal pelvis	25/523 (4.8) versus 1/169 (0.6)	20/515 (3.9) versus 1/149 (0.7)	–
Kennedy <i>et al.</i> (1998) ^b	No controls used	–	–	5/35 (14.3)

Values in parentheses are percentages.

^aDisease defined as endometriosis and/or adenomyosis.

^bDisease defined as radiological evidence of lesions >1 cm in diameter, excluding adenomyosis.

Table II. Candidate gene association studies (published)

	Polymorphism	Affecteds	Controls	
GALT N314D				
Cramer <i>et al.</i> (1996b)	Heterozygotes	10/33 (30)	15/111 (14)	<i>P</i> = 0.02
Morland <i>et al.</i> (1998)	Heterozygotes	14/78 (18)	42/248 (17)	NS
Hadfield <i>et al.</i> (1999)	Heterozygotes	22/148 (15)	6/53 (11)	NS
GSTM1 null mutation				
Baranov <i>et al.</i> (1996)	Homozygotes	34/42 (81)	26/67 (39)	<i>P</i> < 0.001
Baranova <i>et al.</i> (1997)	Homozygotes	43/50 (86)	33/73 (46)	<i>P</i> < 0.0001
Baxter <i>et al.</i> (2001)	Homozygotes	40/84 (48)	107/219 (49)	NS
GSTT1 null mutation				
Baranova <i>et al.</i> (1999)	Homozygotes	13/65 (20)	7/72 (10)	NS
NAT2 polymorphisms				
Baranova <i>et al.</i> (1999)	Slow acetylator	39/65 (60)	28/72 (39)	<i>P</i> < 0.02
ER gene polymorphisms				
Georgiou <i>et al.</i> (1999)	<i>Pvu</i> II (TA) _n repeats	Median = 15	Median = 20	<i>P</i> < 0.001

Values in parentheses are percentages.

GALT = galactose-1-phosphate uridyl transferase; GSTM1 = glutathione *S*-transferase M1; GSTT1 = glutathione *S*-transferase T1; NAT2 = arylamine *N*-acetyl transferase; ER = oestrogen receptor.

(Kennedy *et al.*, 1996); (iv) there is a 6–9-fold increased prevalence (see Table I) among the 1st-degree relatives of women with all disease severities compared to the general population (Simpson *et al.*, 1980; Coxhead and Thomas, 1993; Moen and Magnus, 1993); (v) disease prevalence, determined using magnetic resonance imaging (MRI), may be as high as 15% in the sisters of women with severe disease (Kennedy *et al.*, 1998), and (vi) certain genetic polymorphisms (see Table II) are more frequent in affected women than in controls (Baranov *et al.*, 1996; Cramer *et al.*, 1996b; Baranova *et al.*, 1997, 1999; Georgiou *et al.*, 1999).

There is also evidence that environmental factors, such as chronic exposure to dioxins, play a role in disease aetiology (Rier *et al.*, 1993; Koninckx *et al.*, 1994a). Endometriosis is therefore

likely to be a complex genetic trait in which multiple genes interact with each other and the environment to produce the disease phenotype (Kennedy, 1999). This raises the possibility of identifying susceptibility genes using a 'reverse genetics' approach (see 'Positional cloning approach' below) which is independent of any prior knowledge about aberrant mechanisms.

Lastly, it has been argued that in many ways ovarian endometriosis and DIE behave like a tumour, especially DIE, which invades the pouch of Douglas mimicking a malignant but self-limiting process. The use of molecular genetic techniques on endometriosis tissue (i.e. identification of clonality and somatic mutations) is therefore an appropriate strategy complementary to germline studies (Quirke and Mapstone, 1999). Ovarian endometriomas are monoclonal, which suggests that they result from the

proliferation of a single cell or endometrial gland (Jimbo *et al.*, 1997; Tamura *et al.*, 1998), but it is not known whether peritoneal lesions and/or DIE are derived from the same clone. Loss of heterozygosity (LOH) on chromosomes 9p (18%), 11q (18%) and 22q (15%) has been reported in ovarian endometriomas (Jiang *et al.*, 1996): in total, 11 of 40 (28%) cases demonstrated LOH at one or more of these loci. There are also some data to suggest that endometriosis is the clonal precursor of endometrioid and clear cell ovarian carcinomas (Jiang *et al.*, 1998). Mapping and identifying susceptibility genes in endometriosis will undoubtedly require a number of such different strategies. The research may also provide valuable information about the early stages of carcinogenesis.

What is a complex genetic trait?

Traits such as diabetes, asthma and hypertension, which do not exhibit classic Mendelian inheritance, are described as 'complex', 'multifactorial' or 'polygenic'. In other words, the trait is determined by many different genes each with allelic variations, although no single gene has a particularly large effect. Traits are typically described as quantitative (e.g. fasting glucose levels) or dichotomous (e.g. diabetes, present or absent based upon a predefined threshold).

The number of genes responsible for predisposing an individual to a disease and the relative contribution of those genes may vary considerably, which is one of the principal reasons why dissecting out the genetic basis of complex traits is so difficult. The other major problem is that a consistent relationship between genotype and phenotype does not always appear to exist. For example, the same genotype can result in different phenotypes depending upon the influence of chance or environmental factors, e.g. an individual who inherits asthma-predisposing genes may not develop the disease unless exposed to environmental pollutants (incomplete penetrance). Another individual may develop asthma as a consequence of some environmental exposure without inheriting any predisposing genes (phenocopy). Sometimes, a phenotype can result from mutations in different genes (genetic heterogeneity) or different mutations at a single locus, e.g. breast/ovarian cancer at the *BRCA1* locus (allelic heterogeneity). Allelic heterogeneity does not interfere with gene mapping because all families will show linkage to the same chromosomal region. Genetic heterogeneity can cause problems, however, because a chromosomal region may be linked to the phenotype in some families and not others.

The extent to which a complex genetic trait clusters within families is defined by its relative risk (λ_R). This is the risk of an affected relative having the disease compared to the risk in the general population (the term λ_S is used in the case of sibs). The power to detect linkage (and therefore the number of affected sib-pairs required) is directly related to the size of λ_R . Generally speaking, the higher the ratio (i.e. >10) the easier it is to identify disease-predisposing genes. Some examples of values for common conditions are Crohn's disease (~20); type 1 diabetes (15), type 2 diabetes (3.5) and rheumatoid arthritis (~6). λ_R is clearly difficult to estimate in endometriosis: the prevalence in the general population is uncertain as the diagnosis is only reliably made at surgery and the disease status of an affected woman's sister will be unknown unless she has had a laparoscopy.

Nevertheless, it is reasonable to suggest that for all disease severities λ_S is 6–9, based upon the three studies in the literature that have reported relative risks in the sisters of affected women (Coxhead and Thomas, 1993; Moen and Magnus, 1993; Simpson *et al.*, 1980). In women with the most severe forms of endometriosis, λ_S may be as high as 15 (Kennedy *et al.*, 1998) based upon the MRI findings in the sisters of women with revised American Fertility Society (rAFS) Stage III-IV disease (Anonymous, 1985). A higher value for λ_S in the sisters of women with more severe disease is consistent with the general finding that severe phenotypic types in most complex traits usually display the highest λ_S ratios.

Positional cloning approach

The 'reverse genetics' or 'positional cloning' approach to unravel the genetic basis of a disease with unknown biochemical defects broadly involves: (i) chromosomal localization of the disease locus; (ii) identification of the gene(s) and mutations; and (iii) determination of abnormal function associated with the mutations. Localization is achieved by studying the segregation of DNA markers with disease in selected families (linkage analysis). Thus, genes are identified by first mapping their approximate chromosomal localization using genetic markers without any knowledge of disease mechanisms. The next stages (which are not the subject of this review) involve identification of a candidate 'positional' gene in that region, often based upon some theoretical or actual knowledge of disease mechanisms, and showing that mutations in that gene occur more frequently in affected individuals than in controls. This has been called the 'positional candidate' approach.

The approach relies upon the fact that reciprocal exchange of DNA sequences occurs by recombination (cross-over) when homologous chromosomes pair at meiosis. The probability that an exchange has occurred is defined by the distance between any two genetic loci. Thus, two loci (such as a disease gene and a genetic marker) that are very close together on a chromosome will rarely be separated by recombination. They will therefore tend to co-segregate within families, i.e. there will be a tendency, depending upon the disease's mode of inheritance, for affected family members to inherit from their parents the same chromosomal region in which the mutated gene and the genetic marker are located. Conversely, the further apart two loci are on a chromosome, the more likely it is that a cross-over will separate them. In other words, a disease gene and an unlinked genetic marker will segregate randomly, i.e. their inheritance will not deviate from simple Mendelian rules. The genetic distance between loci is measured by the recombination fraction (θ): two loci that show 1% recombination are defined as 1 centimorgan (cM) apart on a genetic map. One cM is approximately equivalent to 1 megabase (Mb), i.e. 1 million base pairs, of DNA although this value varies between chromosomes and between the sexes as recombination is more frequent in female meiosis.

In principle, any genetic marker that is inherited in a Mendelian fashion can be used provided that its chromosomal location is known and it is sufficiently informative or polymorphic (i.e. multiple alleles exist at the marker locus, thereby increasing the probability that any individual will be heterozygous). Restriction-fragment length DNA polymorphisms (Botstein *et al.*, 1980) and sequence polymorphisms defined by variable numbers of tandem

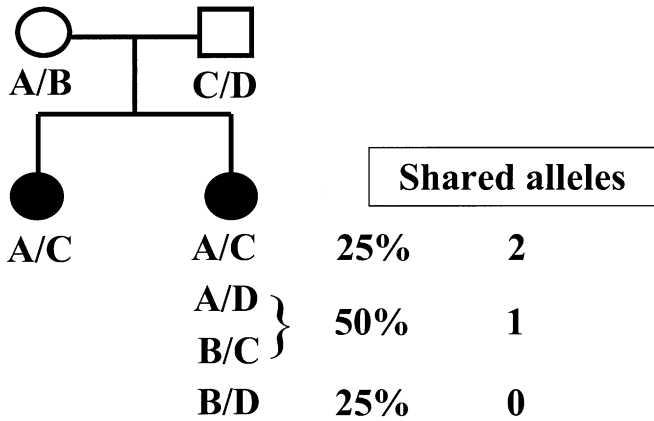


Figure 1. Allele sharing in affected sib-pairs. Two affected sisters can share two, one or zero alleles at any marker locus with 25, 50 and 25% likelihoods in the absence of linkage, i.e. with random segregation. In the simplified example above, there are four alleles (A, B, C and D) at a hypothetical marker locus. Both parents are heterozygous (mother's genotype = A/B and father's genotype = C/D). Assuming that the older affected sister has inherited the A/C genotype, then there is a 25% chance in the absence of linkage that the younger sister has the A/C genotype; a 50% chance of the A/D or B/C genotypes, and a 25% chance of the B/D genotype. Evidence of linkage of the disease gene to the marker locus would exist if the affected sisters shared the two alleles A and C, identical-by-descent from their parents, more frequently than expected by chance alone.

repeats (Nakamura *et al.*, 1987) provided a set of genetic markers that were for the first time sufficiently informative and evenly spaced throughout the genome to make linkage analysis and gene mapping on a large scale feasible. Southern blotting was originally used to study such polymorphisms; since then, the use of the polymerase chain reaction to amplify DNA from the region of interest before fragment analysis has made the process of genotyping much easier and more rapid.

For single gene disorders, localization was a relatively straightforward process. The approach was successful in diseases such as Duchenne muscular dystrophy (Monaco and Kunkel, 1988) and cystic fibrosis (Kerem *et al.*, 1989). However, it proved more difficult to apply the methods to the identification of susceptibility loci in complex traits given that multiple genetic loci are involved. The available markers were not sufficiently informative and the techniques were too laborious to permit large numbers of individuals to be genotyped. The discovery of microsatellites—highly polymorphic dinucleotide, mostly (CA)_n, tandem repeats—considerably increased the number of markers available (Weber and May, 1989) and the development of automated DNA fragment analysis with fluorescently labelled PCR primers made high-throughput genotyping a reality (Wilson *et al.*, 1990; Reed *et al.*, 1994). The number of microsatellites available has since dramatically increased: in 1992, Weissenbach *et al.* reported on the development of 600 microsatellite markers covering a linkage distance spanning ~90% of the genome (Weissenbach *et al.*, 1992). Then, in 1996, a map consisting of 5264 microsatellites was published spanning a sex-averaged genetic distance of 3699 cM with an average interval size of 1.6 cM (Dib *et al.*, 1996). The latest generation of markers are bi-allelic, single nucleotide polymorphisms (SNP). SNP can be studied on solid-state arrays, which allows for very high throughput genotyping; however, their

use at present is confined to fine-mapping experiments, not genome-wide screens.

Affected sib-pair method

The affected sib-pair (ASP) method is the statistical approach that is most commonly used to identify susceptibility loci in complex genetic traits (Weeks and Lange, 1988). This non-parametric method differs from classical linkage analysis in that it makes no assumptions about the mode of inheritance of a trait. It merely aims to demonstrate whether identical chromosomal regions are inherited by affected sibs more commonly than would be expected by chance alone. Affected sibs and their parents are genotyped to determine how often marker alleles in the affected sibs are shared identical-by-descent (IBD), i.e. inherited from a common parent (see Figure 1). Allele sharing at a chromosomal region that is occurring more frequently than expected by chance, i.e. that is not consistent with random Mendelian segregation, is indicative of linkage. Predisposing loci in multifactorial diseases such as type 1 diabetes (Davies *et al.*, 1994), asthma (Sandford *et al.*, 1993) and inflammatory bowel disease (Satsangi *et al.*, 1996) have been mapped using this strategy. In reproductive medicine, the two diseases that are being investigated in this way are polycystic ovary syndrome and endometriosis.

The advantages of the ASP method in a disease such as endometriosis include: (i) only affected individuals are recruited (i.e. there is no need to identify at laparoscopy those women in a family who are unaffected); (ii) families with only two affected sisters are much easier to find than large pedigrees; and (iii) the parents of women with endometriosis are usually alive, which is not the case in conditions such as hypertension and type 2 diabetes.

The likelihood that two loci are linked is typically measured by the 'lod' score. This is defined as the logarithm (base 10) of the odds ratio between the likelihood that the loci are linked (with recombination fraction = θ) and not linked (with recombination fraction = 0.5). For Mendelian diseases, a lod score greater than +3 is highly suggestive of linkage, and one less than -2 is evidence against linkage. A lod score of 3 corresponds to a significance level of 0.05 at a single point, i.e. a 0.05 chance of a false positive result. If multiple markers are used, as in a genome-wide screen, the chance of a spurious positive result is greater than if only one marker is used. It follows that the definition of statistical significance needs to be more robust in complex genetic traits than in Mendelian diseases, i.e. higher lod scores are required, and that the results of a genome-wide screen need to be replicated in separate data sets before a linkage can be considered confirmed. This is a complicated issue that is beyond the remit of this review (Lander and Schork, 1994; Risch, 2000). However, it is important to emphasize that the power of the ASP method to detect linkage is related to many factors including the sample size (i.e. the number of ASP families genotyped); the marker allele frequencies; the size of λ_s , and the number of susceptibility loci and their relative contributions to the value of λ_s . Consequently, it is likely that thousands of ASP families will be needed to identify susceptibility genes with small effects in certain diseases.

Disease phenotype for OXEGENE study

The Oxford Endometriosis Gene (OXEGENE) Study is a collaboration between the Oxford Group (led by Prof. David Barlow, Helen Mardon and Stephen Kennedy), a number of international researchers and Oxagen Ltd, an Oxford-based genomics company that was formed in 1997 as a spin-off from the Wellcome Trust Centre for Human Genetics at the University of Oxford. The aim of the study is to identify susceptibility genes in endometriosis using a positional cloning approach.

It is vital to have a reliable definition of disease phenotype in such a study. Without a clear and consistent definition it may be impossible to determine the relative contributions of genetic and environmental influences on disease aetiology and progression. Defining phenotype precisely has been difficult in some conditions, particularly psychiatric disorders because diagnoses are often subjective and they have a tendency to vary over time (Kelsoe *et al.*, 1989). In many respects, the phenotype is much easier to define in endometriosis because, whether disease is present or not at surgery, it is usually obvious and histological confirmation of disease is often obtained as well. What is uncertain, however, is whether all disease severities and types should be included under one phenotypic classification.

One approach to this problem is to define the phenotype as unequivocally as possible, i.e. as severe endometriosis, whilst realizing that such a definition may have its limitations. The other approach is to define all disease severities and types as a single phenotype and then subcategorize if necessary at a later date on the basis of genotype findings. Initially, it was decided to use surgically confirmed, rAFS Stage III-IV endometriosis (termed Stage B disease for linkage purposes) as the phenotype. This has necessitated confirming the disease status of every woman recruited by writing to the relevant gynaecologist(s) to obtain copies of the surgical records with histological confirmation if available. One person (S.K.) has reviewed the records to assign disease severity; if the woman has had more than one operation, the phenotype is defined on the basis of the most severe disease found at surgery. Women have not been included in the study if: (i) the records were unobtainable; (ii) only rAFS Stage I-II endometriosis (termed Stage A disease) has been found, or (iii) the diagnosis was questionable. In addition, whenever it was difficult to determine whether Stage A or B disease was found at surgery, the woman has been assigned to the lesser category.

Recently, however, the family collection strategy has changed and women with less severe forms of endometriosis are now being included, provided that they have at least one affected relative with Stage B disease and they have more than just a few peritoneal implants, i.e. some ovarian disease and/or some pelvic adhesions must have been found at surgery. The main reasons for the alteration in strategy are that, within Stage B families, many women give a history of progression over time from rAFS Stage I-II to Stage III-IV disease and it is more consistent with the approach being adopted by other research groups investigating the genetic basis of endometriosis.

It is readily accepted that it can be difficult to classify disease retrospectively on the basis of the surgical records alone as they often contain limited details. However, there is no other choice because it would be virtually impossible to collect the large numbers of affected families required for a linkage study in a

prospective study. Inevitably, the results of linkage analysis may lead to a new way of defining phenotype that may or may not justify the hypotheses made in this study about disease classification.

OXEGENE Study Family Collection

To recruit families, two different approaches are being used: 'bottom-up' and 'top-down'. The 'bottom-up' approach has involved systematic searches through the surgical records of collaborating clinicians in Belgium (Prof. Philippe Koninckx, Leuven); Britain (Prof. Chris Sutton, Guildford); Ireland (Prof. Colm O'Herlihy and Mary Wingfield, Dublin), and USA (Matthew Peterson, Salt Lake City). Women with surgically confirmed endometriosis in the clinic databases are asked whether they have any affected sisters (or other relatives) who might be willing to participate. If they do not have any relatives with surgically confirmed disease, blood is taken from the single affected women and their parents for future fine mapping studies.

The 'top-down' approach involves finding sister-pairs: (i) by placing articles in women's magazines and the newsletters of national endometriosis self-help groups, e.g. the American Endometriosis Association; (ii) via the OXEGENE web-site (www.medicine.ox.ac.uk/ndog/oxegene/oxegene.htm) which has been a rapid and cost-effective means of recruiting families (Suchard *et al.*, 1998); and (iii) with the collaboration of leading research centres throughout the world (see web-site for details).

Women are invited to participate in the OXEGENE Study via the approaches outlined above. In the UK, the study has received both MREC (Multi-centre Research Ethics Committee) and LREC (Local Research Ethics Committee) approval; the appropriate LREC approval has also been obtained in Belgium, Ireland and the USA. If a woman is willing to participate, she is sent detailed information about the purpose of the research and the extent of her and her family's involvement. Based upon the written information sent, she is asked to sign a consent form allowing access to her medical records and the use of her DNA to investigate the genetic basis of endometriosis. Assurance is given that her DNA will not be used for any other research purposes. If the woman feels that she has not been provided with sufficient information to enable her to give informed consent, she can speak to a member of the research team by ringing a free-phone number (available to international participants as well). Those women who have responded via the web-site can provide all the necessary information by e-mail or by completing an electronic form.

Each proband is asked to provide simple demographic data and information about her affected sisters (and/or other relatives) having asked them for permission beforehand to allow their details to be passed on. The affected family members are then contacted and asked to provide the same information as the probands. Once disease status has been confirmed, 10 ml blood samples are collected in plastic EDTA tubes from the affected sisters and their parents and sent in specially designed polystyrene boxes by routine post (or collected by courier in North America). If one parent is dead, blood is also taken from as many 1st-degree relatives of that dead individual as possible (irrespective of their sex) to allow his/her genotype to be estimated. A detailed epidemiological questionnaire is also sent to all affected family

Table III. Disease stage combinations and family numbers per combination in the 85 OXEGENE families genotyped in the first genome-wide screen^a

Stage of disease combinations	Family numbers per combination
BB	41
BBB	10
BBBB	3
BBBBA	1
BBAA	2
BBA	6
BAA	2
BA	11
B?	8
AA	1

^aFor example, BBB = three relatives or three BB sib-pairs with Stage III-IV disease. In eight families, the disease stage in one relative could not be determined accurately enough from the surgical records.

members at the same time as the blood kits. DNA is extracted from the blood samples and stored at -80°C for subsequent analyses.

Current status of OXEGENE and other linkage studies

Through these various family collection strategies, more than 1500 probands have been identified in the OXEGENE Study principally from the UK and USA. Those women and their family members are at various stages of the recruitment process. The first 85 families containing 128 affected sister-pairs have been genotyped using an ABI 377 automated sequencer. The families were collected by the Oxford group ($n=57$) or by collaborators in Utah ($n=12$), Norway ($n=3$), Russia ($n=3$), Germany ($n=3$), Belgium ($n=2$) and Australia ($n=4$). The disease stages in these families are shown in Table III. Other affected relative-pairs within the families, which will be used in later analyses, are aunt-niece ($n=10$), first cousin ($n=8$), and grandmother-granddaughter ($n=1$) pairs. Marker data were generated (~ 400 markers at ~ 10 cM) and analysed for a total of 121 affected sister-pairs (seven families were excluded from the final analysis because of misinheritance). There is preliminary evidence of linkage in these families to five chromosomal regions, with non-parametric linkage (NPL) scores >2 . Genotyping to replicate these linkages in a second data-set in collaboration with Oxagen Ltd is currently in progress (Oxford Endometriosis Group, 2000).

Lastly, a similar strategy has been adopted by a Collaborative Australian Group (involving the Department of Epidemiology at the Queensland Institute of Medical Research; The Cooperative Research Centre for Discovery of Genes for Common Human Disease; The Walter and Eliza Hall Institute for Medical Research, and the Institute for Molecular Bioscience, University of Queensland). To date, marker data have been generated (~ 400 markers at ~ 10 cM) and analysed for a total of 289 families containing 374 sister-pairs plus other affected relatives. Suggestive linkage has been identified for one locus, and possible linkage for five other loci (Treloar *et al.*, 2000).

Table IV. Genehunter version 1.3 (multipoint analysis)

Marker	Heterozygosity (%)	No. of alleles observed	Genehunter P value
<i>DIS2635</i>	85	13	NS
<i>DIS2844</i>	83	11	NS
<i>DIS2762</i>	82	12	NS

Candidate functional genes

An alternative but highly speculative approach to positional cloning is the investigation of candidate 'functional' genes. These are chosen based upon knowledge of their putative or actual role in mechanisms involved in disease development. Candidate genes are often investigated in association studies by comparing the frequency of marker alleles in affected individuals and controls. Differences in frequency arise if the marker is causally implicated in the disease or if it is in linkage disequilibrium with a susceptibility locus, i.e. haplotype combinations of alleles at different loci occur more frequently than would be expected from random association. Examples of association reported in complex traits include apolipoprotein E in Alzheimer's disease (Pericak and Haines, 1995) and the insulin gene in type 1 diabetes (Bennett *et al.*, 1995; Bennett and Todd, 1996). However, spurious associations can arise by chance because of ethnic variations in allele frequencies, too small a sample size, and/or the use of inappropriate control groups. Another approach therefore is to investigate candidate genes using the ASP method. The examples below illustrate how genes implicated in the aetiology of endometriosis through association studies have been investigated in linkage studies.

The N314D polymorphism in the gene encoding for galactose-1-phosphate uridyl transferase (GALT) has previously been associated with premature ovarian failure (Cramer *et al.*, 1995), Müllerian anomalies (Cramer *et al.*, 1996a) and ovarian cancer (Cramer *et al.*, 1994). GALT is an enzyme involved in galactose metabolism. The N314D polymorphism results from an A to G transition in exon 10 of the gene, substituting aspartate for asparagine. The GALT gene locus has been mapped to the short arm of chromosome 9p21, a region where loss of heterozygosity at candidate ovarian tumour suppressor loci has been reported in endometriotic tissue (Jiang *et al.*, 1996).

Cramer *et al.* (1996b) investigated the prevalence of the variant in endometriosis given the known association between the disease and Müllerian anomalies. They reported that 30% (10/33) of women with endometriosis (rAFS Stages I-IV) carried at least one N314D allele compared to 14% (15/111) of controls, consisting of pre-menopausal women from the general population (excluding those with a past history of endometriosis, hysterectomy or bilateral oophorectomy). Six of the 11 (55%) women with Stage III-IV disease carried at least one N314D allele. However, these findings have not been replicated in the UK (Morland *et al.*, 1998; Hadfield *et al.*, 1999) or Icelandic populations (Geirsson *et al.*, 2000). In an additional study, 203 women with endometriosis (rAFS Stages I-IV) and their relatives from 59 Icelandic families were genotyped using 30 microsatellite markers on chromosome

9, but no evidence was found for linkage to this region (Geirsson *et al.*, 2000).

Enzymes belonging to the glutathione *S*-transferase (GST) family are involved in the detoxification of polycyclic aromatic hydrocarbons (found in tobacco smoke, food and combustion fumes) and pesticides. Null mutations occurring in two genes, *GSTM1* and *GSTT1*, result in loss of enzyme activity, which has been associated with environmentally induced cancers (Seidegard *et al.*, 1990; Kempkes *et al.*, 1996).

Mutations in the *GST* gene have been investigated in endometriosis because of the evidence suggesting that exposure to environmental pollutants may be a risk factor (Rier *et al.*, 1993). In a Slavic sample, 81% (34/42) of women with endometriosis (disease severity not specified) had the null mutation compared with 39% (26/67) of healthy male and female controls. (Baranov *et al.*, 1996). The prevalence of *GSTM1* homozygotes in a group of French women with endometriosis (rAFS Stages I-IV) was higher than in controls: 86% (43/50) versus 46% (33/72) (Baranova *et al.*, 1997). More affected women (rAFS Stages I-IV) were reported as *GSTT1* homozygotes than controls [20% (13/65) versus 10% (7/72)] but the difference was not statistically significant (Baranova *et al.*, 1999). More recently (published in abstract form only) the same group has suggested that the presence of the *GSTM1* mutation may reduce the efficacy of treatment (Baranova *et al.*, 2000). In the UK population, however, there is no evidence of a significant association: 47.6% (40/84) of affected women carried the null mutation compared to 48.9% (107/219) of controls (Baxter *et al.*, 2001).

On the basis of the positive association studies, the Oxford Group genotyped 192 individuals from 43 families containing 52 sister-pairs with rAFS Stage III-IV endometriosis using three microsatellite markers selected from the region of chromosome 1p13 to which *GSTM1* has been mapped (R.M.Hadfield *et al.*, unpublished data). The markers (*DIS2844*, *DIS2635* and *DIS2762*) were chosen for their high heterozygosity to maximize the information content.

The three PCR products from each patient were pooled and run on an ABI 377 (PE Biosystems). Analysis was conducted using Genescan™ and Genotyper™ software (PE Biosystems, USA). The proportion of patients successfully genotyped was 99% for *DIS2635*, 93% for *DIS2762* and 91% for *DIS2844*. The accuracy of the genotyping data was checked by comparing the calculated heterozygosity of each marker to published values from Généthon. The observed and expected values were not significantly different for any of the three markers. The observed heterozygosity was 85% for *DIS2635* (published value 87%), 83% *DIS2762* (published value 81%) and 82% for *DIS2844* (published value 81%). No significant deviations from expected values (Table IV) were observed using the program Genehunter (Kruglyak *et al.*, 1996), which suggests that this chromosomal region may not be linked to endometriosis, but the numbers are clearly too small to draw any firm conclusions.

Conclusion

The identification of disease susceptibility loci through a positional cloning approach using affected sib-pair analysis should in time lead to a clearer understanding of the molecular and cellular mechanisms involved in the aetiology of endome-

triosis. Clearly, this can only be achieved through the collection of very large numbers of well-characterized families. It is unlikely that endometriosis is a single disease entity and it should also be possible to categorize the disease into sub-types based upon genotype data. The strategy will, it is hoped, lead to the discovery of new drug targets in this disease which will be a major advance given that treatment to date has focused upon very unsophisticated manipulation of the hypothalamic-pituitary-ovarian axis. Lastly, it is possible that informing a young woman that she is at increased risk of developing the disease, based upon the information gained from a prognostic test, will enable her to make important life decisions such as whether to have children earlier than her planned career would have permitted.

Acknowledgements

The OXEGENE Study was initially funded by the Medical Research Council with a ROPA grant and subsequently by Oxagen Ltd. Related academic-based research is being funded by The Wellcome Trust, the Medical Research Council, WellBeing, the Oxford and Anglia R&D Fund, the Oxford Endometriosis Trust Fund and an educational grant from Searle, USA.

References

- Anonymous (1985) Revised American Fertility Society classification of endometriosis. *Fertil. Steril.*, **43**, 351–352.
- Baranov, V.S., Ivaschenko, T., Bakay, B. *et al.* (1996) Proportion of the *GSTM1* 0/0 genotype in some Slavic populations and its correlation with cystic fibrosis and some multifactorial diseases. *Hum. Genet.*, **97**, 516–520.
- Baranova, H., Bothorishvilli, R., Canis, M. *et al.* (1997) Glutathione *S*-transferase M1 gene polymorphism and susceptibility to endometriosis in a French population. *Mol. Hum. Reprod.*, **3**, 775–780.
- Baranova, H., Canis, M., Ivaschenko, T. *et al.* (1999) Possible involvement of arylamine *N*-acetyltransferase 2, glutathione *S*-transferase M1 and T1 genes in the development of endometriosis. *Mol. Hum. Reprod.*, **5**, 636–641.
- Baranova, H., Jarmolinskaya, M., Ivaschenko, T. *et al.* (2000) *GSTM1* and *NAT2* detoxification system gene polymorphisms. Feasible predictive test for the treatment efficiency response in endometriosis patients. *Cytogenet. Cell Genet.*, **85**, 98 (Abstract).
- Baxter, S.W., Thomas, E.J. and Campbell, I.G. (2001) *GSTM1* null polymorphism and susceptibility to endometriosis and ovarian cancer. *Carcinogenesis*, **22**, 63–65.
- Bennett, S.T. and Todd, J.A. (1996) Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu. Rev. Genet.*, **30**, 343–370.
- Bennett, S.T., Lucassen, A.M., Gough, S.C.L. *et al.* (1995) Susceptibility to human type 1 diabetes at *IDDM2* is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.*, **9**, 284–292.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, **32**, 314–331.
- Coxhead, D. and Thomas, E.J. (1993) Familial inheritance of endometriosis in a British population: a case control study. *J. Obstet. Gynaecol.*, **13**, 42–44.
- Cramer, D.W., Muto, M.G., Reichardt, J.K. *et al.* (1994) Characteristics of women with a family history of ovarian cancer. I. Galactose consumption and metabolism. *Cancer*, **74**, 1309–1317.
- Cramer, D.W., Xu, H. and Harlow, B.L. (1995) Family history as a predictor of early menopause. *Fertil. Steril.*, **64**, 740–745.
- Cramer, D.W., Goldstein, D.P., Fraer, C. and Reichardt, J.K. (1996a) Vaginal agenesis (Mayer-Rokitansky-Kuster-Hauser syndrome) associated with the N314D mutation of galactose-1-phosphate uridylyl transferase (GALT). *Mol. Hum. Reprod.*, **2**, 145–148.
- Cramer, D.W., Hornstein, M.D., Ng, W.G. and Barbieri, R.L. (1996b) Endometriosis associated with the N314D mutation of galactose-1-phosphate uridylyl transferase (GALT). *Mol. Hum. Reprod.*, **2**, 149–152.
- Davies, J.L., Kawaguchi, Y., Bennett, S.T. *et al.* (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature*, **371**, 130–136.
- Dib, C., Faure, S., Fizames, C. *et al.* (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**, 152–154.

- Geirsson, R.T., Einarsdottir, E., Gudnason, A. *et al.* (2000) Neither linkage nor association to the *GALT* gene in chromosome 9 is seen in Icelandic endometriosis patients. Poster at World Endometriosis Congress, London.
- Georgiou, I., Syrrou, M., Bouba, I. *et al.* (1999) Association of estrogen receptor gene polymorphisms with endometriosis. *Fertil. Steril.*, **72**, 164–166.
- Hadfield, R.M., Mardon, H.J., Barlow, D.H. and Kennedy, S.H. (1997a) Endometriosis in monozygotic twins. *Fertil. Steril.*, **68**, 941–942.
- Hadfield, R.M., Yudkin, P.L., Coe, C.L. *et al.* (1997b) Risk factors for endometriosis in the rhesus monkey (*Macaca mulatta*): a case-control study. *Hum. Reprod. Update*, **3**, 109–115.
- Hadfield, R.M., Manek, S., Nakago, S. *et al.* (1999) Absence of a relationship between endometriosis and the N314D polymorphism of galactose-1-phosphate uridyl transferase in a UK population. *Mol. Hum. Reprod.*, **5**, 990–993.
- Houston, D.E., Noller, K.L., Melton, L.J. *et al.* (1987) Incidence of pelvic endometriosis in Rochester, Minnesota, 1970–1979. *Am. J. Epidemiol.*, **125**, 959–969.
- Jansen, R.P. and Russell, P. (1986) Nonpigmented endometriosis: clinical, laparoscopic, and pathologic definition. *Am. J. Obstet. Gynecol.*, **155**, 1154–1159.
- Jiang, X., Hitchcock, A., Bryan, E.J. *et al.* (1996) Microsatellite analysis of endometriosis reveals loss of heterozygosity at candidate ovarian tumor suppressor gene loci. *Cancer Res.*, **56**, 3534–3539.
- Jiang, X., Morland, S.J., Hitchcock, A. *et al.* (1998) Allelotyping of endometriosis with adjacent ovarian carcinoma reveals evidence of a common lineage. *Cancer Res.*, **58**, 1707–1712.
- Jimbo, H., Hitomi, Y., Yoshikawa, H. *et al.* (1997) Evidence for monoclonal expansion of epithelial cells in ovarian endometrial cysts. *Am. J. Pathol.*, **150**, 1173–1178.
- Kelsoe, J.R., Ginns, E.I., Egeland, J.A. *et al.* (1989) Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature*, **342**, 238–243.
- Kempkes, M., Golka, K., Reich, S. *et al.* (1996) Glutathione S-transferase GSTM1 and *GSTT1* null genotypes as potential risk factors for urothelial cancer of the bladder. *Arch. Toxicol.*, **71**, 123–126.
- Kennedy, S. (1999) The genetics of endometriosis. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, **82**, 129–133.
- Kennedy, S., Mardon, H. and Barlow, D. (1995) Familial endometriosis. *J. Assist. Reprod. Genet.*, **12**, 32–34.
- Kennedy, S., Hadfield, R., Mardon, H. and Barlow, D. (1996) Age of onset of pain symptoms in non-twin sisters concordant for endometriosis. *Hum. Reprod.*, **11**, 403–405.
- Kennedy, S., Hadfield, R., Westbrook, C. *et al.* (1998) Magnetic resonance imaging to assess familial risk in relatives of women with endometriosis. *Lancet*, **352**, 1440–1441.
- Kerem, B., Rommens, J.M., Buchanan, J.A. *et al.* (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science*, **245**, 1073–1080.
- Koninckx, P.R., Braet, P., Kennedy, S.H. and Barlow, D.H. (1994a) Dioxin pollution and endometriosis in Belgium. *Hum. Reprod.*, **9**, 1001–1002.
- Koninckx, P.R., Oosterlynck, D., D'Hooghe, T. and Meuleman, C. (1994b) Deeply infiltrating endometriosis is a disease whereas mild endometriosis could be considered a non-disease. *Ann. NY Acad. Sci.*, **734**, 333–341.
- Koninckx, P.R., Barlow, D. and Kennedy, S. (1999) Implantation versus infiltration: the Sampson versus the endometriotic disease theory. *Gynecol. Obstet. Invest.*, **47** (Suppl. 1), 3–9.
- Kruglyak, L., Daly, M.J., Reeve, D.M. and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.
- Lander, E. S. and Schork, N. J. (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.
- Moen, M.H. (1994) Endometriosis in monozygotic twins. *Acta Obstet. Gynecol. Scand.*, **73**, 59–62.
- Moen, M.H. and Magnus, P. (1993) The familial risk of endometriosis. *Acta Obstet. Gynecol. Scand.*, **72**, 560–564.
- Monaco, A.P. and Kunkel, L.M. (1988) Cloning of the Duchenne/Becker muscular dystrophy locus. *Adv. Hum. Genet.*, **17**, 61–98.
- Morland, S.J., Jiang, X., Hitchcock, A. *et al.* (1998) Mutation of galactose-1-phosphate uridyl transferase and its association with ovarian cancer and endometriosis. *Int. J. Cancer*, **77**, 825–827.
- Nakamura, Y., Leppert, M., O'Connell, P. *et al.* (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, **235**, 1616–1622.
- Oxford Endometriosis Group, Oxegene Collaborative Group and Oxagen (2000) Preliminary results from the Oxegene study: genome-wide screen for endometriosis susceptibility loci in over 100 affected sister-pairs. Poster at World Endometriosis Congress, London.
- Pericak, V.M. and Haines, J.L. (1995) Genetic susceptibility to Alzheimer disease. *Trends Genet.*, **11**, 504–508.
- Quirke, P. and Mapstone, N. (1999) The new biology: histopathology. *Lancet*, **354** (Suppl. 1), S126–S131.
- Reed, P.W., Davies, J.L., Copeman, J.B. *et al.* (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automatic genome mapping. *Nat. Genet.*, **7**, 390–395.
- Rier, S.E., Martin, D.C., Bowman, R.E. *et al.* (1993) Endometriosis in rhesus monkeys (*Macaca mulatta*) following chronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Fundam. Appl. Toxicol.*, **21**, 433–441.
- Risch, N. J. (2000) Searching for genetic determinants in the new millennium. *Nature*, **405**, 847–856.
- Sandford, A.J., Shirakawa, T., Moffatt, M.F. *et al.* (1993) Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q. *Lancet*, **341**, 332–334.
- Satsangi, J., Parkes, M., Louis, E. *et al.* (1996) Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat. Genet.*, **14**, 199–202.
- Seidegard, J., Pero, R.W., Markowitz, M.M. *et al.* (1990) Isoenzyme(s) of glutathione transferase (class Mu) as a marker for the susceptibility to lung cancer: a follow up study. *Carcinogenesis*, **11**, 33–36.
- Simpson, J.L., Elias, S., Malinak, L.R. and Buttram-VC, J. (1980) Heritable aspects of endometriosis. I. Genetic studies. *Am. J. Obstet. Gynecol.*, **137**, 327–331.
- Spitsyn, V.A., Andreeva, E.N., Adamian, L.V. *et al.* (1996) Genetic aspects of endometriosis: features of the distribution of polymorphic gene frequencies. *Genetika*, **32**, 1693–1699.
- Stefansson, H., Geirsson, R.T., Jonsson, H. *et al.* (2000) Familial clustering of endometriosis evident from a population based study. Poster at World Endometriosis Congress, London.
- Suchard, M.A., Hadfield, R., Elliott, T. and Kennedy, S. (1998) Beyond providing information: the Internet as a research tool in reproductive medicine. *Hum. Reprod.*, **13**, 6–7.
- Tamura, M., Fukaya, T., Murakami, T. *et al.* (1998) Analysis of clonality in human endometriotic cysts based on evaluation of X chromosome inactivation in archival formalin-fixed, paraffin-embedded tissue. *Lab. Invest.*, **78**, 213–218.
- Treloar, S.A., O'Connor, D.T., O'Connor, V.M. and Martin, N.G. (1999) Genetic influences on endometriosis in an Australian twin sample. *Fertil. Steril.*, **71**, 701–710.
- Treloar, S.A., Bahlo, M., Ewen, K. *et al.* (2000) Suggestive linkage for endometriosis found in genome-wide scan. *Am. J. Hum. Genet.*, **67** (Suppl. 2), 318 (Abstract).
- Weber, J.L. and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**, 388–396.
- Weeks, D.E. and Lange, K. (1988) The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.*, **42**, 315–326.
- Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A second-generation linkage map of the human genome. *Nature*, **359**, 794–801.
- Wilson, R.K., Chen, C., Avdalovic, N. *et al.* (1990) Development of an automated procedure for fluorescent DNA sequencing. *Genomics*, **6**, 626–634.

Received on November 7, 2000; accepted on February 14, 2001