NEWS RELEASE
Friday August 19, 2005

Queensland scientists lead international study discovering Genes Behind Endometriosis

A team of international scientists headed by The Queensland Institute of Medical Research (QIMR) has made a major genetic discovery in unlocking some of the mysteries of endometriosis.

Endometriosis, which causes up to 50% of female infertility, is a gynaecological condition where endometrium-tissue (normally found lining the uterus) actually grows outside the uterus and on pelvic organs. The condition disables women in their reproductive years and causes pelvic pain, menstrual disturbance, scarring and tissue damage, and fertility problems.

It has long been thought to have a familial or hereditary basis and the QIMR study, which has been going for more than 10 years and was co-founded with Brisbane gynaecologist Dr. Daniel O’Connor, revealed chromosome 10 as containing the area in which the most significant genes for endometriosis are located.

“Now that we have pin-pointed the chromosomal area, we can start to examine which genes on the chromosome are actually responsible for this disease, which does run in families and has a significant impact on health and reproduction,” said Chief Investigator, Dr Susan Treloar from QIMR’s Genetic Epidemiology Group.

The major study, which was published in the *American Journal of Human Genetics*, achieved one of the largest collections of sibling pairs (sisters in this case) for a complex disease that has ever been achieved anywhere in the world. In total, around 4,000 women with surgically confirmed endometriosis have participated in the study.

“Apart from taking DNA from the women with endometriosis, we also included their parents and other family members – bringing the total to around 10,000 people across Australia. The extracted DNA samples are now stored in our freezers, and our team led by Dr. Grant Montgomery is using them for different tests of association with endometriosis,” said Dr Treloar.

The first four pages of the paper follow this media release, issued by QIMR for national Endometriosis Awareness week August 22-28 2005.
Genomewide Linkage Study in 1,176 Affected Sister Pair Families Identifies a Significant Susceptibility Locus for Endometriosis on Chromosome 10q26

Susan A. Treloar,1,4 Jacqueline Wicks,1,4 Dale R. Nyholt,1,4 Grant W. Montgomery,1,4 Melanie Bahlo,1,2 Vicki Smith,6 Gary Dawson,6 Ian J. Mackay,6 Daniel E. Weeks,7 Simon T. Bennett,6 Alisoun Carey,6 Kelly R. Ewen-White,3 David L. Duffy,1,4 Daniel T. O’Connor,9 David H. Barlow,8 Nicholas G. Martin,1,4 and Stephen H. Kennedy8

1Cooperative Research Centre for Discovery of Genes for Common Human Diseases, 2Walter and Eliza Hall Institute, and 3Australian Genome Research Facility, Melbourne; 4Queensland Institute of Medical Research and 5Queensland Endometriosis Research Institute, Brisbane, Australia; 6Oxagen, Abingdon, United Kingdom; 7Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh; and 8Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, United Kingdom

Endometriosis is a common gynecological disease that affects up to 10% of women in their reproductive years. It causes pelvic pain, severe dysmenorrhea, and subfertility. The disease is defined as the presence of tissue resembling endometrium in sites outside the uterus. Its cause remains uncertain despite >50 years of hypothesis-driven research, and thus the therapeutic options are limited. Disease predisposition is inherited as a complex genetic trait, which provides an alternative route to understanding the disease. We seek to identify susceptibility loci, using a positional-cloning approach that starts with linkage analysis to identify genomic regions likely to harbor these genes. We conducted a linkage study of 1,176 families (931 from an Australian group and 245 from a U.K. group), each with at least two members—mainly affected sister pairs—with surgically diagnosed disease. We have identified a region of significant linkage on chromosome 10q26 (maximum LOD score [MLS] of 3.09; genomewide \( P < 0.047 \)) and another region of suggestive linkage on chromosome 20p13 (MLS = 2.09). Minor peaks (with MLS > 1.0) were found on chromosomes 2, 6, 7, 8, 12, 14, 15, and 17. This is the first report of linkage to a major locus for endometriosis. The findings will facilitate discovery of novel positional genetic variants that influence the risk of developing this debilitating disease. Greater understanding of the aberrant cellular and molecular mechanisms involved in the etiology and pathophysiology of endometriosis should lead to better diagnostic methods and targeted treatments.

Introduction

Endometriosis (MIM 131200) is a common gynecological disease that causes pelvic pain, severe dysmenorrhea (painful periods), and subfertility. It is defined as the presence of tissue resembling endometrium in sites outside the uterus, most commonly the pelvic peritoneum, ovaries, and rectovaginal septum (Giudice and Kao 2004). The main pathological processes associated with the disease are peritoneal inflammation and fibrosis and the formation of adhesions and ovarian cysts.

The diagnosis is usually made by visual inspection of the pelvis at laparoscopy, because noninvasive diagnostic tools, such as ultrasound scanning, can reliably detect only severe forms of the disease—that is, ovarian endometriotic cysts. Therefore, the population prevalence is difficult to measure. The best estimates indicate that endometriosis affects 8%–10% of women in their reproductive years (Eskenazi and Warner 1997), and it has been suggested that, in North America alone, >5.5 million women are affected (National Institute of Child Health and Human Development 2002). These data are compatible with the prevalence estimate of 7.2% from our own study of a large community sample of Australian twins (Treloar et al. 1999).

The cause of endometriosis remains uncertain despite >50 years of hypothesis-driven research, and thus the therapeutic options are limited. However, there is now convincing evidence that the disease is inherited as a complex genetic trait (Simpson and Bischoff 2002; Kennedy 2003; Giudice and Kao 2004). Genetic factors accounted for 52% of the variation in liability to endometriosis in our Australian twin study (Treloar et al. 1999). Familial aggregation has been reported in humans (Kennedy et al. 1995; Stefansson et al. 2002) and...
with endometriosis was significantly higher than that identified for the women identified in a Icelandic population study, in nonhuman primates with spontaneous disease (Zondervan et al. 2004). In an Icelandic population study, the average kinship coefficient for the women identified with endometriosis was significantly higher than that calculated for 1,000 sets of 750 matched controls (Stefansson et al. 2002). The genetic relative-recurrence risk for sibs (λs) was estimated to be 2.34 in our Australian sample of twins and their families (Treloar et al. 1999).

The most widely accepted theory to explain endometriosis is that viable endometrial cells reach the peritoneal cavity through retrograde menstruation along the fallopian tubes (Sampson 1927). Some cells then adhere to the peritoneal surface and proliferate. However, it is well established that menstrual debris is present in the peritoneal cavity of 90% of menstruating women, suggesting that endometrial cells from only some women are capable of establishing ectopic endometrial implants. There are several possible explanations for such susceptibility, including differences in genetic predisposition, increased exposure to menstrual debris, abnormal eutopic endometrium, altered peritoneal environment, reduced immune surveillance, and increased angiogenic capacity (Healy et al. 1998; Viantier et al. 2001; Treloar et al. 2002; Varma et al. 2004).

Various functional candidates have been tested for association as disease-susceptibility genes, but many of these case-control studies have lacked power and/or adequate controls; the results have therefore been inconclusive (Zondervan et al. 2001). Several researchers have adopted a positional-cloning approach to identify loci for endometriosis (Kennedy 2003). One unpublished study reported modest linkage to an unidentified chromosome 10 candidate region in 32 Puerto Rican families (Flores et al. 2004). Linkage to two candidate genes—GSTM1 on chromosome 1p13 (Hadfield et al. 2001) and GALT (MIM 606999) on 9p13 (Stefansson et al. 2001)—has been excluded, but no positive linkage regions from genomewide mapping have been reported to date.

We recruited >1,000 families—mainly affected sister pair (ASP) families—for a positional-cloning approach in our International Endogene Study (Treloar et al. 2002), which resulted from the merger of two independent groups: the Australian Genes Behind Endometriosis Study and the United Kingdom–based, international Oxford Endometriosis Gene (OXEGENE) Study; this ensured that the combined resource had 80% power to detect loci of modest effect (λs = 1.3) (Treloar et al. 2002), which is consistent with current expectations for most complex diseases. Here, we report results from the genomewide linkage scan in 1,176 families, one of the largest ASP linkage studies conducted for any disease to date.

### Methods

#### Family and Sample Collection

From 1995 to 2002, the Australian and U.K. groups recruited affected families with the use of protocols that have been described in detail elsewhere (Treloar et al. 2002). All women classified as affected had surgically confirmed endometriosis. In both studies, disease severity was assessed retrospectively from medical records by use of the revised American Fertility Society (rAFS) classification system (American Fertility Society 1985), which assigns patients to one of four stages (I–IV) on the basis of the extent of the disease and the associated adhesions present. The study families mainly comprised ASPs, although parents and other affected relatives were also recruited. If one or both parents were unavailable, sibs were recruited to increase the identity-by-descent (IBD) information, but they were assigned “unknown” disease status. There were three or more affected sisters in 104 (9%) of the sibships, which added power for linkage detection. The final data set comprised 1,176 families (931 from the Australia group and 245 from the U.K. group), with 1,242 independent affected sibships (i.e., a sibship of size S is equivalent to S – 1 independent sib pairs [Suarez and Hodge 1979]) that included two or more siblings. The mean number of affected family members in both Australian and U.K. kindreds was 2.3. Their ethnic backgrounds were similar, and >95% of subjects were white. Affected women completed a questionnaire about their pain symptoms and fertility history (table 1).
Approval for obtainment of medical records and collection of blood for DNA extraction and for all questionnaires and interview schedules was obtained from the Human Research Ethics Committee of the Queensland Institute of Medical Research. In the United Kingdom, the study received approval from the regional Multi-Centre Research Ethics Committee, from local research ethics committees, and from collaborating centers in Leuven and Dublin. All participants gave written informed consent.

Blood Sample Collection and Storage

In both studies, DNA was extracted from peripheral blood lymphocytes (Miller et al. 1988) or from buccal swabs with the use of Microcon Centrifugal Filter Devices (Amicons) and was stored at 4°C at a concentration of 300–600 μg/ml.

Genome Scans

In total, 4,985 individuals were genotyped, including 2,709 women with endometriosis. The genome scans for the Australian families were performed by the Australian Genome Research Facility (Ewen et al. 2000) and by Oxagen, United Kingdom. Both groups used the 400 dinucleotide microsatellite markers from Linkage Mapping Set version 2 (LMSV2 [PE Biosystems]) to provide ∼10-cM coverage of the genome on the autosomes and the X chromosome. To save cost and time, the final 79 Australian families were genotyped using only the 113 markers on chromosomes 9, 10, 11, 19, 20, 21, 22, and X. The U.K. study genotyped a small number of additional microsatellite markers from the ABI Prism LMS +HD5 set (Applied Biosystems).

At both sites, markers were amplified individually and were later combined into appropriate electrophoretic running panels, with up to 20 PCR products combined for one panel of markers (LMSV2 [PE Biosystems]). The pooled PCR products were electrophoresed through polyacrylamide gels on a PE Biosystems 377 (Australia) or 3700 (United Kingdom) Automated Sequencer, with the use of the recommended gel conditions and run protocols. PszI–restriction-digested lambda-phage DNA labeled with 6-carboxy rhodamine (GS500-ROX [PE Biosystems]) was included in each lane as a size standard (Ewen et al. 2000).

Combining Genotype Data from the Two Studies

Linkage analyses used allele frequencies calculated from pedigree founders by use of Sib-Pair (Duffy 2001). One approach to combining the data for common markers in different studies is to pool the genotypes, but doing so for samples typed at different facilities can be problematic because of allele binning and frequency differences. To avoid such difficulties, markers common to both studies were treated as different markers located at almost the same location (0.01 cM between them), which thereby kept the data from the two studies separate by a method we called “merging.” Hence, the separate allele frequencies were maintained—that is, allele frequencies were based on the Australian sample for the Australian pedigrees and were based on the U.K. sample for the U.K. pedigrees. Analyses of the markers for one chromosome with this merging method and with the much more labor-intensive pooling of alleles between studies produced near-identical linkage results, indicating that this is a valid approach.

Data Integrity

In both studies, sibling and family relationships were confirmed using GRR (which gives a graphical representation of relationships) (Abecasis et al. 2001). Data were cleaned prior to analysis of Mendelian inconsistencies with the use of Sib-Pair (Duffy 2001), PedCheck (O’Connell and Weeks 1998), and MERLIN (Abecasis et al. 2002). If a Mendelian inconsistency was detected, data for that marker were dropped for all family members. Two families common to both data sets (which occurred because the U.K. group recruited a small number of families in Australia) were removed from the Australian data set prior to analysis of combined data.

Statistical Analysis

For linkage analyses, we used a version of GENEHUNTER (v2.1) (Kruglyak et al. 1996) recompiled to handle the very large numbers of families. The statistic chosen for the analysis was the “possible triangle” maximum LOD score (MLS) that restricts maximization to the set of possible IBD-sharing probabilities for ASPs (Holmans 1993), where sibships of size S were weighted to be equivalent to S − 1 independent sib pairs. This approach has been shown to be the most powerful use of available data that maintained the correct type I error rate (Suarez and Hodge 1979; Davis and Weeks 1997). We opted for a nonparametric statistic because (1) there was no a priori reason to assume a particular disease model and (2) the assignment of the status “unaffected” is problematic (because a surgical procedure is required to exclude endometriosis). The MLS statistic was chosen because it has more-consistent power across disease models in ASP studies than do single-parameter statistics such as the nonparametric linkage (NPL) score (Cordell 2004). However, the U.K. data set contained several non-ASP affected relatives (ARPs). Huang and Vieland (2001) recently showed that the possible triangle MLS statistic is approximately equivalent to the ordinary parametric heterogeneity LOD (HLOD) score under the assumption of a simple recessive model (Huang and Vieland 2001). Therefore, we also compared results ob-
tained from parametric HLOD scores calculated under a simple (i.e., with no phenocopies) recessive mode of inheritance (which yields HLOD-R scores) (Greenberg et al. 1998), assuming 50% penetrance (Hodge et al. 1997). As recommended by Pal et al. (2001), we specified a disease gene frequency of 0.1. The X chromosome was analyzed using an earlier version of the GENEHUNTER program (X-GENEHUNTER) because the chromosome is not included in later versions. In the final analyses, we used published deCODE map positions for microsatellite markers when available and used Marshfield sex-averaged positions when deCODE positions were not available.

In accordance with the now widely accepted practice (Abecasis et al. 2004a, 2004b; Song et al. 2004) to obtain empirical estimates of genomewide significance levels, simulations were performed using the pedigree structures in the data set and any missing genotypes. These simulations allow us to take account of uneven marker spacing and informativeness (see the work of Kruglyak and Daly [1998] for a discussion of the utility of empirical significance levels in linkage analysis). Simulations were run on the data prior to the addition of the fine markers but after the addition of the last 79 families. Data for 1,000 genome scans were generated using MERLIN (Abecasis et al. 2002), under the assumption of no susceptibility loci, and were analyzed using our modified version of GENEHUNTER. The empirical significance level of an MLS peak was then estimated by counting the proportion of genome scans containing one or more peaks of that size. The cutoff for suggestive linkage (MLS = 1.88) was calculated as the mean of the genomewide MLS from each genome scan, which determines the maximum peak size expected once per genome scan by chance alone. The significant linkage threshold (MLS = 3.08) was defined as the MLS occurring with probability 0.05 in a genome scan (i.e., 50 peaks of equal or greater size observed in the 1,000 simulations).

Stratified analyses of the combined genome scan data were performed for stage of disease, age at symptom onset, age at surgical diagnosis, presence of pelvic pain, and problems in conceiving (subfertility). The number of affected women per family (zero, one, or more than one) who had the more extreme phenotype was the general criterion used to allocate families to one of three strata. For stage of disease, the three disease-severity strata were defined by the presence of zero, one, or more than one individual per family who had stage B (rAFS stages III–IV) endometriosis. Families were stratified by age at onset and age at diagnosis by use of the thresholds <20 years and <27 years of age, respectively. For subfertility, the relevant subphenotype for defining three strata was the reported lifetime history of any problem in conceiving. Since pelvic pain was so prevalent, we defined only two strata—zero or one family member and two or more family members reporting pelvic pain. To assess the significance of evidence for linkage in a phenotypic subset, compared with that in the entire sample, we randomly chose from the total sample the same number of families (1,000 replicates) in the subset of interest and calculated the MLS under the null hypothesis, which is analogous to the empirical estimation of genomewide significance levels.

Results

For the final merged Australian and U.K. data, the highest MLSs were on chromosome 10q (3.09 at D10S587) and chromosome 20p (2.09 at D20S889/D20S116). In our simulation study, 47 of the 1,000 replicates had peaks ¥3.09, and 360 had peaks ¥2.09, which implied significant linkage on chromosome 10 (genomewide empirical P = .047) and suggestive linkage on chromosome 20. The linkage (MLS) peaks for all chromosomes are shown in figure 1, and table 2 gives the positions and nearest markers for all MLS peaks ¥1.0.

To better define these peaks, we genotyped four additional markers under each of the chromosome 10 and 20 peaks in the Australian families. One of these markers (D20S116) had already been typed in the U.K. families. These markers changed the shape of the peaks but had minimal effects on the peak MLS: it increased from 3.09 to 3.16 for chromosome 10 and decreased from 2.09 to 1.86 for chromosome 20. The MLS curves are shown in figure 2. The summit for the chromosome 20 region moved marginally, from 8.49 cM to 12.09 cM, and the location of the chromosome 10 peak shifted from 145.44 cM to 148.75 cM—that is, from D10S387 toward D10S1656 in the telomeric direction. Both Australian and U.K. families contributed to the chromosome 10 linkage, with respective peaks of MLS = 2.06 and MLS = 1.48 when data were analyzed separately (fig. 3A). Both sets of families also contributed to the chromosome 20 peak (fig. 3B). Additional markers are shown in figures 2 and 3.

Given the large sample size from which our linkage statistics were calculated, conversion of the distribution of the likelihood-ratio test statistic of linkage versus no linkage into a $\chi^2$ distribution will be asymptotically valid (Walling et al. 2000). An MLS of 1.0 (which is a mixture of $\chi^2$ distributions with 0, 1, and 2 df) corresponds to an asymptotic (two-sided) $\chi^2$ with a P value of .05 (Nyholt 2000). The end points of the 95% CI are calculated by finding the locations on either side of the 148.75 cM peak on chromosome 10 that have an MLS of 2.16 (i.e., 1.0 less than the peak at 3.16). On the basis of our fine-mapping results, the 95% CI for the peak at 148.75 cM spans 15 cM, from 139.49 (10q26.11) to 154.77 cM (10q23.33).