ARTICLE

On the origin and diffusion of *BRCA1* c.5266dupC (5382insC) in European populations

Nancy Hamel[1,2], Bing-Jian Feng[3], Lenka Foretova[4], Dominique Stoppa-Lyonnet[5,6], Steven A Narod[7], Evgeny Imyanitov[8], Olga Sinilnikova[9,10], Laima Tihomirova[11], Jan Lubinski[12], Jacek Gronwald[12], Bohdan Gorski[12], Thomas v O Hansen[13], Finn C Nielsen[13], Mads Thomassen[14], Drakoulis Yannoukakos[15], Irene Konstantopoulou[15], Vladimir Zajac[16], Sona Ciernikova[16], Fergus J Couch[17], Celia MT Greenwood[18], David E Goldgar[3] and William D Foulkes*[,1,2]

The *BRCA1* mutation c.5266dupC was originally described as a founder mutation in the Ashkenazi Jewish (AJ) population. However, this mutation is also present at appreciable frequency in several European countries, which raises intriguing questions about the origins of the mutation. We genotyped 245 carrier families from 14 different population groups (Russian, Latvian, Ukrainian, Czech, Slovak, Polish, Danish, Dutch, French, German, Italian, Greek, Brazilian and AJ) for seven microsatellite markers and confirmed that all mutation carriers share a common haplotype from a single founder individual. Using a maximum likelihood method that allows for both recombination and mutational events of marker loci, we estimated that the mutation arose some 1800 years ago in either Scandinavia or what is now northern Russia and subsequently spread to the various populations we genotyped during the following centuries, including the AJ population. Age estimates and the molecular evolution profile of the most common linked haplotype in the carrier populations studied further suggest that c.5266dupC likely entered the AJ gene pool in Poland approximately 400–500 years ago. Our results illustrate that (1) *BRCA1* c.5266dupC originated from a single common ancestor and was a common European mutation long before becoming an AJ founder mutation and (2) the mutation is likely present in many additional European countries where genetic screening of *BRCA1* may not yet be common practice.
European Journal of Human Genetics (2011) 19, 300–306; doi:10.1038/ejhg.2010.203; published online 1 December 2010

Keywords: 5385insC; 5382insC; founder mutation; c.5266dupC; *BRCA1*; Europe

INTRODUCTION

Germline mutations in the breast cancer predisposition genes *BRCA1* and *BRCA2* account for a substantial fraction of hereditary breast cancer. Founder populations such as the French Canadian (FC) population of Quebec, the Icelandic population and the Ashkenazi Jewish (AJ) population have relatively frequent, well characterized founder mutations in the *BRCA* genes.[1]

According to the Breast Cancer Information Core database,[2] the two most frequently reported mutations in *BRCA1* are *BRCA1*: c.68_69delAG (traditionally known as 185delAG or 187delAG, 1980 reports) and c.5266dupC (also known as 5382insC or 5385insC, 1063 reports). Both mutations are known founders in the AJ population, with c.68_69delAG being the most frequent with approximately 0.9% of all AJ individuals being carriers.[3] *BRCA1*:c.68_69delAG is found most frequently in individuals of AJ descent but is also observed in some Hispanic populations, likely owing to historical gene flow between these two populations in Europe and America.[4] In contrast, c.5266dupC is less frequent in the AJ population (0.13%)[5] and is also observed in a wide range of other populations, primarily in Europe. Historically, AJ individuals rarely married outside their faith, raising the question of whether the mutation arose independently multiple times in the course of history or whether all mutation carriers share a single common ancestor. Neuhausen et al[6] reported early on that 21 mutation carrier families, including some families with Jewish ancestry, shared a common haplotype at markers near *BRCA1*, favouring the second hypothesis.

The age of a founder mutation can theoretically be estimated by determining the size of the conserved region surrounding the

[1]Program in Cancer Genetics, Departments of Oncology and Human Genetics, McGill University, Montreal, Quebec, Canada; [2]Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; [3]Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT, USA; [4]Department of Cancer Epidemiology & Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic; [5]Department of Biology, Institut Curie, Paris, France; [6]Medicine Faculty, University Paris Descartes, Paris, France; [7]Center for Research on Women's Health, Toronto University, Toronto, Ontario, Canada; [8]N.N. Petrov Institute of Oncology, Pesochny-2, St Petersburg, Russia; [9]Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon/Centre Léon Bérard, Lyon, France; [10]Laboratoire de Génétique Moléculaire, Signalisation et Cancer, UMR5201 CNRS, Université de Lyon, Lyon, France; [11]Latvian Biomedical Research and Study Centre, Riga, Latvia; [12]Department of Genetics and Pathomorphology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland; [13]Section of Genomic Medicine, Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark; [14]Odense University Hospital, Department of Clinical Genetics, Odense, Denmark; [15]Molecular Diagnostic Laboratory, IRRP, National Centre for Scientific Research 'Demokritos', Athens, Greece; [16]Department of Cancer Genetics, Cancer Research Institute, Slovak Academy of Sciences, Bratislava, Slovak Republic; [17]Department of Laboratory Medicine & Pathology and Health Sciences Research, Mayo Clinic, Rochester, MN, USA; [18]Genetics and Genome Biology, Hospital for Sick Children Research Institute, University of Toronto, Toronto, Ontario, Canada
*Correspondence: Dr WD Foulkes, Program in Cancer Genetics, Departments of Oncology and Human Genetics, McGill University, 1650 Cedar avenue Room L10-120, Montreal, Quebec, Canada. H3G 1A4. Tel: +514 934 1934 Ext. 44121; +514 938 8273; E-mail: william.foulkes@mcgill.ca
Received 25 June 2010; revised 13 October 2010; accepted 19 October 2010; published online 1 December 2010

mutation. In the case of a recent founder mutation, carriers will typically share a relatively large region of DNA surrounding the mutation where identical alleles will be observed at many loci in all carriers. As time passes, recombination events will create different chromosomal arrangements in selected individuals, and the region of shared homology around the mutation among carriers will become progressively smaller. In this study, we set out to first confirm using a large cohort of mutation carriers whether all c.5266dupC carriers indeed share a common haplotype background. Using the genetic information collected, we then attempted to estimate the number of generations since the appearance of the mutation in each population studied in the hopes of gaining some insight into where and when c.5266dupC arose and how the mutation may have spread throughout Europe to reach its current distribution, including its designation as an AJ founder.

Table 1 Mutation carrier families genotyped

Population group	Number of families	Number of samples	Contributing centres
Russian	26	28	St-Petersburg
Latvian	25	25	Riga
Ukrainian	5	6	Toronto
Czech	53	111	Brno
Slovak	5	8	Bratislava
Polish	27	81	Szczecin, Paris, Toronto
Ashkenazi Jewish	27	34	Toronto, Montreal, Paris
Danish	22	37	Copenhagen, Odense
Dutch	2	2	Toronto
French	41	41	Paris
German/Italian	2	4	Toronto
Greek	7	10	Athens, Toronto
Brazilian[a]	3	3	Toronto
	245	390	

[a]Families were recruited in Brazil and the samples are housed in Toronto.

METHODS

Subjects

A total of 390 DNA samples derived from c.5266dupC carrier families were genotyped, representing 245 families. For each participating family, a sample from one mutation carrier (index carrier, $n=245$) and available relatives ($n=145$) were obtained from collaborating research centers in Greece, Slovakia, Latvia, the Czech Republic, Russia, France, Poland, Denmark and Canada. All participants provided informed consent for use of their genetic material in research as well as self-reported population group membership. A summary of the participating subjects is presented in Table 1.

Genotyping

We initially genotyped a subset of 130 index cases and 75 of their relatives for 15 microsatellite markers (short tandem repeat markers, or STRs) and performed a series of preliminary analyses. Characteristics of the region studied and markers analyzed are presented in Table 2. From these initial analyses it was evident that some markers were too far from *BRCA1* to contribute useful information, whereas others were in partial linkage disequilibrium with nearby markers and provided mostly duplicate information. Thus, we selected a subset of seven STR markers covering 5.2 cM in 6.98 Mb of DNA that were highly heterozygous and which captured the haplotype diversity. These seven markers were then genotyped in a further 185 samples that were obtained subsequent to the original data collection and analysis. Genotyping data are presented in Supplementary Table 1. All genotyping was performed by deCODE genetics (Reykjavik, Iceland).

Age estimates using the maximum likelihood method

In order to estimate the age of the mutation (or more precisely, the number of generations since the most recent common ancestor, MRCA), we used the method that was first used to estimate the age of several *BRCA1* mutations including c.5266dupC[6] and was then extended and applied to *BRCA2* mutations.[7] This method uses maximum likelihood and allows for both recombination and mutational events at the marker loci as means of altering a presumed ancestral haplotype. Phased haplotypes were used if these could be inferred from available family data; otherwise, all possible haplotypes were constructed from multilocus genotype data and weighted according to their probability. For each value of G (the number of Generations since the MRCA), the relative likelihood that each haplotype is descended from the ancestral haplotype through mutation and recombination is calculated compared with the

Table 2 Microsatellite markers genotyped in carrier families

STR marker	Genomic position (GRCh37-hg19)	Distance to mutation (bp)	No. Alleles observed	Reported heterozygosity[a]
D17S1795	47925162	6 715 879	9	0.71
D17S1827	46638614	5 429 331	12	0.82
D17S931[b]	44995888	3 786 605	8	0.73
D17S934	43057681	1 848 398	11	0.84
D17S1804	43015297	1 806 014	9	NA
D17S1861[b,c]	42807097	1 597 814	12	0.82
D17S951[b,c]	41820209	610 926	10	0.77
D17S1327	41375525	166 242	13	NA
c.5266dupC	41209283	—	—	—
D17S855[b,d]	41204819	4464	9	0.82
D17S1147	40780067	429 216	7	NA
D17S1801[b,c]	40415909	793 374	9	0.63
D17S1299[b,c]	38994464	2 214 819	7	0.70
D17S1814[b]	38118881	3 090 402	10	0.78
D17S1818	37163238	4 046 045	13	0.84
D17S1867	35379550	5 829 733	8	0.67

Abbreviation: NA, not available.
[a]Values from GeneLoc Marker Cards (http://genecards.weizmann.ac.il/geneloc/index.shtml).
[b]STRs used for maximum likelihood age estimates are shown in bold.
[c]STRs used for single marker age estimates.
[d]STR located in intron 20 of *BRCA1*.

likelihood that it is a totally independent haplotype (ie, an independent recurrent c.5266dupC mutation on a different haplotype background). The value of G which maximizes this likelihood is obtained through iterative search. In all, 95% support intervals were constructed by identifying those points GL and GU where the likelihood differed from the maximum by 0.86 (corresponding to a χ^2 likelihood ratio statistic of 3.84, eg, $P=0.05$). In order to examine the likely genetic history of the c.5266dupC mutation, we analyzed separately each of several defined subgroups in which a sufficient number of samples were available for analysis: (a) AJ; (b) Russian (St Petersburg); (c) Polish (Szczecin and Paris only); (d) French; (e) Danish; (f) Czech/Slovak; (g) Latvia; (h) other.

Assumed genetic map
The recombination rates between markers were assumed to be those estimated in Kong *et al.*[8] Physical positions of the STRs and SNPs were those from the Human Reference sequence, build 3.7. For markers present on the deCODE map, we used the genetic positions in centimorgans as reported there, whereas for those not on the deCODE map, we estimated the genetic position from the proportion of physical distance between the known markers and then translated this to the genetic scale. This has the effect of using locally defined relationships between physical and genetic distance and thus can accommodate the reported recombination suppression in this region.[9]

Marker mutation rates
As a baseline we used the rates for the six dinucleotide and single tetranucleotide microsatellite markers as estimated from CEPH data by Weber and Wong[10] of 0.0006 and 0.002, respectively, for a mutation of a single repeat unit. We assumed the probability of changes of n repeat units in a given meiosis was $(0.0006 \text{ or } 0.002)^n$ for $n=2,3,4$ and that for more than four repeats was taken to be equal to that for four repeats units. Because of the imprecision of these rates (and model) we introduced another parameter into the likelihood and jointly estimated the number of generations and a multiplier of the assumed marker-mutation rates described above. Thus, to a certain extent, we let the data inform the proper marker mutation rates. In addition to the true underlying marker mutation rates, this also allows for potential genotyping errors to be accounted for in the model. We found that the best fit to our data was when the recombination rate was $2.75\times$ that of Weber and Wong.[10]

Allele frequencies
Our method uses marker allele frequencies in the calculation of the likelihood (Supplementary Table 2). We estimated frequencies from the unlinked allele of the chromosomes in the sample. Because the AJ population often has different allele frequencies at many genetic markers, the AJ frequencies were separately estimated from a sample of 30 controls and used for the likelihood calculations of the AJ multi-locus genotype/haplotype data.

Age estimates using single markers method
In an attempt to corroborate age estimate results obtained using the maximum likelihood method described above, we also estimated the time since MCRA using four markers (D17S1299, D17S1801, D17S951 and D17S1861) analyzed individually in three populations (Czech/Slovak, Polish and Danes) where we had the largest number of families with known phase for the markers linked to the mutation. The single marker method was implemented as described previously in Greenwood *et al.*[11] The Labuda correction for population-growth rate was assumed to be 1.5 and applied as previously described. Because this method does not consider marker mutations, which likely have a significant role in a region where there is documented recombination suppression such as *BRCA1*,[9] this method will not be as well-suited to our dataset as the maximum likelihood method, but can nevertheless serve to test the robustness of our original estimates.

RESULTS
Literature review of the frequency, distribution and morbidity of BRCA1:5266dupC
Figure 1 provides a comprehensive summary of the distribution and relative frequency of c. 5266dupC throughout Europe as reported in the literature over the past 15 years. In an attempt to obtain as accurate an estimate of relative frequency as possible for each population, we focused on reports where the entire *BRCA1* gene was screened for mutations using methods such as SSCP, dHPLC and sequencing, and did not include studies where only selected mutations were tested. For several countries, indicated in the figure legend, there were limited data available in the literature and frequency estimates may not accurately reflect actual mutation frequencies. Nevertheless, the compiled data clearly shows that *BRCA1*:c.5266dupC is not merely an AJ founder mutation but in fact appears to be the most common *BRCA1* mutation in several European countries.

While reviewing the literature, we also compiled available data from reports investigating mutation frequencies in unselected or consecutive cohorts of breast and/or ovarian cancer cases in an attempt to estimate the contribution of c.5266dupC to the incidence of breast and ovarian cancer. These data are presented in Table 3 and show that in Slavic countries,[12] where the frequency of the mutation is highest (Figure 1), carrier status of this single mutation is associated with a remarkably high proportion of reported ovarian cancers (9.4%) and a lesser, but still notable, percentage of breast cancers (2.2%).

Common ancestry of mutation carriers
Genotyping of index carriers was performed in two phases. We initially genotyped 15 STR markers within and flanking *BRCA1* over a total region of 12.5 Mb in 130 index carriers and 75 of their relatives (see Table 2 for marker information). Population/ethnic groups represented included AJ individuals, Czechs, Slovaks, Latvians, Greeks as well as a small number of Dutch, Ukrainian, Germans, Italians and Brazilians (Table 1). Genotypes from relatives were used to assign allelic phase and identify the haplotype in *cis* with the mutation in index carriers. Although evidence of recombination at markers further away from *BRCA1* was observed, there was clear conservation immediately flanking the mutation in all mutation carriers that was consistent with the theory that all c.5266dupC carriers share a single common ancestor. In individuals where phase could not be confirmed, genotypes consistent with a single conserved linked haplotype were always present. Based on these preliminary results, we used the seven most informative markers to genotype 115 additional index carriers and 70 of their relatives from regions of Europe where the presence of c.5266dupC was well documented (France, Poland, Denmark and Russia) and that were not represented the first phase of genotyping. These additional data confirmed that the common ancestry of c.5266dupC extends to all populations studied.

Origins of c.5266dupC
The number of generations since the last common shared ancestor was estimated using the maximum likelihood method for all index carriers combined, as well as for several defined subgroups where there were sufficient numbers of individuals to allow for separate analysis. Results are presented in Table 4 and suggest that c.5266dupC most likely originated in Northern Europe, specifically Russia or possibly Denmark, between 1800 and 1500 years ago (72 and 61 generations of 25 years, 95% CI: 49–107 and 40–89, respectively). Overlapping confidence intervals associated with the age estimates prevent us from establishing a conclusive chronology of exactly how the mutation spread among European countries. However, the conserved haplotype found within the AJ population is significantly younger (27 generations, 95% CI: 10–31), consistent with the mutation entering the AJ population more recently. Using the premise that molecular evolution from the original founder haplotype to the haplotypes observed in each population should be achieved using a minimal number of
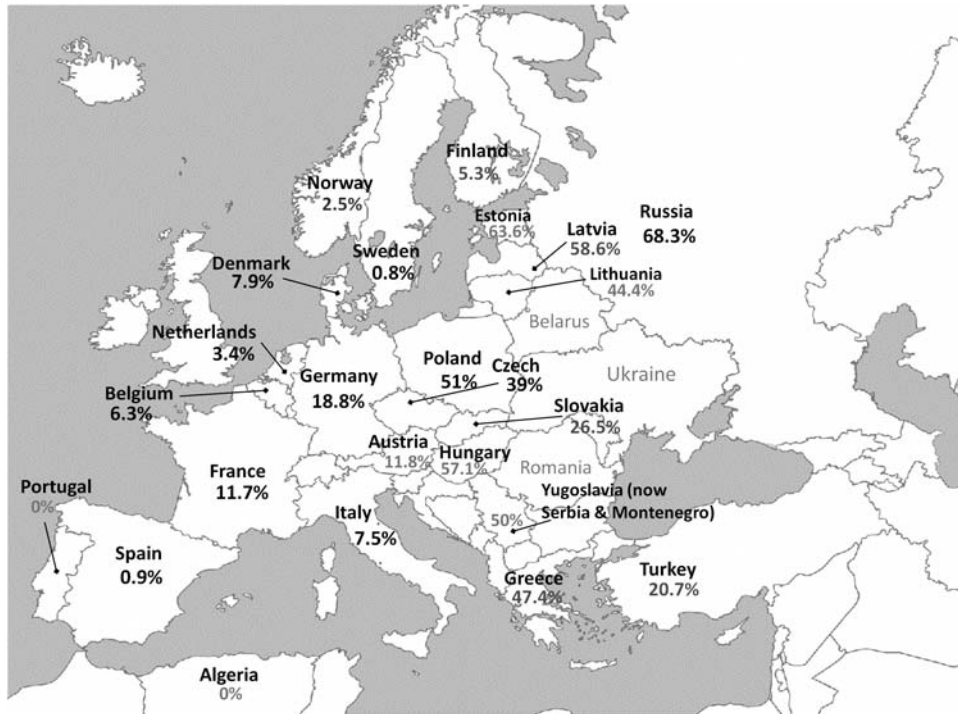
Figure 1 Map of Europe showing the proportion of all *BRCA1* mutations reported in the population accounted for by c.5266dupC per country. Only studies where the entire gene was screened for mutations were included. Countries marked * have between 20–50 total *BRCA1* mutations reported in the literature whereas countries marked ** have less than 20 total *BRCA1* mutations reported; therefore, frequency values for these countries have a high degree of uncertainty and should not be considered definitive values. *Slovakia [41,42]; Czech Republic [43–48]; Russia [49–53]; **Estonia [54]; Poland [55–63]; **Yugoslavia [64]; **Austria [65]; **Hungary [66,67]; **Lithuania [68]; *Latvia [69,70]; Germany [71–78]; Italy [79–89]; *Greece [90–94]; Netherlands [95,96]; Belgium [97–99]; *Norway [100]; Sweden [101–106]; Denmark [107–109]; **Finland [110–112]; Spain [113–119]; **Portugal [120]; France [121–126]; **Algeria [127]; *Turkey [128–133]. The references cited in this legend (41–133) are available in Appendix 1 as Supplementary Material.

Table 3 Contribution of BRCA1:c.5266dupC to cancer incidence per region

Country/ Region	5266dupC/BC cases tested	%	5266dupC/OC cases tested	%	Publications
Scandinavia[a]	1/823	0.12	2/606	0.33	21–24
Slavic Countries[b]	52/2322	2.2	81/864	9.4	25–34
Germany	8/800	1.0	NA	—	35
Hungary	7/500	1.4	1/90	1.1	36
Greece	13/1049	1.2	NA	—	37,38
Turkey	0/63	0	2/122	1.6	39,40

Abbreviations: BC, breast cancer; NA, not available; OC, ovarian cancer.
[a]Netherlands; Denmark; Norway; Sweden.
[b]Eastern Slavic: Russia, Latvia; Western Slavic: Lithuania, Poland, Czech Republic.[12]

Table 4 Estimated time to most recent common ancestor for c.5266dupC

Population group	Number of families	Number of generations	95% CI
Russia	26	72	49–107
Denmark	22	61	40–89
Czech/Slovak	58	53	42–66
Latvia	25	49	33–72
France	41	49	35–68
Poland	24	45	30–64
Ashkenazi Jewish	27	19	10–31
Other	22	68	44–101
Combined[a]	245	62	55–69

Abbreviation: CI, confidence interval.
Estimates obtained using the maximum likelihood method.
[a]Estimates obtained under best estimate of marker mutation rate=2.75× the rate in Weber and Wong.[10] One generation is approximately 25 years.

proposed marker mutation and recombination events, we attempted to reconstruct the most likely molecular evolution scenario of how the mutation spread from Russia or Scandinavia to other carrier populations (Figure 2a). Based on this reconstruction, the haplotype observed in the AJ population is most consistent with having its origin in Poland 400–500 years ago.

In addition, we tested the robustness of the above estimates using an average of single marker likelihoods in the Czech/Slovak, Polish and Danish population groups, where the largest numbers of probands with confirmed haplotypes were available. We obtained values of 43.5, 78.2 and 106.7 generations (averaged over four markers for each group) compared with 53 (95% CI: 42–66), 45 (95% CI: 30–64)

and 61 (95% CI: 40–89) generations using the maximum likelihood method. These results are in general agreement in that the Danish haplotype remains clearly older than the other haplotypes. In addition, estimates using the two methods are consistent for the Czech population, where 48 probands had phase information for at least two of the four markers tested, compared with only 16 and 13 probands with partial- or complete-phase information for the Polish and Danish groups, respectively, where more divergent results were obtained.
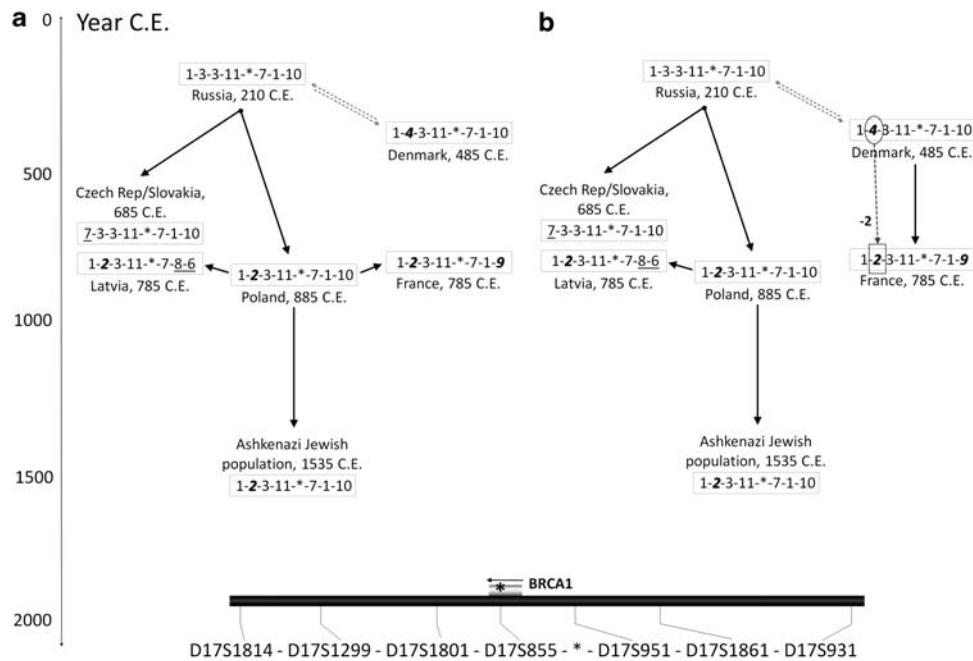
Figure 2 Reconstructed history of the c.5266dupC mutation based on age estimates from the maximum likelihood model. Underlined are potential recombination events (*n*=2), and highlighted bold are presumed stepwise marker mutations (*n*=3). *marks the position of the c.5266dupC mutation. (**a**) 'Best fit' reconstruction based on molecular data. The model uses the most common haplotype observed in each population group and relies on the smallest number of potential recombination and stepwise marker mutation events to account for existing observed haplotypes. (**b**) Alternate scenario based on historical considerations where the mutation could have been carried by Viking raiders to distant countries such as France. An additional, biologically unlikely two-step marker mutation from *allele 4* (encircled) of the Danish haplotype to *allele 2* of the French haplotype is introduced and would argue against this scenario. However, Danish *allele 4* is linked in 42% of our mutation carriers, whereas Danish *allele 3* is linked in 25%. Similarly, French *allele 3* is linked in 43% of mutation carriers, compared with French *allele 2* in 57%. Thus, *allele 3* is the second most commonly linked allele in both populations today and could have been a common allele transmitted from the Danish ancestor to the French ancestor in 785 CE before both population groups continued to diverge, making this an alternate scenario worth considering.

DISCUSSION

c. 5266dupC originated in Northern Europe between 200–500 common era

In the first centuries of the Common Era (CE), borders in Northern Europe were ill defined. Scandinavia encompassed sparsely populated north-eastern regions including present-day Norway, Denmark and Sweden as well as Finland and Iceland (uninhabited at the time), whereas the region that is now Russia, Central Asia and Ukraine had been occupied by Scythian tribes for several centuries. The following 600 years saw the Scythians conquered by the passage of Huns, Goths and Turks, leaving surviving Slavic tribes to spread throughout Central and Eastern Europe. Slavic tribes were also periodically harassed by Northmen, but the extent of potential genetic exchanges between Scandinavian and Slavic tribes during this period, whether through raids or trade, is not known. Russia's documented history only begins in the ninth century after Rurik, the 'great ruler of Novgorod', founded the first Russian dynasty. Historians suggest Rurik may in fact have been a Viking hailing from Sweden, creating further genetic ties between Scandinavians and Russians.[13] It is thus ethnically and geographically challenging to pinpoint the first c.5266dupC carrier, and overlapping confidence intervals surrounding our age estimates do not allow us to conclusively establish which of the Russian or Danish haplotypes preceded the other.

Regardless of its precise origin, the best fit molecular scenario suggested by our data and requiring the smallest number of mutation/ recombination events as shown in Figure 2a favours a sequence of events where the mutation slowly spread west and south to the rest of Europe from the Russian plains with the Slavic migrations.[14] Although

the spread of the mutation to adjacent areas such as Latvia or Poland through marriages seems a natural development, it is more difficult to comprehend how the mutation could have passively spread from Russia over such a short period of time to countries as distant as France or Turkey at frequencies high enough to become established in these populations and persist to this day. One possibility is that today's mutation carriers in these countries represent a subset of the population who relatively recently emigrated from Slavic countries, bringing the mutation with them; however, this explanation cannot satisfactorily explain how the mutation appears to have spread so thoroughly throughout Europe.

History suggests an alternate scenario. By the end of the 8th century, Viking seamen, who were essentially Scandinavian merchants turned to opportunistic looting, were raiding Christian communities and monasteries far and wide and could easily have spread the mutation directly and simultaneously to all corners of Europe.[15,16] For instance, c.5266dupC is also observed in the Yorkshire region of Northern England (GR Taylor, personal communication) where Vikings raids had been routine well before England came under the rule of Norman kings in 1066 CE. The Normans were themselves descendants of a group of Viking raiders that were allowed to settle in Northern France (Normandy) in 911 CE against the promise of protection from further raids on the local population, thereby providing several avenues for direct genetic admixture between Scandinavians and French and English locals.[17,18] Although historically highly plausible, our genotype data make this second scenario appear less likely because it requires a two allele slip at marker D17S1299 (from *allele 4* in the Danes to *allele 2* in the French, Figure 2b) rather than a single, stepwise-allele

change from the Russian *allele 3* to the Polish/French *allele 2*, the latter being far more likely genetic. However, it is important to note that although *allele 4* is the allele most commonly linked to the mutation in today's Danish population, only 42% of our mutation carriers carry this allele, whereas *allele 3* is also represented at appreciable frequency in this population (25%) as well as in the French population (43%, compared with 57% who carry *allele 2*). Thus, *allele 3* may well have existed at higher frequency in ancestors from both population groups back in 785 CE before diverging in later years. Investigation of haplotype data in mutation carriers from intervening countries such as Italy, Germany, Austria and Hungary, and of course from other countries such as England, Sweden and Norway, would be of great interest to refine the molecular data and attempt to elucidate this question.

It is interesting to note that the mutation does not appear to have been carried to North America by French colonists as was the case for several other well-characterized founder mutations found today in the FC population.[19] One possible explanation is that c.5266dupC was restricted to a subset of the French population who did not participate in the colonization process to the America in the 16th century. Another possibility is that settlers did bring the mutation to North America, but that in this instance founder effect acted to remove it from the genetic pool so that it did not become established in later generations.

c.5266dupC entered the AJ population in Poland near 1500–1600 CE

Despite the uncertainty regarding its origin and the manner in which the mutation spread throughout Europe, one clear conclusion emerging from our data is that c.5266dupC entered the AJ population much more recently, around 1500–1600 CE. In addition, the dominant AJ haplotype is identical to the dominant Polish haplotype, suggesting the mutation was likely acquired through admixture in Poland. This is a highly plausible scenario historically. Jews were a minority everywhere and were culturally and genetically isolated in medieval Europe where all rulers were Christians. In the year 1500 CE, there were perhaps 50 000 Jews living in Poland and Lithuania, but in the following years, Jews who were expelled from surrounding Christian countries such as England, Germany, Italy, Portugal and Spain were welcomed to Poland, thanks to their strong contribution to the Polish economy. By 1650, Poland counted many settled Jewish communities and in spite of often difficult relations between Christian and Jewish Poles, the Polish Jewish population had grown to 500 000, nearly 30% of the world's Jewish population.[20] This rapid population expansion would have significantly improved the odds of admixture with the local Polish population, even for an otherwise relatively genetically isolated group, and conceivably facilitated the acquisition of c.5266dupC in the Jewish gene pool where it persisted and became established as a low frequency founder mutation alongside the more frequent Jewish mutations *BRCA1*:c.68_69delAG and *BRCA2*:c.5946delT (traditionally known as 6174delT).

CONCLUSIONS

As demonstrated recently in a study of three AJ founder mutations, age estimate results are strongly dependent on assumptions made about recombination and mutation history that cannot be verified, and results may vary depending on the method used, especially in the case of older mutations such as the one studied here.[11] By using a method relying on molecular mutation and recombination rates informed partly by the data itself, we attempted to minimize the number of assumptions used in our model to get as accurate a picture of the molecular history of the mutation as possible. Although the exact origin and manner of dissemination of *BRCA1*:c.5266dupC may never be precisely elucidated, we were able to establish conclusively that all mutation carriers inherited the mutated chromosome from a single common ancestor who lived well before the establishment of current political boundaries. In addition, the current frequency distribution of the mutation coincides well with expectations from historical records. It is therefore likely that it will be found in several additional European countries sharing ancestries with the populations studied here, but where genetic testing and reporting in the literature has not been commonly performed to date. Furthermore, given the significant contribution of this mutation to the ovarian cancer burden in countries where it is found at high frequency, systematic screening of all ovarian cancer cases for *BRCA1*:c.5266dupC would be highly beneficial to the risk management of affected families.

1 Narod SA, Foulkes WD: BRCA1 and BRCA2: 1994 and beyond. *Nat Rev Cancer* 2004; **4**: 665–676.
2 National Human Genome Research Institute, NIH: Breast Cancer Information Core Database. National Human Genome Research Institute, NIH, 2009. Accessed 10 February 2010; http://research.nhgri.nih.gov/bic/.
3 Struewing JP, Abeliovich D, Peretz T *et al*: The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. *Nat Genet* 1995; **11**: 198–200.
4 Ah Mew N, Hamel N, Galvez M, Al-Saffar M, Foulkes WD: Haplotype analysis of a BRCA1: 185delAG mutation in a Chilean family supports its Ashkenazi origins. *Clin Genet* 2002; **62**: 151–156.
5 Roa BB, Boyd AA, Volcik K, Richards CS: Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat Genet* 1996; **14**: 185–187.
6 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**: 271–280.
7 Neuhausen SL, Godwin AK, Gershoni-Baruch R *et al*: Haplotype and phenotype analysis of nine recurrent BRCA2 mutations in 111 families: results of an international study. *Am J Hum Genet* 1998; **62**: 1381–1388.
8 Kong A, Gudbjartsson DF, Sainz J *et al*: A high-resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
9 Liu X, Barker DF: Evidence for effective suppression of recombination in the chromosome 17q21 segment spanning RNU2-BRCA1. *Am J Hum Genet* 1999; **64**: 1427–1439.
10 Weber JL, Wong C: Mutation of human short tandem repeats. *Hum Mol Genet* 1993; **2**: 1123–1128.
11 Greenwood CM, Sun S, Veenstra J *et al*: How old is this mutation? – a study of three Ashkenazi Jewish founder mutations. *BMC Genet* 2010; **11**: 39.
12 'Slav' in Encyclopædia Britannica: *Encyclopædia Britannica Online* 2010. Accessed 14 June 2010; http://www.britannica.com/EBchecked/topic/548156/Slav.
13 Parmele MP: *A Short History of Russia*. Middlesex: The Echo Library, 2008.
14 Barford PM: *The Early Slavs*. Ithaca: Cornell University Press, 2001.
15 Viking Age Europe - 793-1050 In: *Concise Atlas of World History Andromeda* 1997. Accessed 4 May 2010; http://www.credoreference.com/entry/andawh/viking_age_europe_793_1050.
16 'Vikings' in Encyclopedia of the Middle Ages. André Vauchez (eds): *Encyclopedia of the Middle Ages (e-reference edition)*. James Clarke & Co, 2001. Accessed 4 May 2010; http://www.oxfordreference.com/.
17 James Clarke & Co: Mathieu Arnoux 'Normandy' in Encyclopedia of the Middle Ages. André Vauchez (eds): *Encyclopedia of the Middle Ages (e-reference edition)*. James Clarke & Co, 2001. Accessed 4 May 2010; http://www.oxfordreference.com/.

18 'Norman Conquest' in World Encyclopedia. *Oxford Reference Online* 2008. Accessed 4 May 2010; http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t142.e8198.

19 Tonin PN: [The limited spectrum of pathogenic BRCA1 and BRCA2 mutations in the French Canadian breast and breast-ovarian cancer families, a founder population of Quebec, Canada]. *Bull Cancer* 2006; **93**: 841–846.

20 Dubnow SN: History of the Jews in Russia and Poland; Translated from Russian by Friedlander, I. Bergenfield: Avotaynu Inc., 2000, vol 1.

21 Soegaard M, Kjaer SK, Cox M *et al*: BRCA1 and BRCA2 mutation prevalence and clinical characteristics of a population-based series of ovarian cancer cases from Denmark. *Clin Cancer Res* 2008; **14**: 3761–3767.

22 Malander S, Ridderheim M, Masback A *et al*: One in 10 ovarian cancer patients carry germ line BRCA1 or BRCA2 mutations: results of a prospective study in Southern Sweden. *Eur J Cancer* 2004; **40**: 422–428.

23 Sorlie T, Andersen TI, Bukholm I, Borresen-Dale AL: Mutation screening of BRCA1 using PTT and LOH analysis at 17q21 in breast carcinomas from familial and non-familial cases. *Breast Cancer Res Treat* 1998; **48**: 259–264.

24 Papelard H, de Bock GH, van ER *et al*: Prevalence of BRCA1 in a hospital-based population of Dutch breast cancer patients. *Br J Cancer* 2000; **83**: 719–724.

25 Suspitsin EN, Sherina NY, Ponomariova DN *et al*: High frequency of BRCA1, but not CHEK2 or NBS1 (NBN), founder mutations in Russian ovarian cancer patients. *Hered Cancer Clin Pract* 2009; **7**: 5.

26 Bermisheva MA, Zinnamullina GF, Gantsev S *et al*: [Frequency of 5382insC mutation of the BRCA1 gene]. *Vopr Onkol* 2008; **54**: 31–33.

27 Smirnova TY, Pospekhova NI, Lyubchenko LN *et al*: High incidence of mutations in BRCA1 and BRCA2 genes in ovarian cancer. *Bull Exp Biol Med* 2007; **144**: 83–85.

28 Jasinska A, Krzyzosiak WJ: Prevalence of BRCA1 founder mutations in western Poland. *Hum Mutat* 2001; **17**: 75.

29 Menkiszak J, Gronwald J, Gorski B *et al*: Hereditary ovarian cancer in Poland. *Int J Cancer* 2003; **106**: 942–945.

30 Gorski B, Cybulski C, Huzarski T *et al*: Breast cancer predisposing alleles in Poland. *Breast Cancer Res Treat* 2005; **92**: 19–24.

31 Romanowicz-Makowska H, Smolarz B, Zadrozny M, Kulig A: Analysis of RAD51 polymorphism and BRCA1 mutations in Polish women with breast cancer. *Exp Oncol* 2006; **28**: 156–159.

32 Csokay B, Tihomirova L, Stengrevics A, Sinicka O, Olah E: Strong founder effects in BRCA1 mutation carrier breast cancer patients from Latvia. Mutation in brief no. 258. Online. *Hum Mutat* 1999; **14**: 92.

33 Janavicius R, Pepalyte I, Kucinskas V: Novel and common BRCA1 mutations in familial breast/ovarian cancer patients from Lithuania. *Breast Cancer Res Treat* 2009; **117**: 467–469.

34 Zikan M, Pohlreich P, Stribrna J: Mutational analysis of the BRCA1 gene in 30 Czech ovarian cancer patients. *J Genet* 2005; **84**: 63–67.

35 Backe J, Hofferbert S, Skawran B *et al*: Frequency of BRCA1 mutation 5382insC in German breast cancer patients. *Gynecol Oncol* 1999; **72**: 402–406.

36 van der Looij M, Szabo C, Besznyak I *et al*: Prevalence of founder BRCA1 and BRCA2 mutations among breast and ovarian cancer patients in Hungary. *Int J Cancer* 2000; **86**: 737–740.

37 Armaou S, Pertesi M, Fostira F *et al*: Contribution of BRCA1 germ-line mutations to breast cancer in Greece: a hospital-based study of 987 unselected breast cancer cases. *Br J Cancer* 2009; **101**: 32–37.

38 Kataki A, Gomatos I, Pararas N *et al*: Identification of germline BRCA1 and BRCA2 genetic alterations in Greek breast cancer moderate-risk and low-risk individuals – correlation with clinicopathological data. *Clin Genet* 2005; **67**: 322–329.

39 Yazici H, Glendon G, Yazici H *et al*: BRCA1 and BRCA2 mutations in Turkish familial and non-familial ovarian cancer patients: a high incidence of mutations in non-familial cases. *Hum Mutat* 2002; **20**: 28–34.

40 Manguoglu AE, Luleci G, Ozcelik T *et al*: Germline mutations in the BRCA1 and BRCA2 genes in Turkish breast/ovarian cancer patients. *Hum Mutat* 2003; **21**: 444–445.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)