

Heterogeneity-based genome search meta-analysis for preeclampsia

Elias Zintzaras · Georgios Kitsios · Gavan A. Harrison · Hannele Laivuori ·
Katja Kivinen · Juha Kere · Ioannis Messinis · Ioannis Stefanidis ·
John P. A. Ioannidis

Received: 4 April 2006 / Accepted: 18 May 2006 / Published online: 26 July 2006
© Springer-Verlag 2006

Abstract Preeclampsia is a pregnancy-related disorder that causes maternal and fetal morbidity and mortality. Its exact inheritance pattern is still unknown, and genome searches for identifying susceptibility loci for preeclampsia have thus far produced inconclusive or inconsistent results. We performed a heterogeneity-based genome search meta-analysis (HEGESMA) that synthesized the available genome scan data on preeclampsia. HEGESMA identifies genetic regions (bins) that rank highly on average in terms of linkage statistics across genome scans (searches). The significance of each bin's average rank and heterogeneity across scans was calculated using Monte Carlo tests. The meta-analysis involved four genome-scans on general preeclampsia and five scans on severe preeclampsia. In general

preeclampsia, 13 bins had significantly high average rank ($P_{\text{rank}} < 0.05$) by either unweighted or weighted analyses, while four of them (2p11.2–2q21.1, 9q21.32–9q31.2, 2p15–2p11.2, 2q32.1–2q35) were formally significant by both analyses. Heterogeneity of bin 2.8 (2q32.1–2q35) was significantly low in both unweighted and weighted analysis ($P_Q < 0.01$). In severe preeclampsia, 10 bins had significantly high average rank by either unweighted or weighted analyses and five of them (3q11.1–3q21.2, 2q37.1–2q37.3, 18p11.32–18p11.22, 2p15–2p11.2, 7q34–7q36.3) were significant by both analyses. Bin 2q37.1–2q37.3 showed marginal low heterogeneity in unweighted and weighted analysis ($P_Q = 0.06$). Results should be interpreted with caution as the p values were modest. Further investigation of

E. Zintzaras (✉) · G. Kitsios
Department of Biomathematics,
University of Thessaly School of Medicine,
Papakyriazi 22, Larissa 41222, Greece
e-mail: zintza@med.uth.gr

G. A. Harrison
Division of Environmental and Life Sciences,
Macquarie University, NSW 2109 Macquarie, Australia

H. Laivuori · J. Kere
Department of Medical Genetics, University of Helsinki,
00014 Helsinki, Finland

H. Laivuori
Department of Clinical Genetics, HUSLAB,
University of Helsinki, 00014 Helsinki, Finland

K. Kivinen · J. Kere
Department of Biosciences at Novum,
Karolinska Institute, Huddinge 14157, Sweden

I. Messinis
Department of Obstetrics and Gynaecology,
University of Thessaly School of Medicine,
Larissa 41222, Greece

I. Stefanidis
Department of Nephrology,
University of Thessaly School of Medicine,
Larissa 41222, Greece

J. P. A. Ioannidis
Department of Hygiene and Epidemiology,
University of Ioannina School of Medicine,
Ioannina 45110, Greece

J. P. A. Ioannidis
Department of Medicine,
Tufts University School of Medicine,
Boston 02111, USA

these regions by genotyping with additional markers and families may help to direct the identification of candidate genes for preeclampsia.

Introduction

Preeclampsia is a major pregnancy-related disorder that causes maternal and fetal morbidity and mortality (Lachmeijer et al. 2002a). Family-based studies have shown that genetic factors play a role in preeclampsia but the exact inheritance pattern is still unknown. Candidate gene studies on preeclampsia have produced inconsistent results so far. As an alternative to the single-gene approach, genome search can provide useful information on chromosomal loci that may be linked to complex diseases (Risch 1990). However, genome scans have also produced inconclusive inferences because linkage signals tend to be rather weak, the number of families and affected sibpairs are relatively small, and the individual genome scans identify linkage in different chromosomal regions (Wise et al. 1999; Dempfle et al. 2004). In an effort to generate more conclusive evidence, several investigators have proposed methods for summing the data from diverse genome scans through meta-analysis (Wise et al. 1999; Fisher et al. 2003).

Genome search meta-analysis (GSMA) is one of the best established methods for meta-analysis of such data (Wise et al. 1999; Levinson et al. 2003; Zintzaras et al. 2005a). GSMA, has already been applied to genome scans of several diseases (Fisher et al. 2003; Levinson et al. 2003; Chiodini et al. 2003; Demenais et al. 2003; Lewis et al. 2003; Segurado et al. 2003; van Heel et al. 2004; Koivukoski et al. 2004; Trikalinos et al. 2005). The methodology has been extended for evaluating also heterogeneity between different genome scans in the context of GSMA (HEGESMA: heterogeneity-based genome-scan meta-analysis). Heterogeneity metrics may offer additional useful insights besides summary estimates (Lau et al. 1998; Engels et al. 2000). In the present meta-analysis, these two methods were applied to the available genome scan data on preeclampsia (Zintzaras and Ioannidis 2005a, b).

Materials and methods

Eligible genome scans

We considered all preeclampsia genome scans published until December 2005. The studies were identified by search of the PubMed database. We used the combination of the following terms as the search criterion: “genome search” or “genome scan” and

“preeclampsia”. Scans restricted to specific individual chromosomes were excluded.

Databases

The following information was extracted for each suitable study: first author, journal, year of publication, country of recruitment, racial descent of study population, criteria of disease diagnosis, number of families, number of affected sibpairs, number of microsatellite markers, linkage statistic, software of linkage analysis. The corresponding investigators of each genome scan were then contacted and asked to provide high quality data figures and details on each genome scan so that information could be standardized similarly across genome scans. Data were separately extracted for two phenotypes: generally-defined preeclampsia (“general preeclampsia”, including gestational hypertension, proteinuric preeclampsia, and eclampsia); and severe preeclampsia, defined to include only proteinuric preeclampsia, and eclampsia.

Genome search meta-analysis and heterogeneity testing

The GSMA starts by splitting the chromosomes into bins of approximately equal length; usually each bin has a width of 30 centiMorgan (cM) giving 120 bins in total for the whole genome (Wise et al. 1999; Levinson et al. 2003; Zintzaras and Ioannidis 2005a). For nomenclature purposes, bin $c.n$ is the n th bin on chromosome c . For each genome scan, the most significant result of the test statistic obtained within the bin is recorded. Then, for each scan the bins are ranked according to their significance of results and the ranks for each bin are summed across scans. The significance of the average rank of each bin is assessed empirically against the distribution of average ranks. Under the null hypothesis of no linkage in any chromosomal bin, the ranks are randomly assigned from each study, then, the probability that the ranks X_i (where $i = 1 - m$ studies) from a specific bin sum to R is:

$$P\left(\sum_{i=1}^m X_i = R\right) = 0 \quad \text{for } R < m,$$

$$P\left(\sum_{i=1}^m X_i = R\right) = \frac{1}{n^m} \sum_{k=0}^{\text{int}[(R-m)/n]} (-1)^k \binom{R - KN - 1}{M - 1} \binom{m}{k} \quad \text{for } m = R = nm,$$

$$P\left(\sum_{i=1}^m X_i = R\right) = 0 \quad \text{for } R > nm.$$

When a bin has a high summed rank, then this is considered as evidence for linkage. If several bins within a study had equal test statistics, they were assigned tied ranks, and negative linkage scores were ranked as zero.

The Q statistic measures heterogeneity in the ranks between studies for each bin. It is the sum of the squared deviations of each study's bin rank from the mean of ranks:

$$Q = \sum_{i=1}^m w_i (R_i - \bar{R})^2,$$

where R_i is the rank of the bin under investigation for study i ($i = 1 - m$ studies), \bar{R} is the mean rank across the available studies, and w_i is the weighting factor for study i . Significantly low between-study heterogeneity is an indication that the results of different studies for the same bin are very consistent among themselves. The presence of low heterogeneity for a specific bin with high ranks in all studies can be interpreted as further supportive evidence for the importance of this bin. Conversely, significantly high between-study heterogeneity would suggest that there may be major differences in genetic effect, design, population, or other factors across the combined studies. Besides the Q statistic, two other heterogeneity metrics (B and Ha) have been proposed, but inferences with these metrics were similar to Q metric (not shown in detail) (Zintzaras and Ioannidis 2005a).

The statistical significance of the average rank and the Q metric were assessed using a Monte Carlo method. The ranks of each study are randomly permuted and the simulated average rank and Q metric is calculated; then the procedure is repeated 50,000 times and a null distribution for the average rank and the Q metric is constructed. The significance level (P_{rank}) of the average rank of bins is the percentage of simulated average ranks greater than or equal to the observed. The statistical significance level (P_Q) for low heterogeneity is the percentage of simulated metrics less than the observed (left-sided p value) (Zintzaras and Ioannidis 2005a). For high between-study heterogeneity, we examined the right-sided p values. Moreover, we performed a Monte Carlo test that generates null distributions separately for each bin, considering only the simulated distributions of the Q metric (Q_{adjusted}) for

bins having similar simulated average ranks (± 2) as the bin being considered each time (Zintzaras and Ioannidis 2005a). This analysis takes into account that the Q statistic may be influenced by the average rank.

Finally, we calculated the probability of observing a given average rank for a bin by chance in bins with the same “place” in the ascending order of average ranks in the runs (ordered ranks) (P_{order}) (Levinson et al. 2003). P_{rank} assesses the significance of each bin independently, whereas P_{order} is based on the distribution of average ranks across all bins simultaneously (Fisher et al. 2003).

We performed both unweighted and weighted analyses. In weighted analysis, the ranks of the bins in each study were weighted by

$$w_i = \sqrt{[(\text{number of individuals})_i \times (\text{number of markers})_i]}$$

where i denotes the study, and the weights were scaled to sum up to one.

The main analysis considered only genome scans where all bins were represented. When a scan had evaluated all chromosomes but with fewer markers than those required to generate information for all 120 bins, then, we considered this scan only in a secondary analysis. The ranks of this scan were adjusted to give a maximum rank of 120 and the meta-analysis was then restricted only to common bins represented across all scans (referred to as “conditional” analysis).

Analyses used the HEGESMA software (<http://www.biomath.med.uth.gr>) (Zintzaras and Ioannidis 2005b). Nominal significance corresponds to one-sided p values < 0.05 that are presented here without any adjustment for multiple comparisons.

Results

Databases

The literature review identified 17 titles in PubMed. The full articles of the retrieved studies were read to assess their suitability for meta-analysis. Five studies met the inclusion criteria (Harrison et al. 1997; Lachmeijer et al. 2001; Arngrimsson et al. 1998; Moses et al. 2000; Laivuori et al. 2003). All studies used similar diagnosis criteria for general preeclampsia and severe preeclampsia, apart from the study by Lachmeijer et al. (2001), which was included only in the analysis for severe preeclampsia. In two studies (Harrison et al. 1997; Laivuori et al. 2003), the authors provided LOD scores and NPL scores, respectively,

for each marker used in the genome scans. In the remaining studies, the data were extracted from the published graphs showing linkage scores per distance for all chromosomes. This information was also occasionally complemented by specific data on observed peaks presented in tables and/or text. Graphs were electronically digitized using appropriate software (Engauge Digitizer, ver 2.12, Mark Mitchell, 2002). In the study by Arngrimsson et al. (1998), we extracted the data from the genome scan that includes two families with a large effect to the LOD score at the 2p13 locus and this may cause some bias towards this locus.

Details on the analyzed studies are shown in Table 1. Briefly, the Finnish study consisted of 381 microsatellite markers genotyped in 15 families with 50 affected cases with severe preeclampsia and 63 cases with general preeclampsia (Laivuori et al. 2003). In the Australian and New Zealand study, the genome scan was performed by genotyping 400 microsatellite markers in 34 families with 87 affected cases for severe preeclampsia and 121 cases for general preeclampsia (Moses et al. 2000). In the Icelandic study for severe preeclampsia, 72 families containing 186 cases were evaluated using 440 microsatellite markers (Arngrimsson et al. 1998). For general preeclampsia, 124 families containing 343 cases were analyzed. All of the above-mentioned studies had used multi-point analysis to assess genome scans. In the Dutch study, the genome scan was performed with 293 microsatellite markers in 38 families with 77 cases for severe defined preeclampsia only (Lachmeijer et al. 2001). Finally, the Australian study consisted of 15 families with 40 and 50 severe and general preeclampsia cases, respectively (Harrison et al. 1997). This study used two-point analysis and only 90 markers, thus only 57 of the 120 bins were represented. A few families in this study partly overlapped with the larger Australian–New Zealand study. The exact number of families that were possibly double-counted was not clarified, the data from the Harrison et al. (1997). Therefore, the data of the smaller Australian study were only considered in the secondary “conditional” analysis. All genome scans have been conducted by considering the inheritance of the maternal genotype. The genetic and chromosomal location of the markers used in the original publications were not re-assigned.

LOD score was used as linkage statistics in all studies, except the Finnish study where NPL score was used instead. The Australian study had the least weight with a weight factor equal to $w = 0.08$ or 0.07 for general or severe preeclampsia, respectively, and the Iceland study had the most weight ($w = 0.46$ or 0.34 for

general or severe preeclampsia, respectively). The other studies had weights ranging from 0.18 to 0.26. The chromosome regions with suggestive linkage identified from each individual genome-scan are shown in Table 1.

General preeclampsia

Figure 1 shows the average ranks for each bin from three general preeclampsia genome scans, (Arngrimsson et al. 1998; Moses et al. 2000; Laivuori et al. 2003) and Fig. 2 shows the respective ranks from four severe preeclampsia genome scans (Lachmeijer et al. 2001; Arngrimsson et al. 1998; Moses et al. 2000; Laivuori et al. 2003). The bins with significant $P_{\text{rank}} < 0.05$ in unweighted or weighted analysis are indicated on the figure plots. The significant bins in the main and conditional analyses, the observed ranks, the GSMA and heterogeneity statistics for each study are shown in Table 2.

In general preeclampsia, 12 bins were found to have $P_{\text{rank}} < 0.05$ by either unweighted or weighted analyses; four of them (bins 2.5, 9.4, 2.4, 2.8) had $P_{\text{rank}} < 0.05$ with both methods (Table 2). Conditional analysis revealed an additional bin (bin 4.7). These bins were not significant in the order statistics for the unweighted or weighted analysis ($P_{\text{order}} > 0.05$). Bin 2.8 showed low heterogeneity that was statistically significant in both unweighted and weighted analysis ($P_Q = 0.006$ and 0.007 , respectively). Even when the Monte Carlo distributions were adjusted for the average rank of each bin, significantly low heterogeneity was seen in bin 2.8 for the unweighted and weighted analysis ($P_{Q\text{adjusted}} = 0.027$ and 0.032 , respectively). Thus, bin 2.8 provided evidence of linkage to preeclampsia in terms of both high average rank and significantly low heterogeneity between genome scans. Two bins (2.5 and 2.4) showed evidence for significantly high between study heterogeneity (right-sided $P_{Q\text{adjusted}} < 0.05$): these bins ranked at the top ranks in the Icelandic and Australian–New Zealand scans, but did not rank that high in the Finnish scan.

Severe preeclampsia

In the main analysis for severe preeclampsia, eight bins were found to have $P_{\text{rank}} \leq 0.05$ by either unweighted or weighted analyses and five of them (bins 3.5, 2.10, 18.1, 2.4, 7.6) had $P_{\text{rank}} < 0.05$ with both methods (Table 3). Another two bins (bins 22.2, 4.7) were significant in the conditional analysis. In all bins, the order statistics was non-significant for the unweighted or weighted analysis ($P_{\text{order}} > 0.05$). Bin 2.10 produced

Table 1 Characteristics and major results of preeclampsia genome-scans

Reference	Country	Affected cases		Pedigrees		Markers (<i>n</i>)		Type of analysis	Linkage statistics	Chromosome location with suggestive linkage	
		General	Severe	General	Severe	General	Severe			General	Severe
Laivuori et al. (2003)	Finland	63	50	15	15	435	381	Multipoint	NPL	2p25, 9p13, 4q32, 9p11, 9q34, 15q11	2p25, 9p13, 4q32
Moses et al. (2000)	Australia–NZ	121	87	34	34	400	400	Multipoint	LOD	2q21.1–2q24.1	11q23–24, 2q21–24
Arngrimsson et al. (1998)	Iceland	343	186	124	72	440	440	Multipoint	LOD, NPL	2p13	2p13
Lachmeijer et al. (2001)	Netherlands	–	77	–	38	–	293	Multipoint	LOD	–	22q13.1, 10q22.1, 11q13, 18p11.2
Harrison et al. (1997)	Australia	50	40	15	15	90	90	Multipoint	LOD	4q	4q

marginally low heterogeneity in unweighted analysis ($P_Q = 0.056$). However, heterogeneity was not significantly low, when analyses were adjusted for average rank, $P_{Q\text{adjusted}} = 0.177$. No bins showed significantly high between study heterogeneity.

Correlation of results for the two phenotypes

Three bins (2.4, 2.6, 4.7) had significant average ranks for both general and severe preeclampsia, although this was not seen in both weighted and unweighted analyses for all of them. Analyses showed consistently significant high average ranks only for bin 2.4, regardless of the outcome definition and regardless of whether weighting was used or not. There was modestly high correlation of the average bin ranks between general and severe preeclampsia (Spearman's $r = 0.50$, $p < 0.00001$ based on unweighted analyses; $r = 0.48$, $p < 0.00001$ based on weighted analyses).

Discussion

The present meta-analysis has identified several chromosomal regions with some evidence for linkage for general and severe preeclampsia. However, only four (bins 2.5, 9.4, 2.4, 2.8) and five (bins 3.5, 2.10, 18.1, 2.4, 7.6) regions were formally significant in both unweighted and weighted analysis for general and severe preeclampsia, respectively. There was modest overlap in the regions (bin 2.4) identified for each phenotype suggesting that the two phenotypes are likely to share some common genetic basis, but may also have phenotype-specific genetic effects. However, the Icelandic study contributed mostly to the significance of bin 2.4 (2p13 locus). Heterogeneity testing revealed low heterogeneity at two regions (bin 2.8 for general preeclampsia and bin 2.10 for severe preeclampsia) identified by the HEGESMA in both weighted and unweighted analysis. Heterogeneity testing adjusting for average rank retained the formal significance for low between scan heterogeneity for bin 2.8 suggesting that results are very consistently pointing to this chromosomal region across diverse scans.

A recent paper by Lewis and Levinson (2006) showed that the Q metric can be underpowered for moderate sibling relative risks, while the restricted test (Q_{adjusted}) may provide a better power to detect low heterogeneity. However, more extensive simulation studies are required to draw definitive conclusions. In addition, the low power of heterogeneity statistics is well known phenomenon in meta-analysis (Lau et al. 1998).

Fig. 1 Unweighted (*open circle*) and weighted (*filled circle*) average ranks from three (Iceland, Australia–New Zealand, Finland) general preeclampsia genome-scans. Bins (*c.n: nth bin on chromosome c*) with significant $P_{\text{rank}} < 0.05$ in unweighted or weighted analysis are shown above the *solid line*

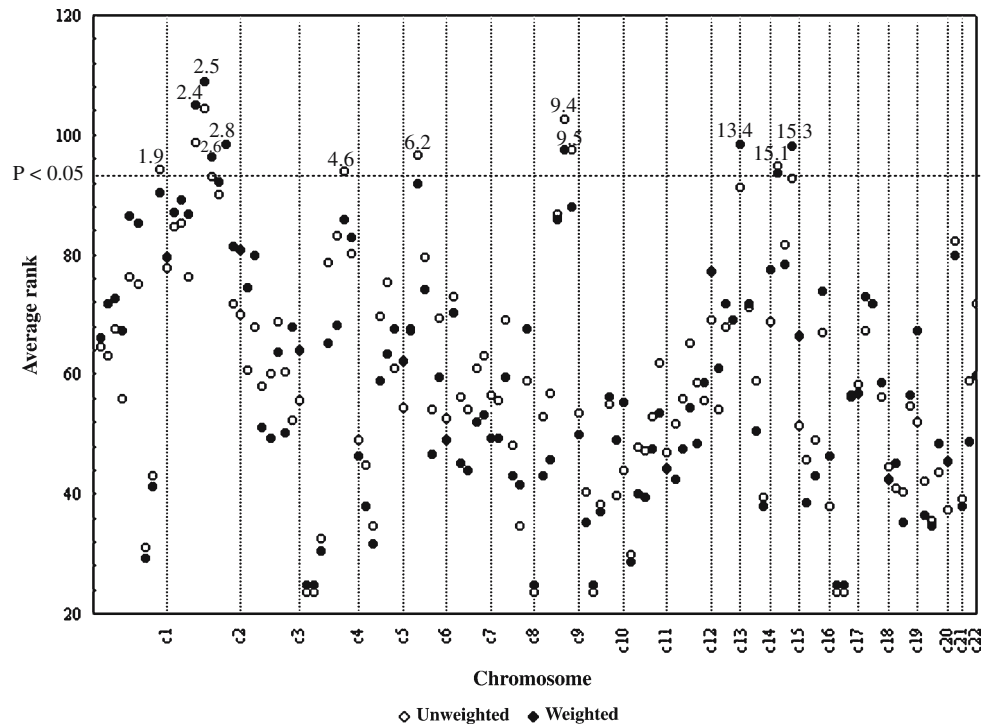
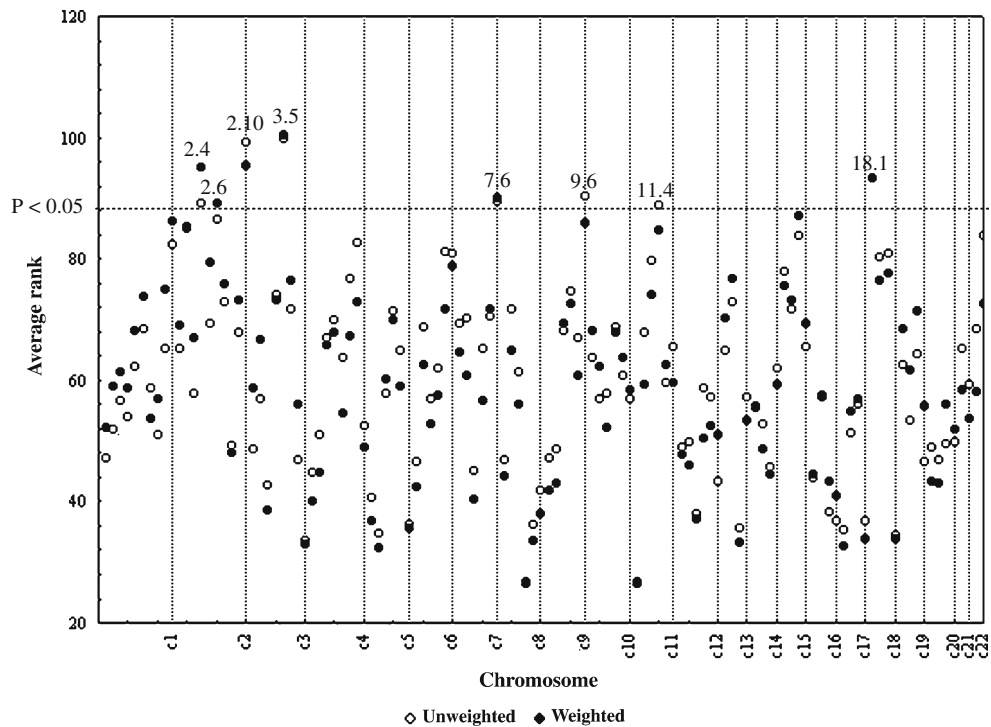


Fig. 2 Unweighted (*open circle*) and weighted (*filled circle*) average ranks from four (Iceland, Australia–New Zealand, Netherlands, Finland) severe preeclampsia genome-scans. Bins (*c.n: nth bin on chromosome c*) with significant $P_{\text{rank}} < 0.05$ in unweighted or weighted analysis are shown above the *solid line*



Nevertheless, quantification of heterogeneity is an important metric in a meta-analysis, and currently, HEGESMA is the only available approach for heterogeneity testing in GSMA.

For general preeclampsia, the present meta-analysis replicated the significance of seven chromosomal regions that were also reported as significant by the individual genome scans (2p11.2–2q21.1, 2p15–2p11.2,

Table 2 General preeclampsia genome-scan meta-analysis results in unweighted and (in parentheses) weighted analyses. Results based on three scans (4 scans on conditional analysis^b)

Bin	Cytogenetic location	Ranks in each genome scan			Average ranks and statistics			Low heterogeneity testing						
		Icel	Aust–NZ	Fin	^a Aust	Observed	P_{rank}	P_{rank}	P_{order}	P_{order}	Q	$Q_{adjusted}$	$Q_{adjusted}$	$Q_{adjusted}$
2.5	2p11.2–2q21.1	120	117	76.5	NA	104.5 (108.8)	0.010 (0.003)	NA	0.743 (0.371)	NA	0.331 (0.292)	0.964 (0.979)	NA	NA
9.4	9q21.32–9q31.2	88	105	115	82	102.6 (97.55)	0.015 (0.032)	0.012 (0.019)	0.559 (0.189)	0.162 (0.332)	0.141 (0.160)	0.442 (0.307)	0.359 (0.19)	0.085 (0.076)
2.4	2p15–2p11.2	119	119	58.5	116	98.83 (105.1)	0.025 (0.010)	0.003 (0.002)	0.643 (0.311)	0.212 (0.156)	0.579 (0.508)	0.986 (0.989)	0.997 (0.787)	0.382 (0.384)
2.8	2q32.1–2q35	103	96	96	27	98.33 (98.53)	0.029 (0.029)	0.118 (0.035)	0.467 (0.716)	0.190 (0.109)	0.006 (0.007)	0.027 (0.032)	0.434 (0.53)	0.592 (0.357)
9.5	9q31.2–9q34.11	63	118	112	NA	97.66 (88.05)	0.032 (0.095)	NA	0.322 (0.363)	NA	0.463 (0.562)	0.913 (0.643)	NA	NA
6.2	6p22.3–6p21.1	79	116	95	59	96.66 (91.93)	0.034 (0.065)	0.056 (0.054)	0.245 (0.231)	0.412 (0.169)	0.218 (0.251)	0.475 (0.365)	0.394 (0.287)	0.232 (0.182)
15.1	15q11.2–15q14	97	80	108	27	95 (93.58)	0.044 (0.054)	0.151 (0.063)	0.251 (0.321)	0.187 (0.095)	0.145 (0.129)	0.255 (0.158)	0.463 (0.484)	0.596 (0.343)
1.9	1q32.3–1q43	86	84	113	27	94.33 (90.24)	0.047 (0.078)	0.162 (0.087)	0.176 (0.314)	0.120 (0.063)	0.174 (0.159)	0.317 (0.173)	0.497 (0.395)	0.607 (0.300)
4.6	4q28.3–4q32.1	66.5	106	109	55	93.83 (85.82)	0.049 (0.119)	0.083 (0.103)	0.105 (0.195)	0.148 (0.072)	0.319 (0.387)	0.636 (0.431)	0.375 (0.343)	0.315 (0.272)
2.6	2q21.1–2q24.1	101	120	58.5	NA	93.16 (96.38)	0.053 (0.039)	NA	0.069 (0.160)	NA	0.495 (0.401)	0.888 (0.692)	NA	NA
15.3	15q22.32–15q26.1	110	112	56.5	NA	92.83 (98.22)	0.055 (0.029)	NA	0.036 (0.289)	NA	0.491 (0.428)	0.88 (0.795)	NA	NA
13.4	13q33.1–13q34	116	104	54	65	91.33 (98.45)	0.060 (0.030)	0.076 (0.023)	0.048 (0.483)	0.233 (0.133)	0.528 (0.499)	0.893 (0.871)	0.422 (0.771)	0.384 (0.506)
4.7	4q32.1–4q35.1	84	115	41.5	120	80.16 (82.91)	0.162 (0.151)	0.037 (0.085)	0.519 (0.302)	0.391 (0.150)	0.623 (0.509)	0.516 (0.517)	0.76 (0.619)	0.610 (0.473)

NZ New Zealand, Aust Australia, Icel Iceland, Fin Finland, NA non-applicable

^aAvailable data for 57 bins

^bResults from four scans based on 57 bins defined by Harrison et al. study (Australian study)

Table 3 Severe preeclampsia genome-scan meta-analysis results in unweighted and (in parentheses) weighted analyses. Results based on four scans (5 scans on conditional analysis^b)

Bin	Cytogenetic location	Ranks in each genome scan					Average ranks and respective statistics					Testing for low heterogeneity				
		Icel	Aust-NZ	NL	Fin	^a Aust	Observed	P_{rank}	^b P_{rank}	P_{order}	^b P_{order}	Q	^b Q	$Q_{adjusted}$	^b $Q_{adjusted}$	
3.5	3q11.1–3q21.2	116	77	89	117	NA	99.75 (100.5)	0.009 (0.006)	NA	0.701 (0.607)	NA	0.150 (0.156)	NA	0.68 (0.713)	NA	
2.10	2q37.1–2q37.3	84	96	109	108	NA	99.25 (95.34)	0.009 (0.019)	NA	0.361 (0.704)	NA	0.056 (0.064)	NA	0.177 (0.143)	NA	
18.1	18p11.32–18p11.22	99	91	117	66	NA	93.25 (93.24)	0.026 (0.027)	NA	0.642 (0.415)	NA	0.172 (0.134)	NA	0.487 (0.32)	NA	
9.6	9q34.11–9q34.3	70	90	107	95	28	90.5 (85.88)	0.037 (0.070)	0.116 (0.077)	0.714 (0.528)	0.066 (0.694)	0.091 (0.105)	0.398 (0.184)	0.148 (0.118)	0.299 (0.153)	
2.4	2p15–2p11.2	119	114	33.5	91	101	89.37 (95.04)	0.043 (0.021)	0.017 (0.012)	0.445 (0.453)	0.661 (0.537)	0.752 (0.699)	0.551 (0.608)	0.763 (0.915)	0.623 (0.853)	
7.6	7q34–7q36.3	104	68	86	100	76	89.5 (90.13)	0.044 (0.042)	0.038 (0.030)	0.633 (0.604)	0.158 (0.575)	0.101 (0.106)	0.047 (0.065)	0.154 (0.177)	0.037 (0.058)	
11.4	11q13.3 – 11q22.1	62	117	118	59	74	89 (84.65)	0.046 (0.081)	0.044 (0.060)	0.319 (0.413)	0.076 (0.747)	0.495 (0.497)	0.338 (0.425)	0.617 (0.501)	0.298 (0.402)	
2.6	2q21.1 – 2q24.1	97	119	33.5	97	NA	86.62 (89.24)	0.061 (0.048)	NA	0.486 (0.536)	NA	0.653 (0.544)	NA	0.593 (0.62)	NA	
22.2	22q12.3–22q13.3	26.5	82	120	107	103	83.87 (72.61)	0.085 (0.243)	0.033 (0.150)	0.619 (0.671)	0.295 (0.275)	0.829 (0.936)	0.687 (0.927)	0.853 (0.862)	0.753 (0.843)	
4.7	4q32.1–4q35.1	26.5	110	79	115	120	82.62 (73.06)	0.098 (0.235)	0.022 (0.129)	0.539 (0.788)	0.388 (0.684)	0.806 (0.932)	0.791 (0.955)	0.808 (0.838)	0.928 (0.902)	

NZ New Zealand, *Aust* Australia, *Icel* Iceland, *Fin* Finland; *NL* Netherlands, *NA* non-applicable

^aAvailable data for 57 bins

^bResults from five scans based on 57 bins defined by Harrison et al. (1997) study (Australian study)

9q31.2–9q34.11, 15q11.2–15q14, 4q28.3–4q32.1, 2q21.1–2q24.1, 4q32.1–4q35.1). Six novel candidate chromosomal regions for general preeclampsia were identified: 9q21.32–9q31.2, 2q32.1–2q35, 6p22.3–6p21.1, 1q32.3–1q43, 13q33.1–13q34 and 15q22.32–15q26.1. For severe preeclampsia, there was a replication for six already reported chromosomal regions: 18p11.32–18p11.22, 11q13.3–11q22.1, 2p15–2p11.2, 2q21.1–2q24.1, 22q12.3–22q13.3 and 4q32.1–4q35.1. Four new candidate regions were identified: 3q11.1–3q21.2, 2q37.1–2q37.3, 9q34.11–9q34.3 and 7q34–7q36.3.

Genes at several of the identified regions may confer susceptibility to disease. Bin 2.8 includes at least two genes investigated for susceptibility to preeclampsia: the INHA and CTLA-4 genes. CTLA-4 has been implicated as a candidate gene for severe preeclampsia in a recent study (Samsami Dehaghani et al. 2005). CTLA-4 is a well established candidate gene also for several diseases with autoimmune pathophysiology and in particular for diabetes mellitus which may share with preeclampsia a pro-thrombotic tendency and relationship to hypertension (Ueda et al. 2003; Kavvoura et al. 2005). Inhibin levels are affected in preeclampsia, but it is unclear whether polymorphisms of the INHA gene affect susceptibility to preeclampsia (Ciarmela et al. 2005). In bin 3.5, the region that had the highest average rank for severe preeclampsia, there is the AGT1R gene, while bin 2.5 contains the IL1B gene, both of them associated to preeclampsia in some studies (Bouba et al. 2003; Lachmeijer et al. 2002b). In bins 2.4 and 2.6 there are no known genes associated with preeclampsia. However, based on the results of the Iceland and Australian–New Zealand study, Fitzpatrick et al. have designated this 74 cM region, overlapping bins 2.4, 2.5 and 2.6 as PREG1 (preeclampsia/eclampsia gene 1) locus. Two positional candidate genes have been analyzed in bin 2.4 (TACR1, TCF7L1) (Fitzpatrick et al. 2004). Bin 1.9 is the region where various known candidate genes are located such as AGT, REN and MTR (Bouba et al. 2003; Also Rallo et al. 2005). Bin 13.4 is the location of factor VII gene promoter (F7), and bin 7.6 contains the NOS3 and TBXAS1 genes which may also be implicated in the development of preeclampsia (Laasanen et al. 2002; Bashford et al. 2001).

The populations involved in the meta-analysis are of Caucasian origin. Therefore any heterogeneity should be attributed more to genuine inconsistency of genetic effects and to differences in the studies' design and conduct, and less to ethnic descent differences across populations. However, we should acknowledge that Finnish and Icelandic populations may have some of the characteristics of genetic isolates.

The Dutch study, under the strict criteria, presented data for 67 families including the diagnosis of HELLP syndrome and for 38 families excluding HELLP syndrome. The authors of this study analysed the data separately, because in the previous scans no HELLP syndrome cases had been recruited (Lachmeijer et al. 2001). The most recent genome scan following the Dutch study is the Finnish study, which did not include HELLP syndrome cases as well (Laiivuori et al. 2003). Consequently, our meta-analysis included the data from the 38 families without HELLP syndrome.

When preeclampsia starts early, i.e. before week 32, it is frequently associated with intrauterine growth retardation due to deterioration of placental function. The potential placental contribution requires special consideration. The fetal gene load may also influence a mother's susceptibility to preeclampsia. A recent analysis indicated that both maternal and fetal genes play an etiological role and estimated the heritability conferred by maternal genes as 35% and that due to fetal genes as 20% (Pawitan et al. 2004). Another recent study (GOPEC trial) with tightly phenotyped mothers with preeclampsia together with suitable family structures (maternal and fetal trios) found that none of the genetic variants tested conferred a high risk of disease (GOPEC Consortium 2005). Even if the meta-analysis could be restricted to the early-onset, severe phenotype, preeclampsia is still likely to be very heterogenous, and placental contribution (through maternally or paternally inherited genes) could vary in different pregnancies. Most typically preeclampsia affects first pregnancies, but in familial preeclampsia it occurs also more often in subsequent pregnancies and the genetic contribution is likely to be stronger.

Conventionally HEGESMA is based on bins with width of 30 cM. An analysis with bin width less than 30 cM was not performed given that marker data would be sparse for this purpose. Moreover, decisions on multiple tests have not been made in this analysis because heterogeneity-based genome scan meta-analysis is an exploratory non-parametric procedure interested in the relative significance of the regions (Rothman 1990). The statistical significance values that we report are nominal and they should be interpreted with caution given the multiple bins tested, even if these bin results are correlated. Another limitation of the analysis is the use of the variable map density within and between studies (Lewis et al. 2003).

Given these limitations, we should caution that the proportion of type I and II errors may be considerable. Type II errors are expected to be less common in the meta-analysis than in single studies, but the total

evidence is not extensive enough to be deemed conclusive even in the combined database. It is possible that some important regions may have been missed, while some of the identified regions may result from type I errors. Genetic effects of single genes are likely to be small. Candidate gene studies highlight that single-gene effects may be variable across populations, may differ between generally-defined preeclampsia and severe preeclampsia or postulated effects may be subject to bias (Kosmas et al. 2003, 2004). Obtaining large-scale evidence with collaborative meta-analyses is a step towards bypassing these problems (Ioannidis et al. 2006).

In conclusion, allowing for these caveats, the genome scan meta-analysis provides some evidence of linkage for several regions in general and severe preeclampsia. Further investigation of these regions by genotyping with additional markers and families may direct the identification of candidate genes for association studies in preeclampsia.

References

- Also-Rallo E, Lopez-Quesada E, Urreiziti R, Vilaseca MA, Laila JM, Balcells S, Grinberg D (2005) Polymorphisms of genes involved in homocysteine metabolism in preeclampsia and in uncomplicated pregnancies. *Eur J Obstet Gynecol Reprod Biol* 120:45–52
- Arngrimsson R, Sigurdardottir S, Frigge ML, Bjarnadttir RI, Jonsson T, Stefansson H, Baldursdottir A, Einarsdottir AS, Palsson B, Snorraddottir S, Lachmeijer AM, Nicolae D, Kong A, Bragason BT, Gulcher JR, Geirsson RT, Stefansson K (1998) A genome-wide scan reveals a maternal susceptibility locus for pre-eclampsia on chromosome 2p13. *Hum Mol Genet* 8:1799–1805
- Bashford MT, Hefler LA, Vertrees TW, Roa BB, Gregg AR (2001) Angiotensinogen and endothelial nitric oxide synthase gene polymorphisms among hispanic patients with preeclampsia. *Am J Obstet Gynecol* 184:1345–1350
- Bouba I, Makrydimas G, Kalaitzidis R, Lolis DE, Siamopoulos KC, Georgiou I (2003) Interaction between the polymorphisms of the renin-angiotensin system in preeclampsia. *Eur J Obstet Gynecol Reprod Biol* 110:8–11
- Ciarmela P, Florio P, Battistini S, Grasso D, Amato T, Boschi S, Marozio L, Benedetto C, Petraglia F (2005) Mutational analysis of the inhibin alpha gene in preeclamptic women. *J Endocrinol Invest* 28:30–33
- Chiodini BD, Lewis CM (2003) Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. *Arterioscler Thromb Vasc Biol* 23:1863–1868
- Demenaif F, Kanninen T, Lindgren CM, Wiltshire S, Gaget S, Dandrieux C, Almgren P, Sjogren M, Hattersley A, Dina C, Tuomi T, McCarthy MI, Froguel P, Groop LC (2003) A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes. *Hum Mol Genet* 12:1865–1873
- Dempfle A, Loesgen S (2004) Meta-analysis of linkage studies for complex diseases: an overview of methods and a simulation study. *Ann Hum Genet* 68:69–83
- Engels EA, Schmid CH, Terrin N, Olkin I, Lau J (2000) Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 19:1707–1728
- Fisher SA, Lanchbury JS, Lewis CM (2003) Meta-analysis of four rheumatoid arthritis genome-wide linkage studies: confirmation of a susceptibility locus on chromosome 16. *Arthritis Rheum* 48:1200–1206
- Fitzpatrick E, Goring HH, Liu H, Borg A, Forrest S, Cooper DW, Brennecke SP, Moses EK (2004) Fine mapping and SNP analysis of positional candidates at the preeclampsia susceptibility locus (PREG1) on chromosome 2. *Hum Biol* 76:849–862
- GOPEC Consortium (2005) Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. *Am J Hum Genet* 77:127–131
- Harrison GA, Humphrey KE, Jones N, Badenhop R, Guo G, Elakis G, Kaye JA, Turner RJ, Grehan M, Wilton AN, Brennecke SP, Cooper DW (1997) A genomewide linkage study of preeclampsia/eclampsia reveals evidence for a candidate region on 4q. *Am J Hum Genet* 60:1158–1167
- Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ, Human Genome Epidemiology Network, the Network of Investigator Networks (2006) A road map for efficient and reliable human genome epidemiology. *Nat Genet* 38:3–5
- Kavvoura FK, Ioannidis JP (2005) CTLA-4 gene polymorphisms and susceptibility to type 1 diabetes mellitus: a HUGe Review and meta-analysis. *Am J Epidemiol* 162:3–16
- Koivukoski L, Fisher SA, Kanninen T, Lewis CM, von Wövern F, Hunt S, Kardia SLR, Levy D, Perola M, Rankinen T, Rao DC, Rice T, Thiel B, Melander O (2004) Meta-analysis of genome-wide scans for hypertension and blood pressure in Caucasians shows evidence of susceptibility regions on chromosomes 2 and 3. *Hum Mol Genet* 13:2325–2332
- Kosmas IP, Tatsioni A, Ioannidis JP (2003) Association of Leiden mutation in factor V gene with hypertension in pregnancy and pre-eclampsia: a meta-analysis. *J Hypertens* 21:1221–1228
- Kosmas IP, Tatsioni A, Ioannidis JP (2004) Association of C677T polymorphism in the methylenetetrahydrofolate reductase gene with hypertension in pregnancy and preeclampsia: a meta-analysis. *J Hypertens* 22:1655–1662
- Laasanen J, Hiltunen M, Punnonen K, Mannermaa A, Heinonen S (2002) Fibrinogen and factor VII promoter polymorphisms in women with preeclampsia. *Obstet Gynecol* 100:317–320
- Lachmeijer A, Arngrimsson R, Bastiaans EJ, Frigge ML, Pals G, Sigurdardottir S, Stefansson H, Palsson B, Nicolae D, Kong A, Aarnoudse JG, Gulcher JR, Dekker GA, ten Kate LP, Stefansson K (2001) A genome-wide scan for preeclampsia in The Netherlands. *Eur J Hum Genet* 9:758–764
- Lachmeijer A, Dekker G, Pals G, Aarnoudse JG, ten Kate LP, Arngrimsson R (2002a) Searching for preeclampsia genes: the current position. *Eur J Obstet Gynecol Reprod Biol* 105:94–113

- Lachmeijer A, Nosti-Escanilla MP, Bastiaans EB, Pals G, Sandkuijl LA, Kostense PJ, Aarnoudse JG, Crusius JB, Pena AS, Dekker GA, Arngrimsson R, ten Kate LP (2002b) Linkage and association studies of IL1B and IL1RN gene polymorphisms in preeclampsia. *Hypertens Pregnancy* 21:23–38
- Laivuori H, Lahermo P, Ollikainen V, Widen E, Haiva-Mallinen L, Sundstrom H, Laitinen T, Kaaja R, Ylikorkala O, Kere J (2003) Susceptibility Loci for Preeclampsia on Chromosomes 2p25 and 9p13 in Finnish Families. *Am J Hum Genet* 72:168–177
- Lau J, Ioannidis JP, Schmid CH (1998) Summing up evidence: one answer is not always enough. *Lancet* 351:123–127
- Levinson DF, Levinson MD, Segurado R, Lewis CM (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. *Am J Hum Genet* 73:17–33
- Lewis CM, Levinson DF (2006) Testing for genetic heterogeneity in the genome search meta-analysis method. *Genet Epidemiol* 30:348–355
- Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lonnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfsson J, Sigmundsson T, Petursson H, Jazin E, Zoega T, Helgason T (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 73:34–48
- Moses EK, Lade JA, Guo G, Wilton AN, Grehan M, Freed K, Borg A, Terwilliger JD, North R, Cooper DW, Brennecke SP (2000) A genome scan in families from Australia and New Zealand confirms the presence of a maternal susceptibility locus for pre-eclampsia, on chromosome 2. *Am J Hum Genet* 67:1581–1585
- Pawitan Y, Reilly M, Nilsson E, Cnattingius S, Lichtenstein P (2004) Estimation of genetic and environmental factors for binary traits using family data. *Stat Med* 23:449–465
- Risch N (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 46:242–253
- Rothman KJ (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–46
- Samsami Dehaghani A, Doroudchi M, Kalantari T, Pezeshki AM, Ghaderi A (2005) Heterozygosity in CTLA-4 gene and severe preeclampsia. *Int J Gynaecol Obstet* 88:19–24
- Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI, Craddock N, DePaulo JR, Baron M, Gershon ES, Ekholm J, Cichon S, Turecki G, Claes S, Kelsøe JR, Schofield PR, Badenhop RF, Morissette J, Coon H, Blackwood D, McInnes LA, Foroud T, Edenberg HJ, Reich T, Rice JP, Goate A, McInnis MG, McMahon FJ, Badner JA, Goldin LR, Bennett P, Willour VL, Zandi PP, Liu J, Gilliam C, Juo SH, Berrettini WH, Yoshikawa T, Peltonen L, Lonnqvist J, Nothen MM, Schumacher J, Windemuth C, Rietschel M, Propping P, Maier W, Alda M, Grof P, Rouleau GA, Del-Favero J, Van Broeckhoven C, Mendlewicz J, Adolfsson R, Spence MA, Luebbert H, Adams LJ, Donald JA, Mitchell PB, Barden N, Shink E, Byerley W, Muir W, Visscher PM, Macgregor S, Gurling H, Kalsi G, McQuillin A, Escamilla MA, Reus VI, Leon P, Freimer NB, Ewald H, Kruse TA, Mors O, Radhakrishna U, Blouin JL, Antonarakis SE, Akarsu N (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part III Bipolar disorder. *Am J Hum Genet* 73:49–62
- Trikalinos T, Karvouni A, Zintzaras E, Ylisaukko-Oja T, Peltonen L, Jarvela I, Ioannidis JP (2005) A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol Psych* 11:29–36
- Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Slink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511
- van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM (2004) Genome Scan Meta-Analysis Group of the IBD International Genetics Consortium. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 13:763–770
- Wise LH, Lanchbury JS, Lewis CM (1999) Meta-analysis of genome searches. *Ann Hum Genet* 63:263–272
- Zintzaras E, Ioannidis JP (2005a) Heterogeneity testing in meta-analysis of genome searches. *Genet Epidemiol* 28:123–137
- Zintzaras E, Ioannidis JP (2005b) HEGESMA: genome search meta-analysis and heterogeneity testing. *Bioinformatics* 21:3672–3673