LETTER TO JMG

Heterozygote excess is repeatedly observed in females at the BRCA2 locus N372H

M D Teare, A Cox, J Shorto, C Anderson, D T Bishop, C Cannings

....

E
vidence of genotype specific selection at the BRCA2 polymorphism N372H was originally reported by Healey et al.[1] These workers found significant evidence of a heterozygous excess in women control samples when studying the polymorphism for breast cancer association. They then genotyped a large series of newborn boys and girls to examine if this effect was also seen at birth. The newborn girls were consistent with the women but the newborn boys were significantly different. The genotypes in the boys appeared to demonstrate a significant deficit of heterozygotes. Healey *et al* suggested that sex differential viabilities resulted in a stable allele frequency, but this was not formally investigated. We wanted to explore further this apparent sex specific selection in two ways; to refine the estimates of fitness, and we examined the mathematical properties of the sex specific selection model to determine if the data were consistent with a stable equilibrium.

METHODS

Three control populations were available for additional genotyping. These were blood donors, colon cancer controls, and mammography screening control women. Anonymous blood donors were recruited from the Sheffield Blood Transfusion Service in 1996, and their ages and sexes were recorded. Blood DNA samples were obtained for the mammography screening series of controls from white women attending for routine mammography screening at the Sheffield Breast Screening Service between September 2000 and August 2002. Women's DNA samples were included in the study if the mammogram showed no evidence of a breast lesion. All women resident in Sheffield and on the Community Health Index, who are between the ages of 50 and 65, are invited to attend for screening every three years, and the uptake in Sheffield is 81.2%. Ethical approval for this study was obtained from the South Sheffield Research Ethics Committee and informed written consent was obtained from all subjects. Colon cancer controls were selected for the north of England and Scotland colon cancer study. An age and sex matched control was identified for each colon cancer patient in the study. A full description of the ascertainment of the controls (and cases) is described by Barrett et al.[2] We genotyped two male samples, 761 blood donors and 231 colon cancer controls, and three female groups, 389 blood donors, 188 colon cancer controls, and 957 mammography screening control women. The genotyping was carried out using the same protocol as specified in Healey et al.[1] We also searched the literature for further reports of this polymorphism with the objective of performing a full joint analysis.

Our literature search (for ''Arg372His'' or ''N372H'' or ''BRCA2 polymorphism'') identified four further publications. One of these[3] presented two female control populations. In a second study of three Chinese populations,[4] frequencies were not reported by sex. The third study[5] used

Key points

- Sex specific selection at the BRCA2 N372H locus has previously been suggested. We have examined the evidence for sex differential selection at this locus.
- Firstly, control samples have been genotyped and these, together with all published data on the N372H polymorphism, have been incorporated into a joint analysis to refine estimates of genotype viability.
- Secondly, we consider the properties of a mathematical model for sex differential selection, to ensure that the estimates are consistent with a stable polymorphism.
- We find viabilities which are consistent with a stable polymorphism.
- Examination of the mathematical model indicates that the dramatic heterozygote disadvantage originally reported in males would not result in a stable polymorphism.
- There is significant evidence consistent with a heterozygote advantage in females, and no evidence of genotype specific selection in males.

controls in a breast cancer study that had previously been reported by Auranen et al[3] in a study of ovarian cancer. The remaining study[6] reported previously unpublished frequencies for Japanese female controls.

Statistical methods and definition of terms

The aim of this study is to refine the estimate of the viabilities using a joint analysis. In the context of this work we use the terms ''fitness'' and ''viability'' to represent the same quantity, probability of survival. To estimate genotype viability a population must be observed from birth to some specific age. Viability is then estimated as the proportion of deaths occurring in that time frame. However, we do not have observations over generations. The samples available to us are all adult controls, who are essentially survivors. We can only estimate viability relative to some expectation. We assume that the population frequency of the polymorphism is at equilibrium and does not change from one generation to the next. We also expect that at conception or birth the genotype frequencies will be consistent with random mating in the adult population. ''Relative viability'' is the viability of one genotype relative to another (the ratio of the two viabilities). These relative viabilities can be estimated from the observed genotypes counts (for explicit formulae see Weir[7]). The relative viabilities reported in this analysis are relative to the heterozygote viability.

Our analysis will involve the use of a random effects model. Such a model imposes a structure in that it is assumed

that the true mean viability for each sampled population is itself sampled from an underlying distribution. Viewing the problem in this way enables us to make inferences about the mean and variance of this underlying distribution. This is effectively our main interest in this analysis, to see if the underlying relative viabilities are significantly different from 1.

These random effects models were fitted using WinBUGS 1.3.[8] Fitting such a model requires specification of prior distributions for all model parameters as follows:

The model assumes that each population genotype viability and population allele frequency has been drawn from an underlying distribution. In female populations let a_i denote the viability (relative to genotype NH) for genotype NN in population i, and let b_i denote the female viability for genotype HH in population i. Similarly c_j and d_j denote the viabilities for genotypes NN and HH relative to genotype NH in males in population j.

The prior distributions and basic model were specified as follows:

The prior distributions of the logarithms of the underlying relative viabilities a and c are as follows, with mean 0, and large variance:

- $\ln(a) \sim N(0, 1/\text{sig}_{ap})$; $\text{sig}_{ap} \sim \text{gamma}(1000, 1000)$
- $\ln(c) \sim N(0, 1/\text{sig}_{cp})$; $\text{sig}_{cp} \sim \text{gamma}(1000, 1000)$

Each population allele frequency is then sampled from the same beta distribution, with x and y sampled from uniform$(0, 1000)$:

- $p_i \sim \text{Be}(x, y)$

The natural logarithm of each population viability is sampled from a normal distribution:

- $\ln(a_i) \sim N(\ln(a), 1/\text{sig}_a)$; $\text{sig}_a \sim \text{gamma}(1000, 1000)$
- $\ln(c_i) \sim N(\ln(c), 1/\text{sig}_c)$; $\text{sig}_c \sim \text{gamma}(1000, 1000)$

RESULTS

The full set of results for this polymorphism, consisting of 12 female and 3 male control samples, is presented in table 1. Fig 1 displays the point estimates and 95% confidence intervals for the estimated viabilities of genotype NN relative to NH. We have estimated the viability assuming that the system is at equilibrium so that the allele frequency does not change from one generation to the next.[7] When both sexes have been studied in the same population, the allele frequencies are remarkably similar with more variation seen between than within samples. If the genotype frequencies observed were as predicted under Hardy-Weinberg equilibrium, then all the genotype viabilities would be close to or equal 1. Fig 1 shows that 10 of the 12 female populations have homozygote viability estimates of less than 1. Of the two additional male samples one shows a heterozygote excess while the other shows a weak heterozygote deficit. What is apparent from the figure is that not one of the female samples alone is statistically significant. Healy et al[1] were able to demonstrate that this heterozygote advantage was a consistent feature by performing a composite likelihood ratio test. The only viability that was significantly different from 1 in a single sample was that for newborn boys.

As there was no evidence that sex specific allele frequencies differ we assumed that allele frequency was the same in both sexes. Therefore the model consists of three random effects, the allele frequency, and the sex specific relative viability of genotype NN. The results of fitting such a random effects model can be seen in fig 2.

The change in the point estimates and confidence intervals for each sample population reflects the structure imposed by the random effects model. The confidence intervals for population specific relative viability are all slightly smaller. This is due to some of the variability now being accounted for by the underlying distribution. There is evidence that the underlying mean female viability for genotype NN, estimated to be 0.95, is significantly different from 1, (95% confidence interval, 0.91–0.99). However the estimate of 1.01 for the mean underlying fitness of NN in men implies no differential viability for men but the confidence interval is very wide (95% confidence interval, 0.89–1.12). The estimate for underlying population allele frequency was 0.73 (95% confidence interval, 0.71–0.74) and the relative viabilities for the genotype HH were 0.87 (95% confidence interval, 0.76–0.97) and 1.02 (95% confidence interval, 0.74–1.31) in women and men respectively.

These analyses have attempted to estimate the underlying sex specific viabilities assuming the system is at equilibrium, in that allele frequencies are not modified at each generation. The model used is in no way restricted so that parameters need to be consistent with a stable equilibrium. We therefore need to explore the parameter space analytically to identify the conditions under which a stable solution is achieved.

Mathematical analysis of conditions necessary for a stable polymorphism

In a simple selection model for a single autosomal locus in an infinite, randomly mating population with constant viabilities that are the same in two sexes, the behaviour is straightforward.[9] Population fitness will increase steadily. With heterozygote advantage, allele frequencies will be driven to a stable polymorphism, while heterozygote disadvantage will always lead to fixation of one or other allele.

In the case where there are different viabilities in the two sexes the dynamics are far more complex. Suppose that in generation t the frequencies of allele N are $p(t)$ and $r(t)$ in women and men respectively, (with $q(t) = 1 - p(t)$ and $s(t) = 1 - r(t)$ the frequencies of allele H). The viabilities of genotypes NN, NH, and HH are a, h, and b in women, and c, k, and d in men. We can rescale so that $h = 1$ and $k = 1$, which leaves essentially four parameters. In the particular case where there is no selection in one sex (for example, $c = k = d$) then it can be proved that there is a stable polymorphism if, and only if, $h > a, b$ and that the system will converge to it.[10] In general, the behaviour is more complex. It has been demonstrated[11] that there can be three internal equilibria either with one of these parameters stable and the others unstable or vice versa.

Depending on the specific values of the viabilities the patterns of stable and unstable equilibria fall into one of six types. Since if we list the equilibria in sequence of increasing p they must be alternately stable and unstable (assuming there are no neutral equilibria which can occur in pathological cases), we can represent the possible systems as a list of Us (unstable) and Ss (stable), with the first entry for $p = 0$ and the last for $p = 1$—that is, US($=$SU), USU, SUS, USUS($=$SUSU), USUSU, and SUSUS where the equalities just reflect swapping alleles N and H and do not create qualitatively different systems. The qualitative dynamic behaviour of $p(t)$ and $r(t)$ is shown in fig 3. It has been shown that the system will converge to an equilibrium, irrespective of the starting position.[12] In general that will be a stable equilibrium but convergence to an unstable equilibrium will occur if the initial allele frequencies are precisely on the separatrix (the lines which separate the basins of attraction for the stable equilibria). Which of the six pictures is appropriate for any particular set of viabilities can be determined relatively easily. The possible internal equilibria

Table 1 Sex specific genotype frequencies and fitness estimates

	Country of origin	NN	NH	HH	Total	Fitness NN	Fitness HH	Allele frequency N
Females								
Healey[1]	UK	138	115	13	266	0.86	0.63	0.73
Healey[2]	UK	631	493	77	1201	0.94	0.84	0.73
Healey[3]	UK	124	94	10	228	0.88	0.64	0.75
Healey[4]	Germany	433	373	60	866	0.92	0.81	0.72
Healey[5]	Finland	277	152	24	453	1.03	1.11	0.77
Healey[6]	UK	590	485	75	1150	0.93	0.81	0.72
Blood donors	UK	192	172	25	389	0.89	0.73	0.71
Colon cancer controls	UK	98	72	18	188	1.10	1.24	0.71
Mammography screening control women	UK	466	410	81	957	0.97	0.93	0.70
Ishitobi	Japan	95	56	3	154	0.86	0.43	0.80
Auranen[1]	Australia	139	123	20	282	0.92	0.80	0.71
Auranen[2]	Australia	439	322	54	815	0.98	0.94	0.74
Males								
Healey[6]	UK	644	435	116	1195	1.15	1.38	0.72
Blood donors	UK	381	329	51	761	0.91	0.78	0.72
Colon cancer controls	UK	115	95	21	231	1.02	1.05	0.70

Healey[1–6] are the data published in Healey et al.[1]
Ishitobi refers to the data published in Ishitobi et al.[6]
Auranen[1][2] are the controls published in Auranen et al.[3]
The fitness is relative to genotype NH and is computed separately for each sex specific group.
Italics indicate which data came from the original first report.

can be derived as the roots in (0,1) of a certain cubic equation (derived from the recurrence equations for $p(t)$ and $r(t)$ and not given here), and the stability of the equilibria at $p = r = 0$ and $p = r = 1$ can easily be obtained.[13]

Equilibrium ($p = r = 0$) is stable if ($h/b+k/d$)<2 and unstable if ($h/b+k/d$)>2.

Similarly the solution ($p = r = 1$) is stable if ($h/a+k/c$)<2 and unstable if ($h/a+k/c$)>2.

It can thus be demonstrated that the point estimates for the sex specific viabilities originally reported[1] (that is, $a = 0.93$, $b = 0.81$, $c = 1.15$, $d = 1.38$) identify two external stable equilibria. Moreover there exists a single internal equilibrium which is therefore unstable (SUS). In fact the viabilities reported originally were a special case in that they presupposed that the allele frequencies at equilibrium were the same in both sexes, which implies that we can write the
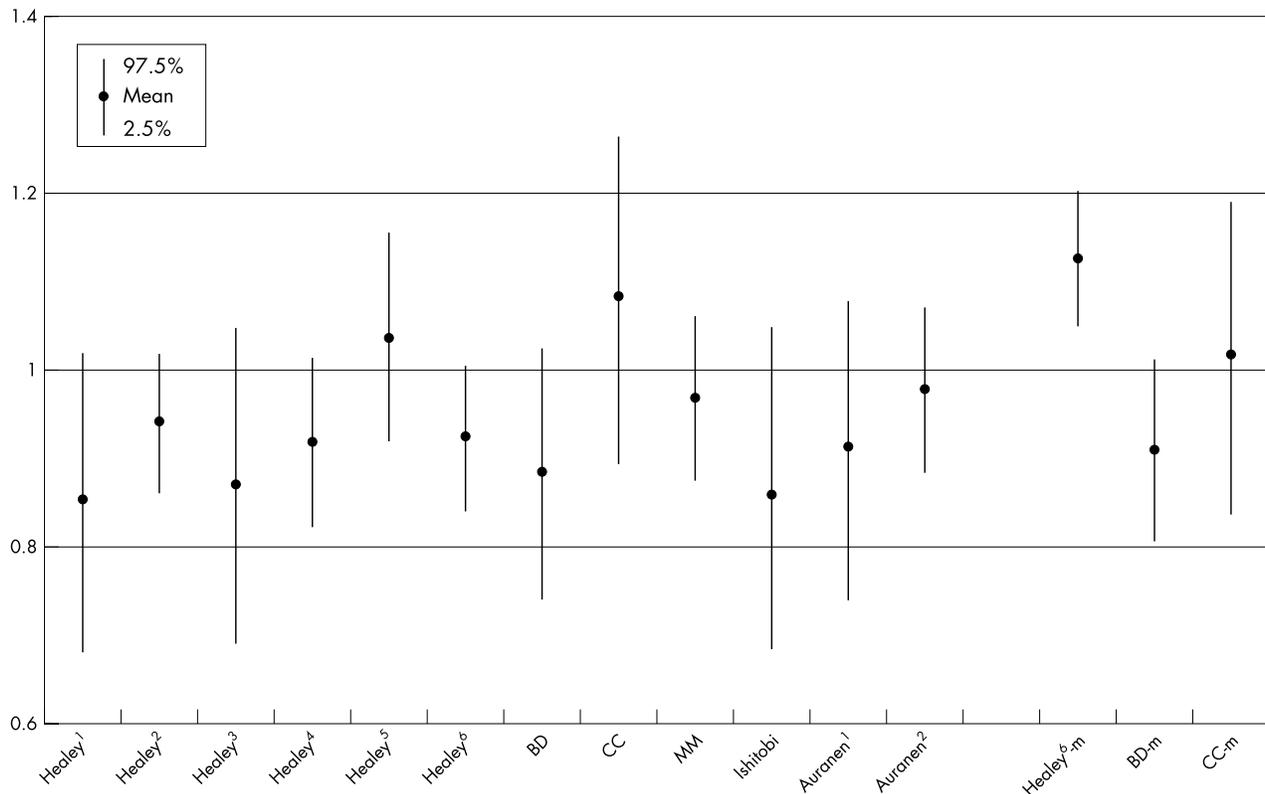

Figure 1 Means and confidence intervals for viability of genotype NN relative to NH. BD, blood donors; CC, colon cancer controls; MM, mammography screening control women; m, male.
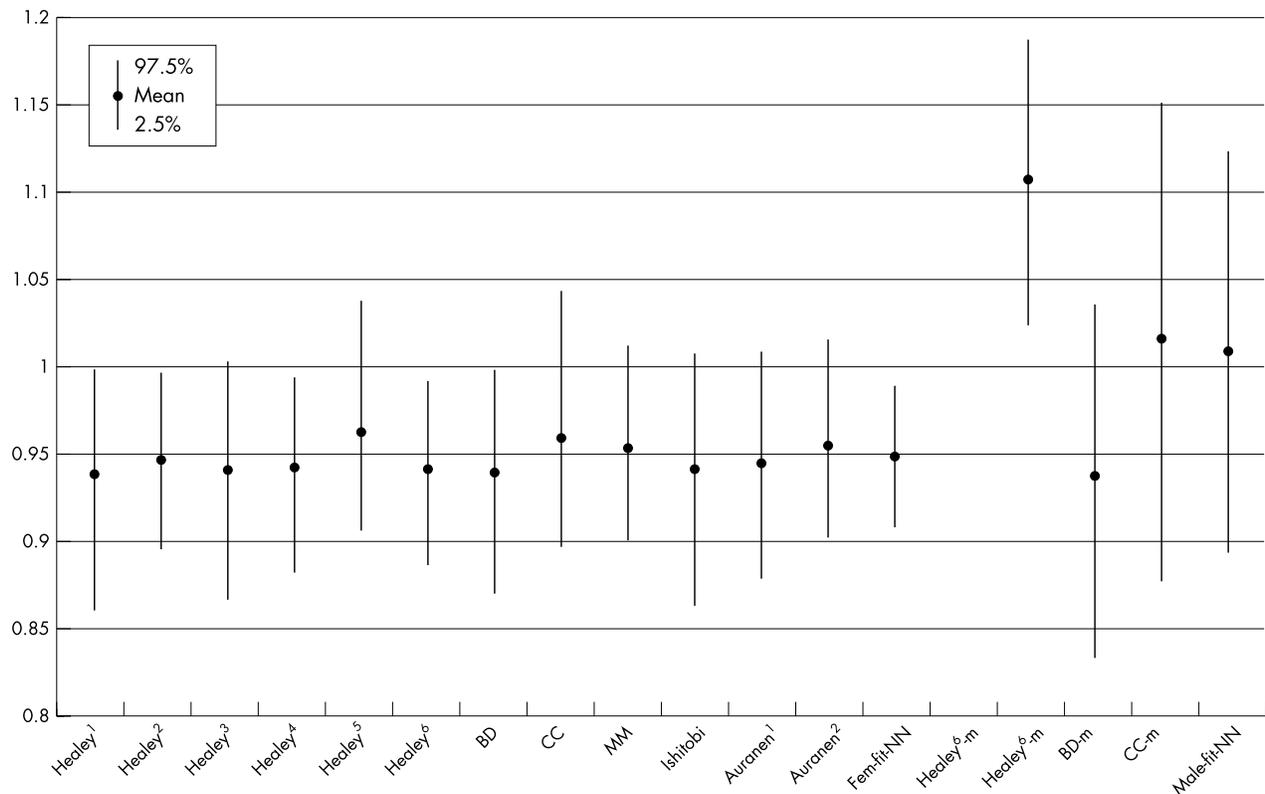
Figure 2 Random effects: means and confidence intervals for sex specific viability of genotype NN relative to NH. Each column represents the mean and confidence interval for the sex specific viability. The first 12 columns correspond to the first 12 rows in table 1, after a random effects model has been superimposed. Column 13 (Fem-fit-NN) represents the underlying relative viability for genotype NN in women. Columns 14–17 represent the underlying relative viability for genotype NN in men, corresponding to the last three rows in table 1, after a random effects model has been superimposed. The final column shows the mean and 95% confidence interval for the underlying relative genotype viability in men. BD, blood donors; CC, colon cancer controls; MM, mammography screening control women; m, male.
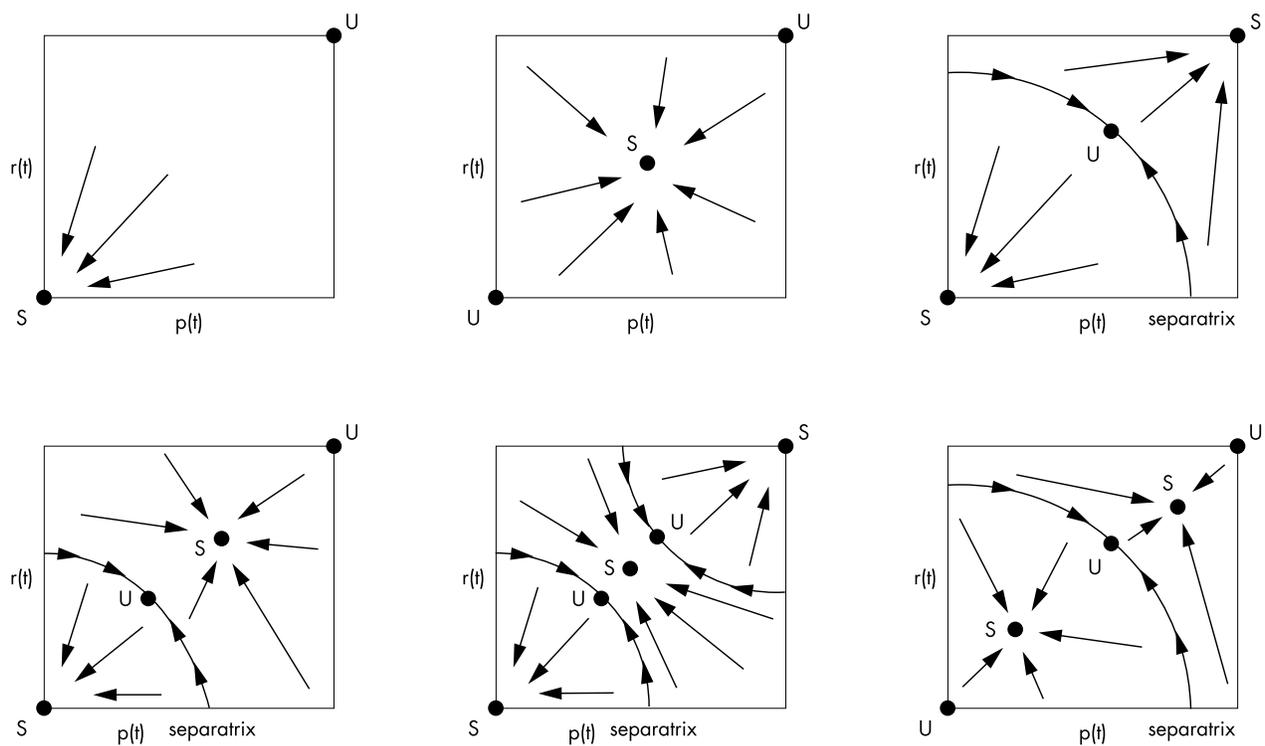

Figure 3 Representation of the six qualitatively distinct dynamical behaviours of ($p(t)$, $r(t)$). Arrows indicate movement of ($p(t)$, $r(t)$) to a stable polymorphism. Arrows along the curves (separatrices) indicate the (pathological) convergence to an unstable polymorphism.

viabilities in terms of three parameters s, t, and u where $a = 1+s$, $h = 1$, $b = 1+t$, $c = 1+us$, $k = 1$, and $d = 1+ut$, which gives $p = r = t/(t+s)$ both at conception and maturity, which is independent of u. Stability of this equilibrium requires that there is heterozygous advantage in at least one sex. Suppose for example that this is true in the female sex and take $-0.5 < s < t < 0$ (the assumption that s and t are both larger than 0.5 fits with the range of values relevant here and makes presentation of the results somewhat easier), then we require $u \leqslant -1/s$ (to ensure that the viabilities in men are non-negative) and the equilibrium is stable (demonstrated by checking the instability of the cases $p = 0$ and $p = 1$) provided $u > -1/(1+2t)$, that is u in $(-1/(1+2t), -1/s)$. This interval always contains the subinterval [0, 1] the extremes of which are no sex specific differential viability, when $u = 1$, and selection in one sex only, when $u = 0$.

The main interest behind the meta-analysis was to refine the estimates for the sex specific viabilities. The point estimates from the full meta-analysis identify two external unstable solutions. Since there was no evidence for differences in allele frequencies between the sexes, the estimates fall within the above special case. Examination of the number of equilibria through the cubic demonstrates that these specific viabilities lead to a single stable equilibrium (USU) with allele frequency of 0.71 in both sexes. In terms of the special case parameters appropriate when allele frequencies are the same at equilibrium we have $s = -0.05$, $t = -0.13$, and $u = -0.2$. There is a unique globally stable equilibrium at $p = r = t/(s+t)$ and unstable end points.

DISCUSSION

This BRCA2 polymorphism has been studied predominantly in female populations. The analysis of our own three datasets plus the nine distinct published female populations presents evidence that there is a consistent heterozygote advantage in women, with viabilities for both homozygotes significantly different from 1. The first report[1] found strong evidence that the frequencies in newborn boys and girls were different. However, our analysis of the three male populations finds no evidence that the relative genotype viabilities are significantly different from 1, and in fact there is no statistical evidence that the male and female viabilities are different from each other.

It could be argued that there is some evidence for differences between the newborn and adult male viabilities. However, no such difference was seen in the female sex and such a strong heterozygote deficit in the newborns could not be swept out of the male population during life as this would amount to an impossibly high number of deaths. So it is most likely that much of the apparent differences between the male samples are due to chance.

What is noteworthy is the apparently stable allele frequency, which is very similar in all populations examined, and exhibits little or no sex specific differences. If male relative viabilities were genuinely different from those seen in women this would most often result in differential sex specific allele frequencies maintained at equilibrium. In fact the allele frequencies can only be the same in both sexes if the selection were balanced such as outlined for the special case above. Therefore, we can state with some confidence that if there is sex differential viability then it cannot be as dramatic as first reported. Although the point estimates from the joint analysis indicate no genotype specific selection in men, the confidence intervals are wide so it is possible that genotype viability may be the same in both sexes. Genotypes from further large samples of men are needed to estimate and model the male genotype viabilities more precisely.

BRCA2 is involved in DNA repair,[14] and it is known that individuals carrying a germline mutation in BRCA2 are at high risk of breast and ovarian cancer.[15] The BRCA2 polymorphism investigated in this paper, specifically the NN genotype, has previously been shown to be associated with a moderate but significant increased risk of breast and ovarian cancer.[1 3 5 6] As no other high frequency or associated single nucleotide polymorphisms have been found in this region Auranen et al[3] suggested that this polymorphism may result in a structural change in protein, which accounts for the increased risk in cancer. This structural change may also be at the root of the suggested genotype specific fetal viability.[1] There are several established examples of heterozygote advantage at various genetic loci,[16 17] but thus far no reports in DNA repair genes or cancer. One possible explanation is that there is a microdeletion in strong linkage disequilibrium with the N allele. Such a microdeletion may be tolerated and possibly favoured in the heterozygote, but survival of homozygotes may be compromised and they may be at increased risk of cancer. The N372H nonconservative polymorphism falls within a region of the gene that has been shown to interact with histone acetyltransferase P/CAF.[18] When more is known about the structural consequences of this polymorphism the mechanism generating these observations may become clearer.

. . . .

Authors' affiliations
M D Teare, A Cox, J Shorto, C Cannings, Division of Genomic Medicine, University of Sheffield, Sheffield, UK
C Anderson, Institute of Cell, Animal and Population Biology, Ashworth Laboratories, West Mains Road, University of Edinburgh, Edinburgh, Scotland, EH9 3JT, UK
D T Bishop, Genetic Epidemiology Division, Cancer Research UK Clinical Centre, St James's University Hospital, Leeds, UK

Correspondence to: M D Teare, PhD, Mathematical Modelling and Genetic Epidemiology, Division of Genomic Medicine, University of Sheffield, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF, UK; m.d.teare@sheffield.ac.uk

REFERENCES

1 **Healey CS**, Dunning AM, Teare MD, Chase D, Parker L, Burn J, Chang-Claude J, Mannermaa A, Kataja V, Huntsman DG, Pharoah PDP, Luben RN, Easton DF, Ponder BAJ. A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability. *Nat Genet* 2000;**26**:362–4.

2 **Barrett JH**, Smith G, Waxman R, Gooderham N, Lightfoot T, Garner RC, Augustsson K, Wolf CR, Bishop DT, Forman D, and the Colorectal Cancer Study Group. Investigation of interaction between N-acetyltransferase 2 and heterocyclic amines as potential risk factors for colorecal cancer. *Carcinogenesis* 2003;**24**:275–82.

3 **Auranen A**, Spurdle AB, Chen XQ, Lipscombe J, Purdie DM, Hopper JL, Green A, Healey CS, Redman K, Dunning AM, Pharoah PD, Easton DF, Ponder BAJ, Chenevix-Trench G, Novik KL. BRCA2 Arg372His polymorphism and epithelial ovarian cancer risk. *Int J Cancer* 2003;**103**:427–30.

4 **Hu N**, Li WJ, Su H, Wang CY, Goldstein AM, Albert PS, Emmert-Buck MR, Kong LH, Roth MJ, Dawsey SM, He LJ, Cao SF, Ding T, Giffen C, Taylor PR. Common genetic variants of TP53 and BRCA2 in esophageal cancer patients and healthy individuals from low and high risk areas of northern China. *Cancer Detect Prev* 2003;**27**:132–8.

5 **Spurdle AB**, Hopper JL, Chen XQ, Dite GS, Cui JS, McCredie MRE, Giles GG, Ellis-Steinborner S, Venter DJ, Newman B, Southey MC, Chenevix-Trench G. The BRCA2 372 HH genotype is associated with risk of breast cancer in Australian women under age 60 years. *Cancer Epidemiol Biomarkers Prev* 2002;**11**:413–6.

6 **Ishitobi M**, Miyoshi Y, Ando A, Hasegawa S, Egawa C, Tamaki Y, Monden M, Noguchi S. Association of BRCA2 polymorphism at codon 784 (Met/Val) with breast cancer risk and prognosis. *Clin Cancer Res* 2003;**9**:1376–80.

7 **Weir BS**. *Genetic data analysis. Methods for discrete population genetic data.* Sunderland, MA: Sinauer Associates, 1990.

8 **Spiegelhalter D**, Thomas A, Best N. *WinBUGS version 1.3 user manual.* Cambridge UK: MRC Biostatistics Unit, 2000.

9 **Edwards AWF**. *Foundations of mathematical genetics.* Cambridge: Cambridge University Press, 1977.

10 **Cannings C**. Unisexual selection at an autosomal locus. *Genetics* 1969;**62**:225–9.

11 **Owen ARG**. A genetical system admitting of two distinct stable equilibria under natural selection. *Heredity* 1953;**7**:97–102.

12 **Selgrade JF**, Ziehe M. Convergence to equilibrium in a genetic model with differential viability between the sexes. *J Math Biol* 1987;**25**:477–90.

13 **Parsons PA**. The initial progress of new genes with viability differences between sexes and with sex linkage. *Heredity* 1961;**16**:103–7.

14 **Powell SN**, Kachnic LA. Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* 2003;**22**:5784–91.

15 **Antoniou A**, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjakoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003;**72**:1117–30.

16 **Krawczak M**, Zschocke J. A role for overdominant selection in phenylketonuria? Evidence from molecular data. *Hum Mutat* 2003;**21**:394–7.

17 **Segal S**, Hill AVS. Genetic susceptibility to infectious disease. *Trends Microbiol* 2003;**11**:445–8.

18 **Fuks F**, Milner J, Kouzarides T. BRCA2 associates with acetyltransferase activity when bound to P/CAF. *Oncogene* 1998;**17**:2531–4.