European Journal of Human Genetics (2001) 9, 787–793

www.nature.com/ejhg

ARTICLE

The western Swedish BRCA1 founder mutation 3171ins5; a 3.7 cM conserved haplotype of today is a reminiscence of a 1500-year-old mutation

Annika Bergman[1], Zakaria Einbeigi[2], Ulrica Olofsson[3], Ziad Taib[3], Arne Wallgren[2], Per Karlsson[2], Jan Wahlström[1], Tommy Martinsson[1] and Margareta Nordling*[,1]

[1]*Department of Clinical Genetics, Göteborg University, Sahlgrenska University Hospital, Göteborg, Sweden;* [2]*Department of Oncology, Göteborg University, Sahlgrenska University Hospital, Göteborg, Sweden;* [3]*Department of Mathematical Statistics, Chalmers University of Technology/Göteborg University, Göteborg, Sweden*

The most recurrent BRCA1/BRCA2 mutation in Sweden is the BRCA1 mutation 3171ins5. In the western part of Sweden this mutation accounts for as much as 77% of identified mutations in these two genes. Our aim was to analyse in detail the haplotype and founder effects of the 3171ins5 and furthermore attempt to estimate the time of origin of the mutation. In the study we included eighteen apparently unrelated families with hereditary breast and/or ovarian cancer. At least one individual in each family had previously tested positive for the 3171ins5 mutation. Polymorphic microsatellite markers were used for the haplotype analyses. The markers were located within or flanking the BRCA1 gene spanning a region of 17.3 cM. We found several different haplotypes both for disease alleles and for the normal alleles. However, a conserved haplotype of 3.7 cM was observed in the 3171ins5 carriers spanning over four markers located within or very close to the BRCA1 gene. As this haplotype was not present in any of the normal controls it is highly likely that this is a mutation identical by descent, i.e. a true founder. The results from the haplotype analyses were used to estimate the age of the mutation. Estimations based on the P_{excess} and linkage disequilibrium gives a first appearance of the mutation sometime around the 6th century, approximately 50 generations ago. *European Journal of Human Genetics* (2001) **9,** 787–793.

Keywords: BRCA1 gene; founder effect; breast neoplasm; ovarian neoplasm; haplotype

Introduction

The cloning of the BRCA1[1] and BRCA2[2] genes has led to the identification of several hundred germline mutations in families with a history of hereditary breast cancer. Some of the mutations in these genes are recurrent in patients of distinct geographic and ethnic populations due to so-called founder effects, i.e. mutations that have spread from a single ancestor.

*Correspondence: M Nordling, Department of Clinical Genetics, Göteborg University, SU/East, S-416 85 Göteborg, Sweden.
Tel: +46 31 343 4162; Fax: +46 31 842160;
E-mail: Margareta.Nordling@obgyn.gu.se
Received 6 April 2001; revised 21 June 2001; accepted 28 June 2001

It is important to identify these founder mutations and determine their geographic distribution in order to be able to design efficient mutational screening. The Ashkenazi Jew and the Icelandic populations are two examples of groups with elevated risks of familiar breast/ovarian cancer due to specific founder mutations. In the Ashkenazi Jewish population three founder mutations – two BRCA1 mutations (185delAG and 5382insC) and one BRCA2 mutation (6174delT) are alone responsible for the greater proportion of the breast/ovarian cancer in high-risk families.[3–5] In the Icelandic population a unique founder mutation has been identified in the BRCA2 gene (999del5), and many families in this population with hereditary breast cancer are carriers of this mutation.[6] This has of course made the mutation analysis less time and cost consuming in this country. In other countries such as Austria,

France, Canada, Norway, Netherlands, Belgium, Hungary, Russia, Scotland and Sweden a number of specific founder mutations have been detected in the population.[7–14]

In Sweden several different germline mutations in BRCA1 and BRCA2 have been identified. Eight of these are likely to be unique founder mutations.[13] The most recurrent BRCA1/BRCA2 mutation in Sweden is the BRCA1 3171ins5 mutation (also denoted 3166ins5), with 33 recorded cases in the Breast Cancer Information Core (BIC) database (URL: http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/) and five more families that are reported but not yet registered (as of March 2001). At least 26 out of the 33 reported families have a Swedish origin. Among breast and ovarian cancer patients with germline mutations in the western Swedish Health Care Region the 3171ins5 mutation is by far the most recurrent mutation and is estimated to account for approximately 77% of identified BRCA1/BRCA2 mutations (unpublished data). The aim of this work was to analyse the haplotype distribution, geographical distribution and founder effects of the 3171ins5 mutation in families from the western part of Sweden, and furthermore, try to estimate the time of origin of the mutation.

Material and methods

Families

We included 18 apparently unrelated families with confirmed 3171ins5 mutation in this study. All affected families originate from the same geographical area along the west coast of Sweden. Blood samples have been collected during the last five years at the Cancer Genetic Counseling Clinic, jointly run by the departments of Clinical Genetics and Oncology at Sahlgrenska University Hospital in Göteborg, Sweden. The mutation status of the individual family members has been determined using a combined Single-Strand Conformation Polymorphism and Heteroduplex technique (SSCP/HD) and confirmed by DNA sequence determination as recently described by us.[15]

Genotyping

Thirteen polymorphic microsatellite markers located on chromosome 17q21, spanning a region of 17.3 cM, were used in genotyping (Figure 1). Primer sequences for amplification of these markers were obtained from the Genome Database (URL: http://www.gdb.org). The forward PCR primers were fluorescent dye labeled with either FAM or TET fluorescent dyes (Applied Biosystem, Foster City, CA, USA) and the amplification products were separated on an ABI310 or on an ABI377 instrument (Applied Biosystems) and analysed using the Gene Scan 2.1 and Genotyper software (Applied Biosystems). The genetic distances between the markers, indicated in Figure 1 were derived from the Marshfield Sex-Averaged linkage map (URL: http://research.marshfieldclinic.org/genetics/) and from Sarantaus *et al*.[16] A physical map has been resolved for three of the markers; D17S1321 – (225 kb) – D17S855 – (400 kb) – D17S1325.[17] These physical distances were converted to centiMorgans assuming that the genome average of 1 cM=1 Mb is approximately accurate in this gene region. The allele frequencies in the west Swedish population were defined using 10 non-carrier family members along with 48 blood donor controls, i.e. based on 116 control chromosomes.

Estimation of founder mutation age and statistical considerations

The following account is the method used for estimating the time since the appearance of the mutation. We assume that the single ancestral chromosome containing the founder disease allele also contained a marker allele denoted '1' at a nearby marker locus. We assume further that the recombination fraction between the disease and marker locus is denoted *r*. Based on the so called p_{excess} and linkage disequilibrium, the number of generations *t*, since the appearance of the mutation can be estimated. A slight modification of formula 4.45 by Sham[18] gives the formula:

$$\hat{t} = \frac{\ln\left(\frac{\hat{p}_{d1} \quad \hat{p}_{n1}}{1 \quad \hat{p}_{n1}}\right)}{\ln(1 \quad r)} \qquad (1)$$

that was derived by Olofsson[19] using the method of moments, based on the theory of Galton–Watson branching processes. $\hat{p}_{d1}$ is the frequency of the founder allele in the disease population and $\hat{p}_{n1}$ is the corresponding frequency in the normal population. $\hat{p}_{d1}$ can be estimated by $\frac{n_{d1}}{n_d}$, n_{dl} being the number of individuals carrying the founder allele in a sample of size n_d of individuals carrying the disease allele. We assume that the normal population is stable, i.e. that the allele frequencies, p_{ni}, of the various marker alleles are unchanged from generation to generation. For translation of map distances into recombination frequencies we used the Haldane map function.[18] To use the above estimate, the founder allele has to be known as well as the haplotypes of the different individuals in the sample. Estimates based on different markers were averaged to obtain a unified estimate of the age of the mutation. The same estimate as above can be arrived at in a variety of other ways. Risch *et al*.[20] proposed the following estimator:

$$\hat{t} = \frac{\log\left[\left(1 \quad p_{n1} \quad \hat{Q}\right)/(1 \quad p_{n1})\right]}{\log(1 \quad r)}$$

where $Q=P$ (the disease chromosome does not carry a progenitor marker allele).

In fact, this is again the same estimator as can be seen by replacing *Q* by its value:

$$Q = 1 \quad p_{1d}$$

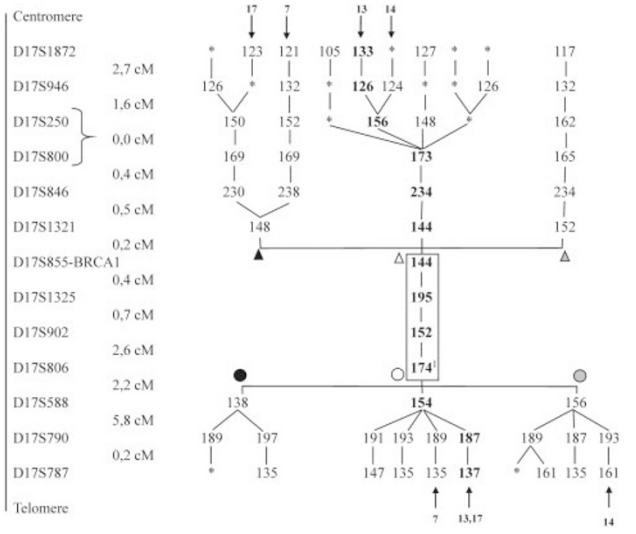

Figure 1 Haplotype branching tree derived from the analysis of haplotypes. The most common haplotype is indicated with boldface numbers. The examples of some family haplotypes from the material are indicated with arrows and the corresponding family number. The markers used are shown together with their relative genetic distance. Three main tracks both distal and proximal to the BRCA1 gene are highlighted by circles and triangles. [1] One family showed an aberrant allele in marker D17S806, allele 172 instead of allele 174, see text.

Lander and Botstein[21] proposed a slightly different estimator:

$$\hat{t} = \frac{1}{r}\ln(1 \quad \hat{p}_{1d})$$

One advantage of using this is that we can estimate its variance:

$$Var(\hat{t}) = \frac{n_{1d}}{n(n \quad n_{1d})r}$$

Notice that since Olofsson's estimate is also a maximum likelihood estimate, the same type of variance can be used. Pankratz' Moran model[22] gives an estimate very similar to Olofsson's, namely:

$$\hat{t} = \frac{1}{r}\ln\left(\frac{\hat{p}_{1d} \quad p_{1n}}{1 \quad p_{1n}}\right)$$

For small values of r, the relationship between this estimate and Olofsson's can be seen through the approximation:

$$e^{\ rt} = (1 \quad r)^t$$

The same estimate is also obtained using a diffusion approximation method due to Xiong and Guo.[23] To summarize, all methods above result in an equivalent estimate of generations, t. In our calculations we used Olofsson's estimate, i.e. formula (1). The method of Rannala and Slatkin[24] uses other parameters such as the population size, and is not convenient in this case. Neuhausen *et al*.[17] have proposed a maximum likelihood method which uses haplotypes, and allows for mutations at each marker locus. Ciotti *et al*.[25] report that this method results in similar estimates to the ones obtained using the other methods and we have not used it here.

It can be difficult to determine which allele is the actual founder allele, however we adopted the strategy of choosing the most frequent allele among individuals having the disease causing mutation as the founder allele. Moreover, the software GENEHUNTER2[26] was used to find the haplotypes for the families used in the study. More specifically, age calculations were based on an estimate of $\hat{p}_{d1}$ involving the haplotypes of 18 disease-allele carrying individuals while $\hat{p}_{n1}$ was estimated using a number of normal chromosomes from the relatives of the affected individuals and from 48 blood donors. All these estimates add to the uncertainty about the age estimate, as do the estimates of the map distances.

Results and discussion

Genotyping and haplotype analysis of BRCA1 3171ins5 carrying families

We analysed the haplotype using polymorphic markers spanning the BRCA1 loci in the 18 families to determine whether carriers from different families harboured the same mutation identical by descent. In 15 of the families at least two family members were genotyped. In the remaining three families only the index case was genotyped. Haplotypes were generated using GENEHUNTER2[26] but some were also corrected manually. Two representative families are shown in Figure 2. Several different haplotypes were present on the normal as well as on the mutated chromosomes. We found, however, that all of the analysed families shared a common 3.7 cM haplotype that included the BRCA1 intragenic marker D17S855 and three markers located close to it; D17S1325, D17S902 and D17S806 (Figure 1). One family carried disease allele 172 instead of the common 174 allele for marker D17S806. This is most likely due to a mutation event rather than a recombination event since the downstream alleles of the family followed the path of the founder alleles in the haplotype tree. The normal chromosomes displayed a large number of different haplotypes for the four markers while the 18 mutation chromosomes shared one haplotype only, i.e. 144-195-152-174 for markers D17S855-D17S1325-D17S902-D17S806. This haplotype was not seen in any of the normal chromosomes. Outside of the common four marker haplotype region the haplotypes diverged into several haplotypes. This has enabled us to try to group the haplotypes into different groups or branches (Figure 1).

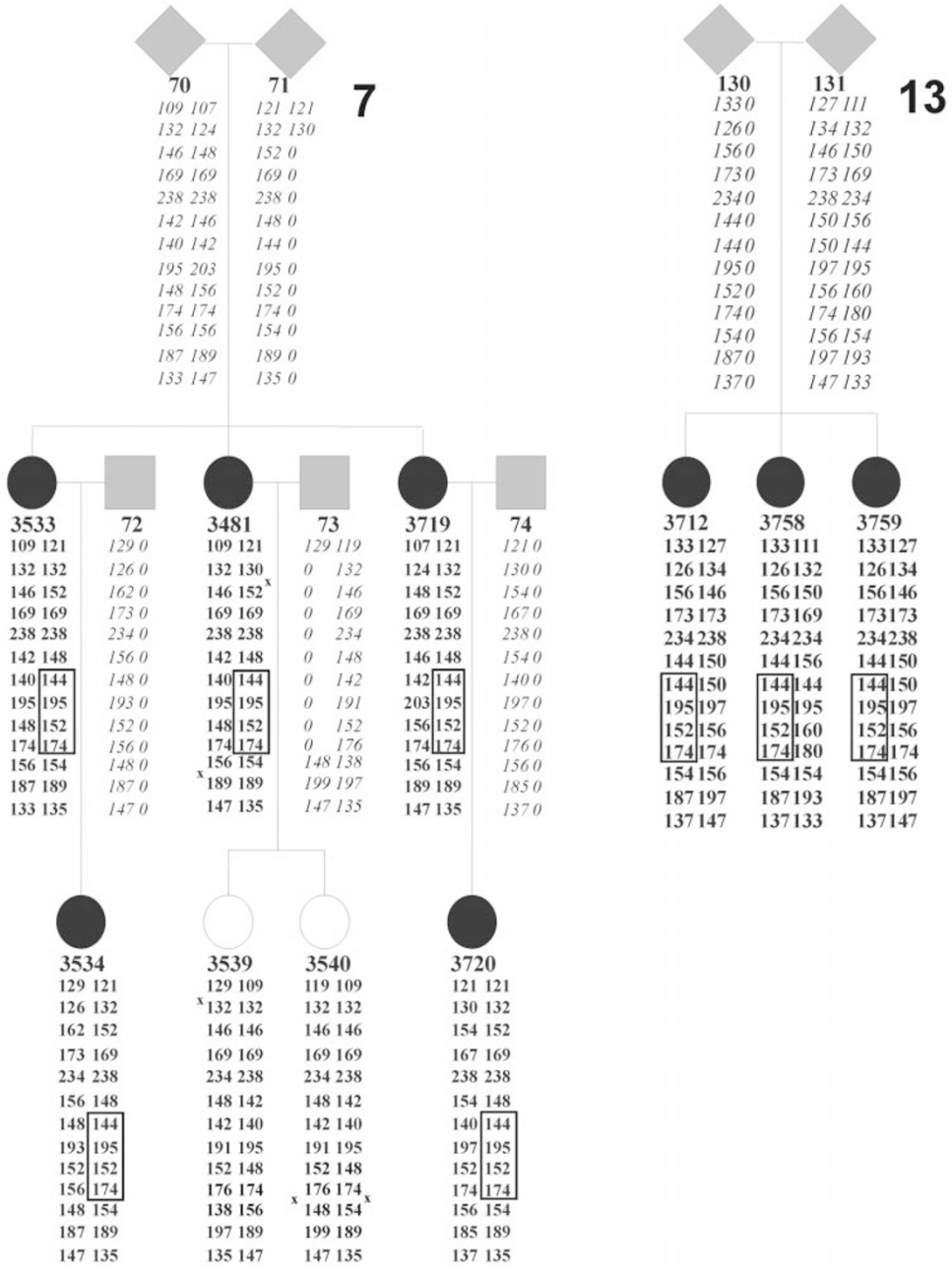

Figure 2 Haplotype analyses of two representative family pedigrees using GENEHUNTER2. The common haplotype region is boxed in each pedigree. The analysed family members are indicated with filled black symbols and normal text, while deduced alleles are in italics. Unaffected individuals are indicated with open symbols and those with uncertain status are indicated with gray symbols. Crossing-over events are denoted by x between the markers where crossing-over has occurred in the parental generation.

From this branched tree it is tempting to speculate on the close relation between the different families. Although the families themselves are unaware of any relation between them, a somewhat closer relation between some of the families is apparent. Two of the families share a common haplotype of 12.1 cM (see families 13 and 17 in Figure 1). However, no obvious correlation between the geographic origin – place of birth of the oldest family member harbouring the 3171ins5 mutation – and genotype was observed, see Figure 3. This is likely to be a reflection of how the population of western Sweden historically has been a migrating people along the west coast even though a migration beyond this distinct geographic area has been limited. Only a few reports of the 3171ins5 mutation have

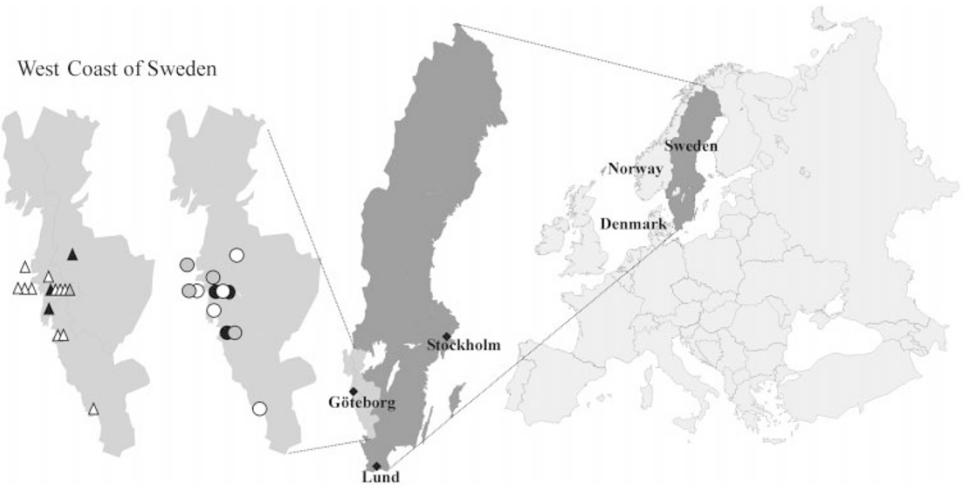

Figure 3 Geographic origins of families with 3171ins5 mutation. The various haplotypes derived from the haplotype analysis are indicated as triangles for branching centromeric of BRCA1 and circles for branching telomeric of BRCA1. The different shadings of the symbols represent the three main groups in the haplotype branching tree in Figure 1.

Table 1 Estimation of the age of the BRCA1-3171ins5 mutation

Marker	*Estimate (no of generations)*	*Distance from BRCA1 (cM)*	*Founder allele*
D17S1872	36.34	5,4	133
D17S946	52.72	2,7	126
D17S250	83.17	1,1	156
D17S800	54.04	1,1	173
D17S846	52.54	0,7	234
D17S588	28.19	5,9	154

been made from other countries (BIC database). It is likely that these families also have their origin in the western part of Sweden.

Haplotype construction and age estimation of 3171ins5
Estimates from the markers are shown in Table 1 and the mean estimated age of the BRCA1 3171ins5 mutation was calculated to be 51 generations, with a 95% confidence interval of [31–71]. The rational behind excluding some markers were; (i) The estimate can only be used when $\hat{p}_{d1} > \hat{p}_{n1}$, i.e. when the founder allele is more common in the disease population than among the normal controls. Thus, the markers D17S790 and D17S787 will not provide any information. (ii) The markers of the conserved haplotype generated a $\hat{p}_{d1}$ value equal to 1, so the method we use results in ln(1)/ln(1-r)=0 as an estimate of the mutation age. The explanation of this is that such a marker is still in perfect linkage disequilibrium with the disease gene as was the case at the time when the mutation occurred. This applies to markers D17S855, D17S1325, D17S902 and D17S806 that are, therefore, not used for estimating the mutation age.

The estimation of mutation age is obviously a rough estimate and the value is prone to changes in several different factors. Some of these factors are; (i) Choice of founder allele – there is no way to be sure of the original founder allele for markers located at some distance from the disease locus. The choice of founder allele is naturally of importance, since the difference between allele frequencies in the two populations is the foundation of the estimate. (ii) Recombination frequencies vary some between different genetic maps and also between different populations – if a marker is set at a location closer to the disease locus than it is in reality, we will get a larger estimate implying an older mutation. The estimate is very sensitive for changes in this parameter. (iii) Haplotypes – errors in the haplotypes will result in errors in the observed allele frequencies and result in an incorrect estimate.

Although rough, the estimate was in the range of 50 generations for this mutation, which dates the first appearance of the mutation to approximately 1500 years ago assuming 30 years per generation.[27] This number of generation is in line with, and sometimes somewhat greater than, other studies, e.g. from founder mutation age analyses, e.g. of Finnish BRCA1 and BRCA2 mutations.[16] For example BRCA1 3744delT was estimated to have first appeared 23–36 generations ago, while BRCA2 9346nt-2A>G was estimated to be only 7–11 generations old. Neuhausen *et al*.[17] analysed the age of several BRCA1 mutations, including the BRCA1 185delAG which is the most common mutation in individuals of Ashkenazi Jewish ancestry. They estimated the age of this mutation to be approximately 46 generations. Thus the age of this mutation is in the same range as that for BRCA1 3171ins5, analysed by us here. There is however a striking difference in the world-wide spread of these two mutations. This is likely to be a reflection of differences in migration

between the Ashkenazi Jews and the population of western Sweden. It appears obvious that the west Swedish people have experienced little migration since the time of the mutation's first appearance, sometime around the 6th century.

One explanation to the very limited migration of the people of western Sweden could be that they had a relatively good standard of living with not only the farming to depend on for survival but also the fishing. Large forests in the inland of Sweden also served as a border towards the east as they were not safe to pass through. The genetic specificity and stability of the population of western Sweden has also been reported for other genetic disorders such as carbohydrate deficient glycoprotein syndrome type 1A (CDG1A)[28] and psoriasis.[29] We would have expected the 3171ins5 mutation to have appeared in Denmark since the geographical closeness is apparent (Figure 3), but also for historical reasons, the south-western parts of Sweden belonged to Denmark from the middle ages until the 17th century. However no 3171ins5 has been detected in Danish breast and/or ovarian cancer families (Dr Åke Borg, personal communication). The mutation has been detected in Stockholm and Lund, and in these cases the families seem to have their origin in the western part of Sweden (data not shown). Two families have been diagnosed with the mutation in Norway, close to the Swedish border (BIC database). Another two individuals carrying the mutation have been reported from Seattle, Washington, USA.[30] A majority of the population that emigrated from Sweden to USA in the 19th century settled down in Minnesota and Washington state. An ancestor from the west coast of Sweden could thus explain the existence of BRCA1 3171ins5 carriers in this part of the world. The very limited spread of this mutation since the 6th century tells us that the population of western Sweden is a genetically unique population and might be well suited for future gene localization research. The reduced heterogeneity of the population of western Sweden population has also greatly facilitated carrier detection in this region as screening of 3171ins5 mutation now is the first analysis of choice for patients and relatives requiring diagnostic testing for BRCA1/BRCA2.

Acknowledgements

This research was funded by the King Gustav V Jubilee Clinic Cancer Research Foundation and the Assar Gabrielsson Fund.

References

1 Miki Y, Swensen J, Shattuck-Eidens D *et al*. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; **266**: 66–71.
2 Wooster R, Neuhausen SL, Mangion J *et al*. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 1994; **265**: 2088–2090.
3 Abeliovich D, Kaduri L, Lerer I *et al*. The founder mutations 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashkenazi women. *Am J Hum Genet* 1997; **60**: 505–514.
4 Tonin P, Weber B, Offit K, *et al*. Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer families. *Nat Med* 1996; **2(11)**: 1179–1183.
5 Levy-Lahad E, Catane R, Eisenberg S *et al*. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am J Hum Genet* 1997; **60(5)**: 1059–1067.
6 Thorlacius S, Olafsdottir G, Tryggvadottir L *et al*. A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nat Genet* 1996; **13(1)**: 117–119.
7 Simard J, Tonin P, Durocher F *et al*. Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. *Nat Genet* 1994; **8(4)**: 392–398.
8 Peelen T, van Vliet M, Petrij-Bosch A *et al*. A high proportion of novel mutations in BRCA1 with strong founder effects among Dutch and Belgian hereditary breast and ovarian cancer families. *Am J Hum Genet* 1997; **60(5)**: 1041–1049.
9 Liede A, Cohen B, Black DM *et al*. Evidence of a founder BRCA1 mutation in Scotland. *Br J Cancer* 2000; **82**: 705–711.
10 Wagner TM, Moslinger R, Zielinski C, Scheiner O, Breiteneder H. New Austrian mutation in BRCA1 gene detected in three unrelated HBOC families [letter]. *Lancet* 1996; **347**: 1263.
11 Ramus SJ, Kote-Jarai Z, Friedman LS *et al*. Analysis of BRCA1 and BRCA2 mutations in Hungarian families with breast or breast-ovarian cancer [letter]. *Am J Hum Genet* 1997; **60**: 1242–1246.
12 Gayther SA, Harrington P, Russell P, Kharkevich G, Garkavtseva RF, Ponder BA. Frequently occurring germ-line mutations of the BRCA1 gene in ovarian cancer families from Russia [letter]. *Am J Hum Genet* 1997; **60**: 1239–1242.
13 Johannsson O, Ostermeyer EA, Hakansson S *et al*. Founding BRCA1 mutations in hereditary breast and ovarian cancer in southern Sweden. *Am J Hum Genet* 1996; **58(3)**: 441–450.
14 Claes K, Machackova E, De Vos M, Poppe B, De Paepe A, Messiaen L. Mutation analysis of the BRCA1 and BRCA2 genes in the Belgian patient population and identification of a Belgian founder mutation BRCA1 IVS5+3A>G. *Dis Markers* 1999; **15(1-3)**: 69–73.
15 Einbeigi, Z *et al*. *Eur J Cancer* 2001; in press.
16 Sarantaus L, Huusko P, Eerola H *et al*. Multiple founder effects and geographical clustering of BRCA1 and BRCA2 families in Finland. *Eur J Hum Genet* 2000; **8(10)**: 757–763.
17 Neuhausen SL, Mazoyer S, Friedman L *et al*. Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. *Am J Hum Genet* 1996; **58(2)**: 271–280.
18 Sham P. Statistics in Human Genetics. London: Arnold, Hodder Headline Group, 1998.
19 Olofsson U. Modelling hereditary diseases using Galton-Watson processes. Licentiate: Chalmers University of Technology, Göteborg University, 2000.
20 Risch N, de Leon D, Ozelius L *et al*. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 1995; **9**: 152–159.
21 Lander ES, Botstein D. Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. Cold Spring Harb Symp Quant Biol 1986; **51**: 49–62.
22 Pankratz S. Stochastic Models and Linkage Disequilibrium: Estimating the Recombination Coefficient. Ph. D.: Rice University, 1998.
23 Xiong M, Guo SW. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. *Am J Hum Genet* 1997; **60**: 1513–1531.

24 Rannala B, Slatkin M. Likelihood analysis of disequilibrium mapping, and related problems. *Am J Hum Genet* 1998; **62**: 459 – 473.
25 Ciotti P, Struewing JP, Mantelli M *et al*. A single genetic origin for the G101W CDKN2A mutation in 20 melanoma- prone families. *Am J Hum Genet* 2000; **67**: 311 – 319.
26 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347 – 1363.
27 Tremblay M, Vezina H. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am J Hum Genet* 2000; **66**: 651 – 658.
28 Bjursell C, Stibler H, Wahlstrom J *et al*. Fine mapping of the gene for carbohydrate-deficient glycoprotein syndrome, type I (CDG1): linkage disequilibrium and founder effect in Scandinavian families. *Genomics* 1997; **39**: 247 – 253.
29 Enlund F, Samuelsson L, Enerback C *et al*. Psoriasis susceptibility locus in chromosome region 3q21 identified in patients from southwest Sweden. *Eur J Hum Genet* 1999; **7**: 783 – 790.
30 Malone KE, Daling JR, Thompson JD, O'Brien CA, Francisco LV, Ostrander EA. BRCA1 mutations and breast cancer in the general population: analyses in women before age 35 years and in women before age 45 years with first-degree family history. *Jama* 1998; **279**: 922 – 929.