An original phylogenetic approach identified mitochondrial haplogroup T1a1 as inversely associated with breast cancer risk in BRCA2 mutation carriers.

Blein, Sophie; Bardel, Claire; Danjean, Vincent; McGuffog, Lesley; Healey, Sue; Barrowdale, Daniel; Lee, Andrew; Dennis, Joe; Kuchenbaecker, Karoline B; Soucy, Penny; Terry, Mary Beth; Chung, Wendy K; Goldgar, David E; Buys, Saundra S; Janavicius, Ramunas; Tihomirova, Laima; Tung, Nadine; Dorfling, Cecilia M; van Rensburg, Elizabeth J; Neuhausen, Susan L; Ding, Yuan Chun; Gerdes, Anne-Marie; Ejlertsen, Bent; Nielsen, Finn C; Hansen, Thomas Vo; Osorio, Ana; Benitez, Javier; Andrés-Conejero, Raquel; Segota, Ena; Weitzel, Jeffrey N; Thelander, Margo; Peterlongo, Paolo; Radice, Paolo; Pensotti, Valeria; Dolcetti, Riccardo; Bonanni, Bernardo; Peissel, Bernard; Zaffaroni, Daniela; Scuvera, Giulietta; Manoukian, Siranoush; Varesco, Liliana; Capone, Gabriele L; Papi, Laura; Ottini, Laura; Yannoukakos, Drakoulis; Konstantopoulou, Irene; Garber, Judy; Hamann, Ute; Donaldson, Alan; Brady, Angela

Published in:
Breast Cancer Research

DOI:

2015

Link to publication

Citation for published version (APA):
An original phylogenetic approach identified mitochondrial haplogroup T1a1 as inversely associated with breast cancer risk in BRCA2 mutation carriers


* Correspondence: david.cox@lyon.unicancer.fr
Deceased
1INSERM U1052, CNRS UMR5286, Université Lyon 1, Centre de Recherche en Cancérologie de Lyon, Lyon, France
2Université de Lyon, 69000 Lyon, France
Full list of author information is available at the end of the article

© 2015 Blein et al; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Abstract

Introduction: Individuals carrying pathogenic mutations in the BRCA1 and BRCA2 genes have a high lifetime risk of breast cancer. BRCA1 and BRCA2 are involved in DNA double-strand break repair, DNA alterations that can be caused by exposure to reactive oxygen species, a main source of which are mitochondria. Mitochondrial genome variations affect electron transport chain efficiency and reactive oxygen species production. Individuals with different mitochondrial haplogroups differ in their metabolism and sensitivity to oxidative stress. Variability in mitochondrial genetic background can alter reactive oxygen species production, leading to cancer risk. In the present study, we tested the hypothesis that mitochondrial haplogroups modify breast cancer risk in BRCA1/2 mutation carriers.

Methods: We genotyped 22,214 (11,421 affected, 10,793 unaffected) mutation carriers belonging to the Consortium of Investigators of Modifiers of BRCA1/2 for 129 mitochondrial polymorphisms using the iCOGS array. Haplogroup inference and association detection were performed using a phylogenetic approach. ALTree was applied to explore the reference mitochondrial evolutionary tree and detect subclades enriched in affected or unaffected individuals.

Results: We discovered that subclade T1a1 was depleted in affected BRCA2 mutation carriers compared with the rest of clade T (hazard ratio (HR) = 0.55; 95% confidence interval (CI), 0.34 to 0.88; \( P = 0.01 \)). Compared with the most frequent haplogroup in the general population (that is, H and T clades), the T1a1 haplogroup has a HR of 0.62 (95% CI, 0.40 to 0.95; \( P = 0.03 \)). We also identified three potential susceptibility loci, including G13708A/rs28359178, which has demonstrated an inverse association with familial breast cancer risk.

Conclusions: This study illustrates how original approaches such as the phylogeny-based method we used can empower classical molecular epidemiological studies aimed at identifying association or risk modification effects.

Introduction

Breast cancer is a multifactorial disease with genetic, lifestyle and environmental susceptibility factors. Approximately 15% to 20% of the familial aggregation of breast cancer is accounted for by mutations in high-penetrance susceptibility genes [1-3], such as BRCA1 and BRCA2. Pathogenic mutations in BRCA1 and BRCA2 confer lifetime breast cancer risk of 60% to 85% [4,5] and 40% to 85% [4,5], respectively. Other genomic variations (for example, in genes encoding proteins interacting with BRCA1 and BRCA2) have been identified as modifiers of breast cancer risk and increase or decrease the risk initially conferred by BRCA1 or BRCA2 mutation [6].

BRCA1 and BRCA2 are involved in DNA repair mechanisms, including double-strand break (DSB) repair by homologous recombination [7,8]. DSBs are considered to be among the most deleterious forms of DNA damage because the integrity of both DNA strands is compromised simultaneously. These breaks can lead to genomic instability resulting in translocations, deletions, duplications or mutations when not correctly repaired [9]. Reactive oxygen species (ROS) are one of the main causes of DSBs, along with exposure to ionizing radiation, various chemical agents and ultraviolet light [10].

ROS are naturally occurring chemical derivatives of metabolism. Elevated levels of ROS and downregulation of ROS scavengers and/or antioxidant enzymes can lead to oxidative stress, which is associated with a number of human diseases, including various cancers [11]. The electron transport chain process, which takes place in the mitochondria, generates the majority of ROS in human cells. Variations in the mitochondrial genome have been shown to be associated with metabolic phenotypes and
oxidative stress markers [12]. Mitochondrial dysfunction recently was shown to promote breast cancer cell migration and invasion through the accumulation of a transcription factor, hypoxia-inducible factor 1α, via increased production of ROS [13].

Human mitochondrial DNA (mtDNA) has undergone a large number of mutations that have segregated during evolution. Those changes are now used to define mitochondrial haplogroups. Some of these changes slightly modify metabolic performance and energy production; thus, not all haplogroups have identical metabolic capacities [14]. It has been hypothesized that the geographic distribution of mitochondrial haplogroups results from selection of metabolic capacities driven mainly by adaptation to climate and nutrition [15,16].

Mitochondrial haplogroups have been associated with diverse multifactorial diseases, such as Alzheimer’s disease [17], hypertrophic cardiomyopathy [18], retinal diseases [19] or age-related macular degeneration [20]. Variations in mtDNA have also been linked to several types of cancer, such as gastric cancer [21] or renal cell carcinoma [22]. Interestingly, variations in mtDNA have been linked to several types of female cancers, including endometrial [23], ovarian [24] and breast cancer [25,26]. A recent study underlined the possibility that mtDNA might be involved in the pathogenic and molecular mechanisms of familial breast cancer [27].

The Collaborative Oncological Gene-environment Study (COGS) is a European project designed to improve understanding of genetic susceptibility to breast, ovarian and prostate cancer. This project involves several consortia: the Breast Cancer Association Consortium (BCAC) [29], the Ovarian Cancer Association Consortium [30], the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) [31] and the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) [32]. CIMBA is a collaborative group of researchers working on genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers. As part of the COGS project, more than 200,000 single-nucleotide polymorphisms (SNPs) were genotyped for BRCA1 and BRCA2 female mutation carriers on the iCOGS chip, including 129 mitochondrial polymorphisms. The iCOGS chip is a custom Illumina “Infinium genotyping array (Illumina, San Diego, CA, USA) designed to test, in a cost-effective manner, genetic variants related to breast, ovarian and prostate cancers.

In this study, we explored mitochondrial haplogroups as potential modifiers of breast cancer risk in women carrying pathogenic BRCA1 or BRCA2 mutations. Our study includes females diagnosed with breast cancer and unaffected carriers belonging to CIMBA. We used an original analytic phylogenetics-based approach implemented in a homemade algorithm and in the program ALTree [33,34] to infer haplogroups and to detect associations between haplogroups and breast cancer risk.

Methods

Ethics statement
A signed informed written consent form was obtained from all participants. All contributing studies involved in CIMBA received approvals from the institutional review committees at their host institutions. Ethical committees that approved access to the data analyzed in this study are listed in Additional file 1.

BRCA1 and BRCA2 mutation carriers
Final analyses included 7,432 breast cancer cases and 7,104 unaffected BRCA1 mutation carriers, as well as 3,989 invasive breast cancer and 3,689 unaffected BRCA2 mutation carriers, all belonging to CIMBA. Supplementary specifications regarding inclusion profiles and studies belonging to CIMBA are available in the reports by Couch et al. [35] and Gaudet et al. [36]. All analyses were conducted separately on CIMBA BRCA1 and BRCA2 mutation carriers (abbreviated pop1 and pop2, respectively). Eligible female carriers were aged 18 years or older and had a pathogenic mutation in BRCA1 and/or BRCA2. Women with both BRCA1 and BRCA2 mutations were included in downstream analyses. Data were available for year of birth, age at study recruitment, age at cancer diagnosis, BRCA1 and BRCA2 mutation carriers, respectively. Information regarding mastectomy was incomplete and was therefore not used as an inclusion or exclusion parameter.

Genotyping and quality filtering
Genotyping was conducted using the iCOGS custom Illumina Infinium array. Data from this array are available to the scientific community upon request. Please see [37] for more information. Genotypes were called using Illumina’s proprietary GenCall algorithm. Genotyping and quality filtering were described previously [35,36]. Initially, 129 mitochondrial SNPs were genotyped for both BRCA1 and BRCA2 mutation carriers. SNPs fulfilling the following criteria were excluded from downstream analyses: monoallelic SNPs (minor allele frequency = 0), SNPs with more than 5% data missing, annotated as triallelic, or having probes cross-matching with the nuclear genome. Heterozygous genotypes were removed from analyses, and we further filtered out SNPs having more than 5% of heterozygous calls to limit the potential for heteroplasmia affecting our results. We also did not retain SNPs representing private mutations. These mutations are rare, often restricted to a few families, and not sufficiently prevalent in the general population to be included in the reference
mitochondrial evolutionary tree (see below). This last step of filtration yielded 93 and 92 SNPs for the pop1 and pop2 analyses, respectively (see Additional file 2). Only individuals with fully defined haplotypes (that is, non-missing genotypes for the 93 and 92 SNPs selected for pop1 and pop2, respectively) were included in downstream analyses (14,536 and 7,678 individuals, respectively).

Mitochondrial genome evolution and haplogroup definition
Analyses were based on the theoretical reconstructed phylogenetic tree of the mitochondrial genome (mtTree) known as PhyloTree [38] (v.15). The mtTree is rooted by the Reconstructed Sapiens Reference Sequence (RSRS).

RSRS has been identified as the most likely candidate to root the mtTree by refining human mitochondrial phylogeny by parsimony [39]. Each haplogroup in mtTree is defined by the set of mtDNA SNPs that have segregated in RSRS until today in the mitochondrial genome. Each haplogroup is fully characterized by the 16,569-bp sequence resulting from the application of all the substitutions that are encoded by the corresponding SNPs in the RSRS sequence.

Haplogroups imputation
The phylogenetic approach used to infer haplogroups is described in Figure 1. Mitochondrial genome sequences can be reconstructed at each node of mtTree, given the

Figure 1 Simplified representation of the phylogenic method used to infer haplogroups. (a) Full-length haplotypic sequences are reconstructed at each node of the reference tree. (b) Haplotypes are then restricted to available loci. Sequences of the same color are identical. (c) Unique short haplotypes are matched directly with the corresponding haplogroup. (d) Sequences that match with several haplogroups are associated with their most recent common ancestor haplogroup. RSRS, Reconstructed Sapiens Reference Sequence.
substitutions that have segregated in RSRS. Each haplogroup therefore has a corresponding full-length mitochondrial sequence. However, the full-length mitochondrial sequence is not available in the data, because the iCOGS platform captured only 93 and 92 SNPs for pop1 and pop2, respectively. Thus, for each of the 7,864 nodes of the phylogenetic tree, the corresponding short haplotype (that is, the full-length sequence restricted to available loci) was defined. Some of the short haplotypes are unique, and they can be matched with their corresponding haplogroup directly. However, most of the time, given the small number of SNPs analyzed, several haplogroups correspond to the same short haplotype. Consequently, a unique haplogroup cannot confidently be assigned to each short haplotype. Therefore, each short haplotype was assigned the most recent common ancestor of all the haplogroups that share the same short haplotype. Once this matching was done, short haplotypes were reconstructed in the same way for each individual in our dataset and were assigned the corresponding haplogroup. The accuracy of the method used was assessed by application to a set of 630 mtDNA sequences of known European and Caucasian haplogroups (see Additional file 3).

Association detection
This phylogenetic approach is based on the identification of subclades in the reference phylogenetic tree of the mitochondrial genome differentially enriched for cases and unaffected controls compared with neighboring subclades. We used ALTree [33,34] to perform association testing. ALTree—standing for Association detection and Localization of susceptibility sites using haplotype phylogenetic Trees—is an algorithm used to perform nested homogeneity tests to compare distributions of affected and unaffected individuals in the different clades of a given phylogenetic tree. The objective is to detect if some clades of a phylogenetic tree are more or less enriched in affected or unaffected individuals compared with the rest of the tree. There are as many tests performed as there are levels in the phylogenetic tree. The P-value at each level of the tree is obtained by a permutation procedure in which 1,000 permutations are performed. Individual labels (“affected” or “unaffected”) are permuted 1,000 times to see to what extent the observed distribution of affected or unaffected is different from a random distribution. A procedure to correct for multiple testing adapted to nested tests [40] is implemented in ALTree. The objective of ALTree is to detect an enrichment difference at the level of the whole tree. To conserve computational time and resources, only the most significant P-value obtained for all tests performed on one tree is corrected.

Handling genetic dependency
ALTree is used to perform homogeneity tests to detect differences in enrichment or depletion of affected or unaffected individuals between clades in the phylogenetic tree. This kind of test can be performed only on independent data. However, because some individuals in the CIMBA dataset belong to the same family, we constructed datasets with genetically independent data by randomly selecting one individual from among all those belonging to the same family and sharing the same short haplotype. To take into account the full variability of our data, we resampled 1,000 times. The results of the analysis pipeline are obtained for each resampling independently and then averaged over the 1,000 resamplings to obtain final results.

Character reconstruction at ancestral nodes
Before the ALTree localization algorithm was launched, ancestral sequences were reconstructed at each internal tree node; that is, short haplotypes were inferred with maximum likelihood at all nodes that were not leaves. We used the software PAML [41] to perform the reconstruction at ancestral nodes using a maximum likelihood method. The phylogeny model used was the general time-reversible model (either GTR or REV).

Localization of susceptibility sites
ALTree also includes an algorithm used to identify which sites are the most likely ones to be involved in the association detected. For each short haplotype observed, the ALTree add-on altree-add-S adds to the short haplotype sequence a supplementary character called S, which represents the disease status associated with this short haplotype. Are individuals carrying this short haplotype more often affected or unaffected? S is calculated based on the affected and unaffected counts, the relative proportion of affected and unaffected in the whole dataset, and sensibility parameter e. e was set to its default value, which is 1. After S character computation, haplotypes including character S are reconstructed at ancestral nodes. Susceptibility site localization is achieved with ALTree by computing a correlated evolution index calculated between each change of each site and the changes of the character S in the two possible directions of change. The sites whose evolution are the most correlated with the character S are the most likely susceptibility sites.

Selected subclades
The analyses were carried out on the full evolutionary tree. However, the more haplogroups there are at each level, the less statistical power homogeneity tests have. Therefore, analyses were also applied to subclades extracted from the tree. Subclades were defined using counts of individuals in each haplogroup of the clade to maximize statistical power. The chosen subclades
and corresponding affected and unaffected counts are presented in Table 2.

Statistical analysis
We quantified the effect associated with enrichment discovered by applying ALTree by building a weighted Cox regression in which the outcome variable is the status (affected or non-affected) and the explicative variable is the inferred haplogroup. Analyses were stratified by country. Data were restricted to the clades of interest. The uncertainty in haplogroup inference was not taken into account in the model. The weighting method used takes into account breast cancer incidence rate as a function of age [42] and the gene containing the observed pathogenic mutation (that is, BRCA1 or BRCA2). Familial dependency was handled by using a robust sandwich estimate of variance (R package survival, cluster() function).

Results

Haplogroup imputation
In Additional file 4, absolute and relative frequencies are recapitulated for each haplogroup imputed in BRCA1 and BRCA2 mutation carriers. For BRCA1 mutation carriers, we reconstructed 489 distinct short haplotypes of 93 loci from the genotypes data. Only 162 of those 489 short haplotypes matched theoretical haplotypes reconstructed in the reference mitochondrial evolutionary tree. These 162 haplotypes represented 13,315 of 14,536 individuals. Thus, 91.6% of BRCA1 mutation carriers were successfully assigned a haplogroup. For BRCA2 mutation carriers, we reconstructed 350 distinct short haplotypes of 92 loci from our genotype data. Only 139 of those 350 short haplotypes matched theoretical haplotypes reconstructed in the reference mitochondrial evolutionary tree. These 139 haplotypes represented 6,996 of 7,678 individuals. Thus, 91.1% of BRCA2 mutation carriers were successfully assigned a haplogroup. Because more BRCA1 than BRCA2 mutation carriers were genotyped (14,536 vs. 7,678 individuals), we logically observed more distinct haplotypes in pop1 than in pop2 (489 vs. 350 haplotypes).

The accuracy of the main haplogroup inference method used was estimated at 82% and reached 100% for haplogroups I, J, K, T, U, W and X. Given the set of SNPs we disposed of, our method has difficulty differentiating between H and V haplogroups (see Additional file 3).

Association results
For both populations of BRCA1 or BRCA2 mutation carriers, as well as for the full tree as for all selected subclades (see Table 1), we extracted the mean corrected P-values for association testing over all resamplings performed (see Table 2). The only corrected P-value that remained significant was that obtained for subclade T (abbreviated T*) in the population of individuals of BRCA2 mutation carriers (P = 0.04).

The phylogenetic tree of subclade T (see Figure 2a) contains only three levels; thus, only three tests were performed within this clade. Raw P-values were examined to determine at which level of the tree ALTree detects a difference of enrichment in affected or unaffected individuals (see Table 3). Only the P-value associated with the test performed at the first level of the tree is significant. We looked more closely at the mean frequencies of affected and unaffected individuals in the tree at this level (see Figure 2b). In the T11a1 subclade, the mean count of affected and unaffected are 32 and 47, respectively. In the T2* subclade, we observed, on average, 217 and 148 affected and unaffected individuals, respectively, whereas in the T subclade, we observed, on average, 13 and 11 affected and unaffected individuals, respectively. The ranges observed for each of these values over the 1,000 resamplings are represented in Figure 2b. On the basis of these observations, we conclude that subclade T11a1 is depleted in affected carriers compared with the neighboring subclades T and T2.

Table 1 Counts of participants in selected subclades

<table>
<thead>
<tr>
<th>Subclade</th>
<th>BRCA1 mutation carriers</th>
<th>BRCA2 mutation carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>U8</td>
<td>1,458</td>
<td>863</td>
</tr>
<tr>
<td>T</td>
<td>1,243</td>
<td>651</td>
</tr>
<tr>
<td>J</td>
<td>1,270</td>
<td>630</td>
</tr>
<tr>
<td>J1</td>
<td>1,043</td>
<td>513</td>
</tr>
<tr>
<td>H</td>
<td>3,706</td>
<td>1,967</td>
</tr>
<tr>
<td>H1</td>
<td>582</td>
<td>337</td>
</tr>
<tr>
<td>U5</td>
<td>868</td>
<td>458</td>
</tr>
<tr>
<td>X1’2’3’</td>
<td>221</td>
<td>103</td>
</tr>
<tr>
<td>K1a</td>
<td>608</td>
<td>364</td>
</tr>
</tbody>
</table>

Table 2 Mean corrected P-values for association testing with ALTree

<table>
<thead>
<tr>
<th>Subclade</th>
<th>pop1 corrected P-value</th>
<th>pop2 corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>0.830</td>
<td>0.681</td>
</tr>
<tr>
<td>U8</td>
<td>0.146</td>
<td>0.626</td>
</tr>
<tr>
<td>T</td>
<td>0.285</td>
<td>0.040</td>
</tr>
<tr>
<td>J</td>
<td>0.718</td>
<td>0.112</td>
</tr>
<tr>
<td>J1</td>
<td>0.621</td>
<td>0.150</td>
</tr>
<tr>
<td>H</td>
<td>0.747</td>
<td>0.930</td>
</tr>
<tr>
<td>H1</td>
<td>0.268</td>
<td>0.804</td>
</tr>
<tr>
<td>U5</td>
<td>0.829</td>
<td>0.747</td>
</tr>
<tr>
<td>X’1’2’3’</td>
<td>0.416</td>
<td>0.629</td>
</tr>
<tr>
<td>K1a</td>
<td>0.170</td>
<td>0.162</td>
</tr>
</tbody>
</table>

*pop1, BRCA1 mutation carrier; pop2, BRCA2 mutation carrier. Bold indicates a significant P-value.
Localization results
We performed a localization analysis with ALTree. The correlated evolution index for all non-monomorphic sites observed in short haplotype sequences of subclade T are displayed in Additional file 5. The higher the correlated evolution index, the more likely it is that corresponding sites will be involved in the observed association. Three short haplotype sites numbered 44, 57 and 72 and corresponding to SNPs T988C, G11812A/rs4154217 and G13708A/rs28359178, respectively, clearly distinguish themselves, with correlation index values of 0.390, 0.324 and 0.318, respectively, whereas the correlation index values of all other sites ranged from −0.270 to −0.101. Table 4 shows the details for these three loci.

Effect quantification
The ALTree method is able to detect an association, but cannot to quantify the associated effect. We estimated the risk of breast cancer for individuals with the T1a1 haplogroup compared with individuals with another T subclade haplogroup in the population of BRCA2 mutation carriers using a more classical statistical method, a weighted Cox regression. We found a breast cancer HR of 0.55 (95% CI, 0.34 to 0.88; \( P = 0.014 \)). We also tested haplogroup T1a1 and compared it with other T* haplogroups and the H haplogroup (the main haplogroup in the general population), and we found a breast cancer HR of 0.62 (95% CI, 0.40 to 0.95; \( P = 0.03 \)).

Discussion
We employed an original phylogenetic analytic method, coupled with more classical molecular epidemiologic analyses, to detect mitochondrial haplogroups differentially enriched for affected BRCA1/2 mutation carriers. We successfully inferred haplogroups for more than 90% of individuals in our dataset. After haplogroup imputation, the ALTree method identified T1a1 in the T clade as differentially enriched in affected BRCA2 mutation carriers, whereas no enrichment difference was found for BRCA1 mutation carriers. The T subclade is present in 4% of African populations compared with 11% in Caucasian and Eastern European populations [43]. In our data, the T subclade represented 9.34% of BRCA1 mutation carriers and 9.30% of BRCA2 carriers. The ALTree method also identified three potential breast cancer susceptibility loci in mtDNA. The main goals of using the phylogenetic method we used were to improve statistical power by regrouping subclades according to genetic considerations, to limit the number of tests performed and to precisely quantify this number. ALTree identified three SNPs of interest. Whereas the association we observed could possibly be driven by a single SNP, no difference was observed between multivariate and univariate cox models including the three SNPs identified by ALTree (data not shown).

In this study, we investigated to what extent mtDNA variability modified breast cancer risk in individuals
Table 4 Description of loci identified as potential susceptibility sites by ALTree

<table>
<thead>
<tr>
<th>Site</th>
<th>SNP name</th>
<th>Position</th>
<th>Direction of change</th>
<th>Correlated evolution index</th>
<th>Major allele</th>
<th>Minor allele</th>
<th>MAF in pop2</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>Mito19900C</td>
<td>9,899</td>
<td>T → C</td>
<td>0.390</td>
<td>T</td>
<td>C</td>
<td>0.016</td>
</tr>
<tr>
<td>57</td>
<td>rs41544217</td>
<td>11,812</td>
<td>G → A</td>
<td>0.324</td>
<td>A</td>
<td>G</td>
<td>0.071</td>
</tr>
<tr>
<td>72</td>
<td>rs28359178</td>
<td>13,708</td>
<td>G → A</td>
<td>0.318</td>
<td>G</td>
<td>A</td>
<td>0.111</td>
</tr>
</tbody>
</table>

*MAF, Mean allele frequency; pop2, BRCA2 mutation carrier.

A large proportion of breast cancer heritability still remains unexplained today [44]. Different methods exist to study genomic susceptibility to a disease, such as linkage analyses (which identified the BRCA1 and BRCA2 susceptibility genes) or genome-wide association studies (GWASs). However, classical linkage analysis cannot be applied to the haploid mitochondrial genome. Furthermore, commercial GWAS chips are not adequately capturing the majority of mtDNA SNPs. A non-genome-wide and mtDNA-focused approach was required to explore how mtDNA variability influences breast cancer risk. Here we have shown that BRCA2 mutation carriers with the subclade T1a1 have between 30% and 50% less risk of breast cancer than those with other clades, which, if validated, is a clinically meaningful risk reduction and may influence the choice of risk management strategies.

The association we observed among BRCA2, but not BRCA1, mutation carriers may reveal a functional alteration that would be specific to mechanisms involving BRCA2-related breast cancer. Today, it is established that BRCA1- and BRCA2-associated breast cancers are not phenotypically identical. These two types of tumors do not harbor the same gene expression profiles or copy number alterations [45]. Breast cancer risk modifiers in BRCA1/2 mutation carriers have already been identified [46]. However, most of them are specific from one or the other type of mutation carried [47]. It is therefore not surprising that this observation is observed in BRCA2 mutation carriers only.

Our inability to assign haplogroups to 9% of study participants could have three main explanations. First, given the high mutation rate in the mitochondrial genome, observed combinations of mtDNA SNPs might have appeared relatively recently in the general population, and the corresponding haplotypes might not yet be incorporated into PhyloTree. Second, only one genotyping error could lead to chimeric haplotypes that do not exist, although, given the quality of our genotyping data, this is unlikely. Third, the mitochondrial reference evolutionary tree PhyloTree is based on phylogeny reconstruction by parsimony, and, for some subclades, it might be suboptimal, especially for haplogroups relying on few mitochondrial sequences, as is the case for African haplogroups [48]. In cases of uncertainty, the choice we made to assign the most recent common ancestor to the studied haplotype enabled us to improve statistical power without introducing a bias in the detected association. For the association detected between T, T1* and T2* subclades, the haplogroup inference method used did not bias the counts of affected and unaffected individuals in these subclades. More details are presented in Additional file 6. Furthermore, on the basis of the haplogroup inference with our method of 630 European and Caucasian mtDNA sequences whose haplogroup is known, we successfully assigned the correct main haplogroup and subhaplogroup of 100% of sequences belonging to T, T2* and T1a1* haplogroups.

We quantified the effect corresponding to the detected association by using a more classical approach. We built a weighted Cox regression including inferred haplogroup as an explicative variable. However, the uncertainty in haplogroup inference was not taken into account in this model. Nevertheless, based on haplogroup assignment and regrouping performed in clade T, affected and unaffected counts of individuals in this clade were not biased.

With only 129 loci genotyped over the 16,569 nucleotides composing the mitochondrial genome, we certainly did not explore the full variability of mitochondrial haplotypes. A characterization of individual mitochondrial genomes would require more complete data acquisition methods to be used, such as next-generation sequencing. However, next-generation sequencing has its own limits and challenges, because some regions of the mitochondrial genome are not easily mappable, owing to a high homology with the nuclear genome, among other factors, and important bioinformatics treatment is necessary to overcome sequencing technology biases. Finally, even for a relatively short genome of “only” 16,569 bp, mtDNA sequencing of more than 20,000 individuals would represent a major increase in cost relative to genotyping 129 SNPs.

ALTree identified T9899C, G11812A/rs41544217 and G13708A/rs28359178 as three potential susceptibility sites for the discovered association (see Additional file 7). These three SNPs are located in the coding part of genes MT-CO3, MT-ND4 and MT-ND5, respectively. When looking at PhyloTree, T9899C seems to be involved in T1 subclade definition, whereas G13708A and A11812G are involved in T2 subclade definition. Whereas T98899C and
G11821/rs41544217 are synonymous SNPs, G10398A leads to a change of amino acid in the final protein (from alanine to threonine). These two synonymous SNPs have never been described in a disease context in the literature. G13708A is also known for being a secondary mutation for Leber’s hereditary optic neuropathy (LHON) and multiple sclerosis [49]. Although the role of secondary mutations in LHON is still controversial, G13708A could be associated with impairment of the respiratory chain in this pathology. G13708A has also been described as a somatic mutation in a breast cancer tumor, whereas it was not present in adjacent normal tissue or in blood leukocytes [50]. A high proportion of mitochondrial somatic tumor-specific variants are also known mtDNA SNPs, which is consistent with the hypothesis that tumor cells are prone to acquire the same mutations that segregate into mtDNA by selective adaptation when humans migrated out of Africa and confronted new environments [51]. Interestingly, the germline variant G13708A has already been shown to be inversely associated with familial breast cancer risk (with the same direction of the association), with a breast cancer odds ratio of 0.47 (95% CI, 0.24 to 0.92) [52]. None of these SNPs have been described in the context of ovarian cancer.

The corrected P-value obtained using ALTREE in studying clade T is 0.02, which is not highly significant. A replication step should be performed to validate these results. However, it will be difficult to include enough women in this replication step, given the specific profile studied here. In fact, the estimations of BRCA2 pathogenic mutations in the general population range from 0.068% [5] to 0.69% [53]. T1a1 represents only a small percentage of European haplogroups (from 1% to 2%). The number of women who have this association is therefore low. However, women carrying such mutations are confronted with drastic choices regarding the prevention of breast cancer, notably prophylactic mastectomy or complete hysterectomy. If breast cancer risk is really reduced by a factor of 2 for women with T1a1, this could be an important fact to take into account for breast cancer prevention.

Conclusions
This study and our results suggest that mitochondrial haplogroup T1a1 may modify the individual breast cancer risk in BRCA2 mutation carriers. For now, this observation cannot be extended to the general population. Further investigation of the biological mechanism behind the associations we observed may further reinforce the hypothesis that the mitochondrial genome is influential in breast cancer risk, particularly among carriers of BRCA2 mutations, and, if validated, is of a level to influence cancer risk management choices.

Additional files

**Additional file 1:** List of ethical committees that approved the access to the data analyzed in this study.

**Additional file 2:** SNPs selected for downstream analyses.

**Additional file 3:** Description and results of the procedure used to estimate the accuracy of our haplogroup inference methodology.

**Additional file 4:** Absolute and relative frequencies of imputed haplogroups by population. Table containing absolute and relative frequencies of imputed haplogroups for BRCA1 and BRCA2 mutation carriers.

**Additional file 5:** Correlated evolution index for all non-monomorphic sites observed in short haplotype sequences of subclade T. Table containing correlated evolution index for all non-monomorphic sites observed in short haplotypes sequences of subclade T.

**Additional file 6:** Details of haplogroups inference results for subclade T.

**Additional file 7:** Methods used to compute coevolution index.

**Abbreviations**

BCAC: Breast Cancer Association Consortium; CI: Confidence interval; CIMBA: Consortium of Investigators of Modifiers of BRCA1/2; COGS: Collaborative Oncological Gene-environment Study; DSB: Double-strand break; GWAS: Genome-wide association study; HR: Hazard ratio; LHON: Leber’s hereditary optic neuropathy; MAF: Mean allele frequency; mtDNA: Mitochondrial DNA; mtTree: Phylogenetic tree of the mitochondrial genome; OCAC: Ovarian Cancer Association Consortium; pop1: BRCA1 mutation carrier; pop2: BRCA2 mutation carrier; PRACTICAL: Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; ROS: Reactive oxygen species; RSRS: Reconstructed Sapiens Reference Sequence; SNP: Single-nucleotide polymorphism.

**Competing interests**
The authors declare that they have no competing interests.

**Authors’ contributions**

**Acknowledgements**

Collaborative Oncological Gene-environment Study (COGS): This study would not have been possible without the contributions of the following: Per Hall (COGS); Kyriaki Michailidou, Manjeet K. Bolla and Qin Wang (Breast Cancer Association Consortium (BCAC)); Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai and Sara Benlloch (PRACTICAL); Alison M. Dunlop, Craig Luccarini, Michael Lush and the staff of the McGill University and Genôme Québec Innovation Centre; and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo College Genotyping Core Facility.

Collaborative Investigators of Modifiers of BRCA1/2 (CIMBA): Maggie Angelakos, Judi Maskiell, Gillian Dite and Helen Tsimiklis; members of and participants in the New York site of the Breast Cancer Family Registry;
members of and participants in the Ontario Familial Breast Cancer Registry for their contributions to the study; Vilis Rudatis, Laimonas Grilkevičius, Drs Janis Eglitis, Anna Kiltova and Aivars Stengrevics; the families who contribute to the BRCA-gene mutations and breast cancer in South African women (BMBSA) study; Chun Ding and Linda Steele; Alicia Barroso, Rosario Alonso, Guillermo Pita, Alessandra Viel and Lara delia Puppa of the Centro di Riferimento Oncologico, IRCCS, Aviano (PN), Italy; Laura Papi of the University of Florence, Florence, Italy; Monica Barile of the Istituto Europeo di Oncologia, Milan, Italy; Liliana Varesco of the IRCCS AOU San Martino – IST Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy; Stefania Torsmarsi, Brunella Pilato and Rossanna Lambo of the Istituto Nazionale Tumori “Giovanni Paolo II” – Bari, Italy; Aline Martayan of the Istituto Nazionale Tumori Regina Bina, Rome, Italy; Maria Grazia Tibilitti of the Ospedale di Circolo-Università dell’Insubria, Varese, Italy; and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy.

Czech Republic); Michal Zíkan, Petr Pohorecký and Zdeněk Klebl (Oncogeneticologic Center and Department of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic); Anne Lincoln and Lauren Jacobs; the National Israeli Cancer Control Center (NICCC) National Familial Cancer Consultation Service team led by Sara Dishon; the laboratory team led by Dr Flavio Letèlicher; and the research field operations team M. Ma Pinchoy; members and participants in the Ontario Cancer Genetics Network for their contributions to the study. Leigha Senter, Kevin Sweet, Caroline Cavine and Michelle O’Connor were instrumental in accrual of study participants, ascertainment of medical records and database management; The Ohio State University Human Genetics Sample Bank; the Merav Breast Cancer team at the Sheba Medical Center, Åke Bor, Håkan Olsson, Helena Jerntorp, Karin Henriksson, Katja Harbst, Maria Soller and UK Krioffferson; Sahlgrenska University Hospital, Gothenburg, Sweden: Anna Överholm, Margareta Nordling, Per Karlsson and Zakaria Einberg; Stockholm and Karolinska University Hospital, Stockholm; Anna von Wachenfeld; Annelie Liljegren, Annika Lindblom, Brita Arver, Gisela Barbany Buxintra and Johanna Ranttä; Umeå University Hospital, Umeå, Sweden: Beatrice Melin, Christina Edwindsdotter Adorff and Monica Emanuelsson; Uppsala University, Uppsala, Sweden: Hanne Persson, Maria Hellstrom Pigged andf Bernard Rosenquist; Linköping University Hospital, Linköping, Sweden: Marie Stenmark-Aksamil and Sigurn Liedgren; Cecilia Zvocer, Qun Niu, and physicians, genetic counselors, research nurses and staff of the Comprehensive Cancer Risk and Prevention Clinic of University of Chicago Medicine, Chicago, IL, USA, for their contributions to this resource; Joyce Seldon, MSGC, and Lorna Rivan, MPH; Dr Robert Nusbaum and the following genetic counselors: Beth Crawford, Kate Loranger, Julie Mak, Nicola Stewart, Robin Lee, Annie Blanco and Peggy Conrad; Salina Chan; Paul DP Pharoah, Simon Gayther, Susan Ramus, Carrie Pye, Patricia Harrington and Eva Wozniak for their contributions to the UK Familial Ovarian Cancer Registry (UFOCRC); Geoffrey Lindeman, Marion Harris, Martin Delatycki of the Victorian Familial Cancer Trials Group; and Sarah Sawyer, Rebecca Driessen and Ella Thompson.

Funding

Higher-level funding: The COGS project is funded through a European Commission Seventh Framework Program grant (agreement number 223175: HEALTH-F2-2009-231737). The COG2M data management and data analysis were supported by Cancer Research UK grants C12292/A11174 and C1287/A10118. SH is supported by a National Health and Medical Research Council (NHMRC) program grant (to GCT). Individual researcher support: ACA is a Cancer Research UK Senior Cancer Research Fellow (C12292/A11174). DFE is a Principal Research Fellow of Cancer Research UK GC, MCS and IC are supported by the National Health and Medical Research Council (NHMRC). BK holds an American Cancer Society Early Detection Professorship (SOP-06-258-01-COLUN). MHG and PLM were supported by funding from the Intramural Research Program of the National Cancer Institute, National Institutes of Health. OIO is an American Cancer Society Clinical Research Professor. JS is Chairholder of the Canada Research Chair in Oncogenetics.

Funding of constituent studies: The Breast Cancer Family Registry (BCFR) was supported by grant UM1 CA164920 from the National Cancer Institute, National Institutes of Health. The content of this article does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government or the BCFR. The Baltic Familial Breast Ovarian Cancer Consortium (BFBOCC) is partly supported by Lithuania (BFBOCC-LT): Research Council of Lithuania grant LIG-07/2012; BFBOCC-LV (Latvia) is partly supported by LSC grant 10.0011.08 and in part by a grant from the European Social Fund number 2009/0220/1DP/1.1.2.0/09/AIP/A/016 and the Liepāja City Council, Liepāja, Latvia; Beth Israel Deaconess Medical Center Cancer Center is supported by the Breast Cancer Research Foundation; BRCA-gene mutations and breast cancer in South African women (BMBSA) was supported by grants from the Cancer Association of South Africa (Cansa) to Elizabeth J van Rensburg; SLN (Bekmark Research Institute, City of Hope, Duarte, CA, USA) was partly supported by the Morris and Horowitz Families Foundation Professorship in Cancer Biology and Outcomes Research; the Copenhagen Breast Cancer Study (EBCS) was supported by the NEYE Foundation; the Spanish National Cancer Research Center (Centro Nacional de Investigaciones Oncológicas (CNIO)) was partly supported by the Spanish Association against Cancer (Asociación Española Contra el Cáncer AECC08), Thematic Network Cooperative Research in Cancer (Red Temática Investigación Cooperativa en Cáncer (RTICC), Centro de Investigación Cáncer, Salamanca, Spain) RTICC 06/0020/1066, Spanish Ministry of Science and Innovation grants FIS PI08 1120 (Fondo de Investigación Sanitaria (FIS)) and SAF2010-20493, and the Foundation Mutua Madrileña (FMM); the City of Hope Clinical Cancer Genetics Community Network and the Hereditary Cancer Research Registry (CON-CGCRN), supported in part by award number RC4A153828 (Principal Investigator: JMM) from the National Cancer Institute and the Office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CONSORZIO Studi Italiani sui Tumori Ereditari Alle Mammella, Italy (CONSIT Team); Funds from Italian citizens who allocated the “5 × 1,000” share of their tax payment in support of the Fondazione RICCS Istituto Nazionale Tumori, according to Italian laws (Istituto Nazionale dei Tumori (INT) institutional strategic project “5 × 1,000”) (to SM); the Italian Association for Cancer Research (AIRC) (to LO); National Centre for Scientific Research “Demokritos” has been cofinanced by the European Union (European Social Fund (ESF)) and Greek national funds through the “Education and Lifelong Learning” operational program of the National Strategic Reference Framework (NSRF) – Research Funding Program of the General Secretariat for Research and Technology: ARISTEA; “Heracleitus II: Investing in knowledge society through the European Social Fund”; the DFKZ study was supported by the Deutsches Krebsforschungszentrum (DFKZ); Epidemiological Study of BRCA1 and BRCA2 Mutation Carriers (EMBRACE) is supported by grants: Cancer Research UK grants C1287/A10118 and C1287/A11190; DGE and FL are supported by a National Institute for Health Research (NIHR) grant to the Biomedical Research Centre, Manchester, UK, the investigators at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust are supported by an NIHR grant to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London; RE and EB are supported by Cancer Research UK grant CS047/ A8358; Kansas University Medical Center: The authors acknowledge support from The University of Kansas Cancer Center (P30 CA168524) and the Kansas Bioscience Authority Eminent Scholar Program; AFG was funded by grants SU01 CA113916 and RO1 CA140323 and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship; The German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC) is supported by German Cancer Aid (grant 109076) (to RKS) and by the Center for Molecular Medicine Cologne (CMMM); GC-HBOC is deeply grateful to Dr Sabine Preisler-Adam for providing information and samples; the GEMO Study was supported by the Ligue Nationale contre le Cancer; the Association “Le cancer du sein, parlons-en!” Award; and the Canadian Institutes of Health Research for the CIHR Team in Familial Risks of Breast Cancer program; G-FAST; KDL is supported by GOA grant BOF10/GOA/19 (Ghent University) and spearhead financing of Ghent University Hospital; the Gynecologic Oncology Group (GOG) was supported by National Cancer Institute grants to the GOG Administrative Office and Tissue Bank (grant CA 27469), the GOG Statistical and Data Center (grant CA 37517) and GOG’s Cancer Prevention and Control Committee (grant CA 101165); HSCC was supported by grants RO1/00369/00012 and 12/00039 from Instituto de Salud Carlos III (ISCII), Madrid, Spain, partially supported by European Regional Development Fund (Fonds européen de développement régional (FEDER)) funds; the Helsinki Breast Cancer Study (HEBCS) was financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266526), the Finnish Cancer Society Study and the Sigrid Juselius Foundation; HEBON is supported by the Dutch Cancer Society grant NK01999-1854, NK02004-3988 and NK02007-3756, the Netherlands Organization of Scientific Research grant NWO 11090024, the Pink Ribbon grant 110005 and Biobanking and Molecular Resource Infrastructure (BBMRI) grant NWO 184.021.007/CP46; HEBON thanks the registration teams of the Comprehensive Cancer Centre Netherlands and Comprehensive Centre South (together the Netherlands Cancer Registry) and PALGA (Dutch Pathology Registry) for part of the data collection, the High Risk Breast Cancer Program (HRBPC) is supported by the Hong Kong Hereditary Breast Cancer Family Registry and the Dr Ellen Li Charitable Foundation, Hong Kong; the Hungarian Breast and Ovarian Cancer Study (HUNBOCS) was supported by Hungarian Research and Technological Innovation Fund (KIA)/Hungarian Scientific Research Fund (Ötiszorgó Tudományos Kutatási Alapprogramok (OTKA)) research grants KTIA-OTKA CK-80745 and KTIA-OTKA K-112288; Institut Català d’Oncologia (ICO) contract grant sponsor: Asociación Española Contra el Cáncer; Spanish Health Research Foundation; Ramón Areces Foundation; Instituto de Salud...
Carlos III (ISCIII); Catalan Health Institute; and Autonomous Government of Catalonia; contract grant numbers ISCIII-RETID06/002/1051, P09/02483, P10/01422, P11/00748, P13/00285, P13/00189 2009SG2090 and P11/00189 2009SG283; the International Hereditary Cancer Center (Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland) was supported in part by the Polish National Science Centre (NCN) grant 2011/03/B/NZ9/04296 and by the Polish Cancer Society (PZKBN); Israeli Cancer Research Foundation (ICRF); and by the Hungarian National Office for Research and Innovation (NKFIH) grant NKFIH125422; the National University Hospital of Iceland was supported by the Icelandic Association “Working for Breast Cancer Research” and by the Landsdalslagi University Hospital Research Fund; The Interdisciplinary Health Research Internal Team Breast Cancer Susceptibility Study (INHERIT) was supported by the Canadian Institutes of Health Research (CIHR) for the “CfHR Team in Familial Risks of Breast Cancer” program, Canadian Breast Cancer Research Alliance grant 019511 and Ministry of Economic Development, Innovation and Export Trade grant PSR-SIRI 701; the Istituto Oncologico Veneto Hereditary Breast and Ovarian Cancer Study (IOVHBOCS) is supported by the Ministero della Salute and a “$ > 1,000” Istituto Oncologico Veneto grant; the Portuguese Oncology Institute–Porto Breast Cancer Study (IPOBACS) was supported in part by Lïga Portuguesa Contra o Cancro; KicenFab is supported by a grant from the National Research Foundation of Korea (2013R1A1A1070659); the Italian National Cancer Institute, Milan, Italy; National Cancer Institute of the United States, Bethesda, MD, USA; National Cancer Research Institute, Cambridge, UK; “Italian Network for Cancer Research (GEMO Study)” is supported by the Italian Association for Cancer Research (AIRC); the American Cancer Society Early Detection Professorship (SO-06-258-01-COUN); GEMO Study: National Cancer Genetics Network UNICANCER Genetic Group, France. The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) coordinating center: Netherlands Cancer Institute, Amsterdam.

Author details

1INSERM U1052, CNRS UMR5286, Université Lyon 1, Centre de Recherche en Cancérologie de Lyon, Lyon, France. 2Université de Lyon, 69000 Lyon, France. 3Université Lyon 1, 69100 Villeurbanne, France. 4UMR CNRS 5558, Laboratoire de Biométrie et Biologie Évolutive (LBBE), “Biométrie et Biologie Évolutive”, Université Claude Bernard Lyon 1, Bâtiment G-gor Mendel, 11 boulevard du 11 novembre 1918, 69622 Villeurbanne, cedex, France. 5Université Grenoble Alpes, UMR 5217, Laboratoire d’Informatique de Grenoble (LIG), équipe-projet Multi-programmation et Ordonnancement sur ressources pour les Applications Interactives de Simulation (MOAIS), 38041 Grenoble, France. 6INRIA Rhône-Alpes, équipe-projet MOAIS, 38334 Saint Ismier, Cédex, France. 7New York for Cancer Genetics, Institute of Medicine and Public Health, New York, NY, USA. 8California Cancer Registry, Berkeley Cancer Prevention and Control Branch, University of California, Berkeley, CA, USA. 9Intramural Research Program of the National Institutes of Health (NIH) grants R01 CA102776 and R01 CA083855; the Breast Cancer Research Foundation; University of California, Los Angeles. 10Mouse Model and Transgenic Core, Basser Center for BRCA; Victorian Family Cancer Trials Group (VFCGT); Victorian Cancer Agency, Cancer Australia, National Breast Cancer Foundation; The Women’s Cancer Program (WCP) at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, is funded by the American Cancer Society Early Detection Professorship.