Published in final edited form as:
Hum Genet. 2011 November ; 130(5): 685–699. doi:10.1007/s00439-011-1003-z.

Haplotype structure in Ashkenazi Jewish *BRCA1* and *BRCA2* mutation carriers

Kate M. Im
Center for Cancer Research, Cancer Inflammation Program, Human Genetics Section, National Cancer Institute, Frederick, MD, USA

Tomas Kirchhoff
Clinical Genetics Service, Department of Medicine, Program in Cancer Prevention and Population Research, Program in Cancer Biology and Genetics (RJK), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Xianshu Wang
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Todd Green
Department of Genetics and of Medicine, Harvard Medical School, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, USA

Department of Molecular Biology, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

Clement Y. Chow
Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

Joseph Vijai
Clinical Genetics Service, Department of Medicine, Program in Cancer Prevention and Population Research, Program in Cancer Biology and Genetics (RJK), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Joshua Korn
Department of Genetics and of Medicine, Harvard Medical School, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, USA

Department of Molecular Biology, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

Mia M. Gaudet
Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

Zachary Fredericksen and **V. Shane Pankratz**
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Candace Guiducci and **Andrew Crenshaw**

B. Gold NCI-Frederick, Box B, Boyles Street Fort Detrick, Frederick, MD 21702, USA golda@mail.nih.gov.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-1003-z) contains supplementary material, which is available to authorized users.

Conflict of interest The authors report no conflicts of interest.

Department of Genetics and of Medicine, Harvard Medical School, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, USA

Department of Molecular Biology, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

Lesley McGuffog, **Christiana Kartsonaki**, and **Jonathan Morrison**
Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Sue Healey
Genetics and Population Health Division, Queensland Institute of Medical Research, Brisbane, Australia

Olga M. Sinilnikova
Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Centre Hospitalier Universitaire de Lyon/Centre Léon Bérard, Lyon, France

Centre Léon Bérard, Université de Lyon, Lyon, France

Phuong L. Mai and **Mark H. Greene**
Clinical Genetics Branch, National Cancer Institute, Rockville, MD, USA

Marion Piedmonte
GOG Statistical and Data Center, Roswell Park Cancer Institute, Buffalo, NY, USA

Wendy S. Rubinstein
NorthShore University Health System, Evanston, IL, USA

HEBON and **Frans B. Hogervorst**
Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, The Netherlands

Matti A. Rookus
Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands

J. Margriet Collée
Department of Clinical Genetics, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands

Nicoline Hoogerbrugge
Hereditary Cancer Clinic, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

Christi J. van Asperen
Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

Hanne E. J. Meijers-Heijboer
Department of Clinical Genetics, VU Medical Center, Amsterdam, The Netherlands

Cees E. Van Roozendaal
Department of Clinical Genetics, University Medical Center, Maastricht, The Netherlands

Trinidad Caldes and **Pedro Perez-Segura**
Molecular Oncology Laboratory, Hospital Clinico San Carlos, Madrid, Spain

Anna Jakubowska, **Jan Lubinski**, and **Tomasz Huzarski**
International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland

Paweł Blecharz

Center of Oncology, Maria Sklodowska-Curie Memorial Institute, Kraków, Poland

Heli Nevanlinna
Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland

Kristiina Aittomäki
Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland

Conxi Lazaro and **Ignacio Blanco**
Hereditary Cancer Program, Catalan Institute of Oncology-IDIBELL, Barcelona, Spain

Rosa B. Barkardottir
Department of Pathology, Landspitali-LSH, Reykjavik, Iceland

Faculty of Medicine, University of Iceland, Reykjavik, Iceland

Marco Montagna
Immunology and Molecular Oncology Unit, Istituto Oncologico Veneto IOV, IRCCS, Padua, Italy

Emma D'Andrea
Immunology and Molecular Oncology Unit, Istituto Oncologico Veneto IOV, IRCCS, Padua, Italy

Department of Oncology and Surgical Sciences, University of Padua, Padua, Italy

kConFab
Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Peter MacCallum Cancer Centre, Melbourne, Australia

Peter Devilee
Department of Human Genetics and Pathology, Leiden University Medical Center, Leiden, The Netherlands

Olufunmilayo I. Olopade
Center for Clinical Cancer Genetics and Global Health, Department of Medicine, University of Chicago Medical Center, Chicago, IL, USA

Susan L. Neuhausen
Department of Population Sciences, The Beckman Research Institute of the City of Hope, Duarte, CA, USA

Bernard Peissel
Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy

Bernardo Bonanni
Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), Milan, Italy

Paolo Peterlongo
Unit of Genetic Susceptibility to Cancer, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy

IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy

Christian F. Singer
Department of Obstetrics and Gynecology, University of Vienna, Vienna, Austria

Comprehensive Cancer Medical Center, University of Vienna, Vienna, Austria

Gad Rennert and **Flavio Lejbkowicz**
CHS National Cancer Control Center and Department of Community Medicine and Epidemiology, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel

Irene L. Andrulis

Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada

Cancer Care Ontario, University of Toronto, Toronto, ON, Canada

Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Gord Glendon
Cancer Care Ontario, University of Toronto, Toronto, ON, Canada

Hilmi Ozcelik
Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada

Cancer Care Ontario, University of Toronto, Toronto, ON, Canada

Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

Ontario Cancer Genetics Network
Cancer Care Ontario, University of Toronto, Toronto, ON, Canada

Amanda Ewart Toland
Department of Molecular Virology Immunology, and Medical Genetics and Internal Medicine, The Ohio State University, Columbus, OH, USA

Maria Adelaide Caligo
Section of Genetic Oncology, Department of Laboratory Medicine, University of Pisa and Santa Chiara University Hospital, Pisa, Italy

SWE-BRCA
Department of Oncology, Lund University, Lund, Sweden

Mary S. Beattie
Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

Salina Chan
Cancer Risk Program, University of California San Francisco, Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA

UKFOCR
Department of Oncology, University of Cambridge, Cambridge, UK

Susan M. Domchek, **Katherine L. Nathanson**, and **Timothy R. Rebbeck**
Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA

Catherine Phelan
Moffitt Cancer Center, Tampa, FL, USA

Steven Narod
Women's College Research Institute, Toronto, ON, Canada

Esther M. John
Cancer Prevention Institute of California, Fremont, CA, USA

John L. Hopper
Centre for Genetic Epidemiology, University of Melbourne, Victoria, Australia

Saundra S. Buys
Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

Mary B. Daly
Fox Chase Cancer Center, Philadelphia, PA, USA

Melissa C. Southey
Department of Pathology, The University of Melbourne, Victoria 3010, Australia

Mary-Beth Terry
Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

Nadine Tung
Division of Hematology-Oncology, Beth Israel Deaconess Medical Center, Brookline Avenue, Boston, MA, USA

Thomas v. O. Hansen
Genomic Medicine, Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

Ana Osorio
Human Genetics Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre, Madrid, Spain

Spanish Network on Rare Diseases (CIBERER), Madrid, Spain

Javier Benitez
Spanish Network on Rare Diseases (CIBERER), Madrid, Spain

Human Genetics Group and Genotyping Unit, Human Cancer Genetics Programme, Spanish National Cancer Research Centre, Madrid, Spain

Mercedes Durán
Institute of Biology and Molecular Genetics, Universidad de Valladolid (IBGM-UVA), Valladolid, Spain

Jeffrey N. Weitzel
Division of Clinical Cancer Genetics, City of Hope Cancer Center, Duarte, CA, USA

Judy Garber
Dana Farber Cancer Institute, Harvard University, Boston, MA, USA

Ute Hamann
Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

EMBRACE, **Susan Peock**, **Margaret Cook**, **Clare T. Oliver**, **Debra Frost**, and **Radka Platte**
Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

D. Gareth Evans
Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Ros Eeles
Oncogenetics Team, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Surrey, UK

Louise Izatt
Clinical Genetics, Guy's and St. Thomas' NHS Foundation Trust, London, UK

Joan Paterson

Department of Clinical Genetics, East Anglian Regional Genetics Service, Addenbrookes Hospital, Cambridge, UK

Carole Brewer
Department of Clinical Genetics, Royal Devon & Exeter Hospital, Exeter, UK

Shirley Hodgson
Clinical Genetics Department, St Georges Hospital, University of London, London, UK

Patrick J. Morrison
Northern Ireland Regional Genetics Centre, Belfast City Hospital, Belfast, UK

Mary Porteous
South East of Scotland Regional Genetics Service, Western General Hospital, Edinburgh, UK

Lisa Walker
Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK

Mark T. Rogers
All Wales Medical Genetics Services, University Hospital of Wales, Cardiff, UK

Lucy E. Side
North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK

Andrew K. Godwin
Department of Pathology and Laboratory, University of Kansas Medical Center, Kansas City, KS, USA

Rita K. Schmutzler and **Barbara Wappenschmidt**
University Hospital of Cologne, Cologne, Germany

Yael Laitman
The Susan Levy Gertner Oncogenetics unit, Institute of Genetics, Sheba Medical Center, Tel Hashomer, Israel

Alfons Meindl
Technical University Munich, Munich, Germany

Helmut Deissler
University Hospital Ulm, Ulm, Germany

Raymonda Varon-Mateeva
Institute of Human Genetics, Charite Campus Virchow Klinikum, Berlin, Germany

Sabine Preisler-Adams
Institute of Human Genetics, University of Münster, Münster, Germany

Karin Kast
Department of Gynaecology and Obstetrics, University Hospital Carl Gustav, Technical University Dresden, Dresden, Germany

Laurence Venat-Bouvet
Department of Medical Oncology, Centre Hospitalier Universitaire Dupuytren, Limoges, France

Dominique Stoppa-Lyonnet
INSERM U509, Service de Génétique Oncologique, Institut Curie, Université Paris-Descartes, Paris, France

Georgia Chenevix-Trench
Genetics and Population Health Division, Queensland Institute of Medical Research, Brisbane, Australia

Douglas F. Easton
Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Robert J. Klein
Clinical Genetics Service, Department of Medicine, Program in Cancer Prevention and Population Research, Program in Cancer Biology and Genetics (RJK), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Mark J. Daly
Department of Genetics and of Medicine, Harvard Medical School, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, USA

Department of Molecular Biology, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

Eitan Friedman
The Susan Levy Gertner Oncogenetics unit, Institute of Genetics, Sheba Medical Center, Tel Hashomer, Israel

Michael Dean
Center for Cancer Research, Cancer Inflammation Program, Human Genetics Section, National Cancer Institute, Frederick, MD, USA

Andrew G. Clark
Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

David M. Altshuler
Department of Genetics and of Medicine, Harvard Medical School, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, USA

Department of Molecular Biology, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

Antonis C. Antoniou
Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Fergus J. Couch
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Kenneth Offit
Clinical Genetics Service, Department of Medicine, Program in Cancer Prevention and Population Research, Program in Cancer Biology and Genetics (RJK), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Bert Gold
Center for Cancer Research, Cancer Inflammation Program, Human Genetics Section, National Cancer Institute, Frederick, MD, USA

Abstract

Abstract Three founder mutations in *BRCA1* and *BRCA2* contribute to the risk of hereditary breast and ovarian cancer in Ashkenazi Jews (AJ). They are observed at increased frequency in the AJ compared to other BRCA mutations in Caucasian non-Jews (CNJ). Several authors have proposed that elevated allele frequencies in the surrounding genomic regions reflect adaptive or balancing

selection. Such proposals predict long-range linkage dis-equilibrium (LD) resulting from a selective sweep, although genetic drift in a founder population may also act to create long-distance LD. To date, few studies have used the tools of statistical genomics to examine the likelihood of long-range LD at a deleterious locus in a population that faced a genetic bottleneck. We studied the genotypes of hundreds of women from a large international consortium of *BRCA1* and *BRCA2* mutation carriers and found that AJ women exhibited long-range haplotypes compared to CNJ women. More than 50% of the AJ chromosomes with the *BRCA1* 185delAG mutation share an identical 2.1 Mb haplotype and nearly 16% of AJ chromosomes carrying the *BRCA2* 6174delT mutation share a 1.4 Mb haplotype. Simulations based on the best inference of Ashkenazi population demography indicate that long-range haplotypes are expected in the context of a genome-wide survey. Our results are consistent with the hypothesis that a local bottleneck effect from population size constriction events could by chance have resulted in the large haplotype blocks observed at high frequency in the *BRCA1* and *BRCA2* regions of Ashkenazi Jews.

Introduction

Mutations in the *BRCA1* and *BRCA2* genes contribute to risk of hereditary breast and ovarian cancers. Three founder mutations in these two genes (*BRCA1* 185delAG, *BRCA1* 5382insC, and *BRCA2* 6174delT) are observed at a relatively high frequency (~2% in total) in the general Ashkenazi Jewish population (Struewing et al. 1997; Tonin et al. 1996) compared with the occurrence of BRCA mutations in the general population. In one study, 10.3% of Ashkenazi women with invasive breast cancer, unselected for family history, were carriers of one of the founder mutations (King et al. 2003).

Recombination, genetic drift and natural selection all act to shape the landscape of pairwise allelic associations, or linkage disequilibrium, in a genomic region. Associations will occur by chance in small populations due to random changes in allele frequencies (Hartl and Clark 2007). When a specific combination of alleles is beneficial for survival, natural selection may increase the frequency of this haplotype, thereby increasing levels of allelic association. Applying the rationale of this model in reverse, if a region is found to have exceptionally long-range haplotypes, then this observation increases the likelihood that the region has undergone recent positive selection (Sabeti et al. 2002; Voight et al. 2006). There are several examples in literature of this type of extended haplotype structure in regions known to be influenced by natural selection. Deficiencies in glucose-6-phosphate dehydrogenase (G6PD) levels cause several blood disorders, but confer resistance to severe malaria. The *G6PD* A allele, defined by two nonsynonymous changes in the coding region of the *G6PD* gene, is found on a conserved haplotype spanning more than 1.6 Mb (Saunders et al. 2005). Tishkoff et al. (Tishkoff et al. 2007) identified three single nucleotide polymorphisms (SNPs) in the minichromosome maintenance 6 (*MCM6*) gene that are associated with the lactase persistence phenotype. In that research, individuals homozygous for the C-14010 allele exhibited a region of haplotype homozygosity of average length 1.8 Mb, whereas individuals homozygous for the G-14010 allele exhibited only a 1.8kb region of homozygosity (Tishkoff et al. 2007). Although these studies provide examples of long haplo-types consistent with positive selection, population genetics theory and data from *Drosophila melanogaster* (which also underwent an out-of-Africa bottleneck with humans) show that bottlenecked demographics can also result in focused regions of the genome having an apparently short coalescence time. The latter strongly resemble long shared haplotypes (Haddrill et al. 2005). Contrasting the merits of natural selection versus neutral explanations for long haplotypes must be done with caution, if the population is known to have faced a demographic history with a bottleneck. In capsule, that is the take-home message from the work presented herein.

For more than 30 years, numerous authors have asserted that the AJ diseases are a consequence of adaptive evolution. In 2000, without doing a formal analysis, but on the basis of estimates of the the time of common ancestry of alleles of *BRCA1* and *BRCA2*, Slatkin and Rannala (Slatkin and Rannala 2000) concluded that these two genes were likely to have been under selection in AJ. In addition, *BRCA1* was suggested to be a target of adaptive evolution along the human and chimpanzee lineages on the basis of now uncorroborated Hardy–Weinberg statistics (Huttley et al. 2000). More recently, in a very controversial review article, Cochran, Hardy and Harpending (Cochran et al. 2006) made a series of arguments that intelligence was being selected for through variation in the homologous recombination pathway, specifically through the function of *BRCA1* and *BRCA2* founder mutations. Then, through the study of a recent small series of 85 Ashkenazi women, the *BRCA1* 185delAG and 5382insC mutations were seen to occur on the two most common haplotypes, found on 15 and 29% of chromosomes, respectively (Pereira et al. 2007). We examined a ~2 Mb region around *BRCA1* or *BRCA2* in hundreds of AJ women to determine if the 185delAG, 5382insC and 6174delT founder mutations were carried on common haplotypes and to attempt to disentangle the effects of demography and natural selection in this genomic region. Approximately, 16% of Ashkenazi chromosomes containing the *BRCA2*-TCCGAAGA allele that is used here as a surrogate for the 6174delT mutation share a long-range haplotype extending more than 1.4 Mb and over half of women with the 185delAG mutation core haplotype share an identical haplotype extending over the entire 2.1 Mb region examined. In contrast, chromosomes carrying alternative core haplotypes showed much shorter haplo-type sharing. The data gathered from the study of *BRCA1* haplotypes are analogous: of the chromosomes sharing the major 17-SNP *BRCA1* haplotype in the CNJ women, approximately 5.8% share a 1.27-Mb haplotype and only 3.3% share an identical 2.1-Mb haplotype. Multiple processes including positive assortative mating (or endogamy), positive selection and a population bottleneck would affect patterns of allelic association and lead to regions of extended linkage disequilibrium. None of the predicted departures from Hardy–Weinberg equilibrium allele frequencies associated with non-random mating was observed and therefore this hypothesis was dismissed. Through simulation, we show that the observed extended haplotypes can occur by genetic drift alone given that a population bottleneck has occurred.

Materials and methods

Subjects and genotyping

Subjects included 372 Ashkenazi Jewish (AJ) women and 1,441 Caucasian, non-Jewish (CNJ) women sampled for a *BRCA2* modifier locus study(Gaudet et al. 2010) within The Consortium of Investigators of Modifiers of BRCA1/2 (Chenevix-Trench et al. 2007). The AJ group consisted of 317 self-reported AJ women and 55 additional women with the 6174delT founder mutation. These women were included despite the fact that they did not indicate AJ heritage, because this mutation is rarely observed outside the AJ population (Struewing et al. 1999). In addition, a principal components analysis (PCA) demonstrates that these 55 women cluster with women self-reported as AJ (data not shown). All women had deleterious mutations in *BRCA2* and were genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 platform. SNPs and individuals were filtered based on the method recently discussed in detail (Gaudet et al. 2010). We extracted a ~2.1-Mb region surrounding *BRCA2* consisting of 1,402 SNPs to investigate LD patterns and haplotype structure.

Subjects from the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA) harboring deleterious *BRCA1* mutations (Antoniou et al. 2010) included 316 and 297 women with the AJ founder 185delAG and 5382insC mutations, respectively, and 2,186 women with other mutations (designated CNJ for future reference). DNA samples from these women were genotyped on the Illumina Infinium 610 K array platform. For more

information on the study subjects and genotyping (including quality control) see Antoniou et al. (2010).

Linkage disequilibrium and core haplotype

Linkage disequilibrium (LD) among pairs of markers and LD block boundaries were estimated from genotype data using the program Haploview (Barrett et al. 2005). The program's default parameters were used except that the minor allele frequency cutoff was set to 0.05.

The core *BRCA2* haplotype, at which we looked for extended homozygosity, is an LD block consisting of 8 SNPs: rs543304, rs206081, rs4942448, rs11571700, rs11571725, rs4942485, rs206146 and rs15869. The block spans approximately 60 kb within the gene.

The core 17-SNP *BRCA1* haplotypes for AJ and the CNJ women spans approximately 274 and 257 kb, respectively, and consists of the following SNPs: rs382571 (only in AJ core), rs9911630, rs11657053, rs8176273, rs8176265, rs1799966, rs3737559 (only in CNJ core), rs1060915, rs16942, rs799917, rs16940, rs799923, rs9646417, rs11651341, rs4534897, rs4793234, rs11657004 and rs9912203.

Haplotype estimation

Haplotypes for all subjects were estimated using SNPHAP (Clayton, http://www-gene.cimr.cam.ac.uk/clayton/software). SNP positions and chromosomes that could not be unambiguously phased were excluded from analysis.

Extended haplotype structure

Markers that were uninformative for LD analysis (MAF<0.05) were removed from the estimated haplotypes. To determine a consensus haplotype and the extent of haplotype structure, we walked SNP by SNP to the left and right of the core haplotype independently. Chromosomes carrying the consensus haplotype were carried forward and the rest were filtered out. SNPs that filtered more than 20% of the remaining chromosomes were considered “wobble” positions and any allele was permitted at these sites. We determined the left- and right-hand border of the extended haplotype as either (1) the position before a 10-SNP window consisting of five or more wobble positions or (2) the position where less than 20% of the chromosomes share the extended haplotype, whichever was closest to the core haplotype. We performed this analysis for the chromosomes carrying the *BRCA2*-TCCGAGGA and those carrying other alleles separately. The same was done for the *BRCA1* cohort data; chromosomes from individuals with the 185delAG mutation carrying the *BRCA1*-GGAGAG GGAGGAGGGGA were analyzed separately from those carrying a different haplotype. We also examined the extended haplotype homozygosity, a measure designed to quantify the decay of association between a core haplotype and other alleles, with the program Sweep (Sabeti et al. 2002).

Simulations

The program msHOT (Hellenthal and Stephens 2007) was used to generate population samples based on both the one-bottleneck and two-bottleneck models of Slatkin (Slatkin 2004). msHOT can take a specific recombination landscape. We specified the local recombination rates (in cM/Mb) for the 2.1 Mb region surrounding *BRCA2* following the estimates of Myers et al. (Myers et al. 2005). The msHOT code for the single bottleneck model can be found in Appendix A. The size of the largest haplotype block from each sample was retained if it exceeded 715 kb, and from the counts of these mutation-drift recombination samples, we obtained the chance of recovering large haplotype blocks by drift in this demographic scenario.

Core haplotype figures

These were generated using the hplot2 perl script written by Xiaoquan Wen and generously provided by him and Sridhar Kudaravalli from the University of Chicago.

Population differentiation metrics

F-statistics and Hardy–Weinberg measures were calculated using the tools developed by Stefanov et al. (Stefanov et al. 2008).

Principal components analysis (PCA)

Principal components analysis was performed using the smartpca program within the EIGENSOFT suite (Price et al. 2006).

Haplotype network diagrams

These were drawn using the median joining (MJ) algorithm as implemented in Network 4.5.1.6., released 31 December 2009 from Fluxus-Engineering.

Tajima's *D* calculation

This was performed using the R-script kindly supplied by Christopher Carlson for his publication on genome-wide characterization (Carlson et al. 2005).

Results

We examined a region of approximately 2 Mb surrounding *BRCA1* or *BRCA2* to explore the linkage disequilibrium (LD) patterns and haplotype structure and to determine whether these regions exhibit signatures of natural selection. Genotype data for two separate cohorts of women carrying deleterious *BRCA1* or *BRCA2* mutations were used to estimate pairwise LD. The *BRCA2* cohort included 372 Ashkenazi Jewish (AJ) women consisting of 317 self-reported Ashkenazi and 55 women who self-reported as Caucasian non-Jews, but carried the AJ founder 6174delT mutation. The *BRCA1* cohort included 316 and 297 women with AJ founder 185delAG and 5382insC mutations, respectively. The *BRCA2* gene is located in a region of high LD, and the major haplotype for the LD block spanning 60 kb of the gene (TCCGAAGA) is found on 55% of chromosomes. We hypothesized that this was a surrogate for the 6174delT mutation, and therefore used it as the core for determining the extent of haplotype structure in the region (Fig. 1). SNPHAP, an Expectation Maximization algorithm haplotype estimation program, was used to generate haplotypes from the genotype data of all 1,402 SNPs from this genomic region. Uninformative markers were removed from the haplotypes, and the remaining 545 SNPs were analyzed. A total of 412 chromosomes representing 328 women carried the *BRCA2*-TCCGAAGA haplotype including 8 of the 13 women who did not indicate that they were of Ashkenazi descent. We separately examined the extended haplotype structure for chromosomes with the *BRCA2*-TCCGAAGA allele and for those chromosomes with other alleles. A total of 18% of the 412 chromosomes carrying the *BRCA2*-TCCGAAGA allele shared a haplotype extending more than 1.4 Mb, while chromosomes with other alleles shared a haplotype approximately half that length (715 kb) (Fig. 2a). A similar pattern was observed when examining women with the *BRCA1* 185delAG mutation. Approximately, half of the chromosomes carried the major 20-SNP haplotype (GGAGAGGGGAGGAAGGGGAG) for the LD block extending across the gene (Supplementary Fig. 1) and, of those, 51.9% shared an identical 2.1-Mb haplotype surrounding *BRCA1*.

To assess the significance of this extended haplotype, we need to evaluate how unlikely it is to have such a large haplotype given the other haplotypes in the region and the demographic

history of Ashkenazim. The general strategy was to produce simulated data sets from a population whose history of bottlenecks and growth matched the historical record of Ashkenazim, and to use these simulations to determine how often haplotypes as long as those observed occur by purely neutral processes. We performed coalescent-based simulations using the popular simulation program msHOT (Hellenthal and Stephens 2007; Hudson 2002) and a set of demographic parameters matching the founding population size and expansion derived from other genetic data (Slatkin 2004). We considered a 2-Mb window, with a recombination landscape for the *BRCA2* region of chromosome 13 based on the Myers et al. (Myers et al. 2005) genetic map of humans. It required 1250 realizations of the simulation to produce a common haplotype block of 715 kb (equal to the length of the haplotype observed on AJ chromosomes without the 6174delT surrogate allele), suggesting that the observation of a haplotype block of this size may be unlikely. However, the correct question is not whether a long haplotype in this particular region will be found by chance; it is rather to ask whether a 1.4-Mb haplotype will be found among some subset of individuals given that others share a 715-kb haplotype. The existence of the 715-kb haplotype indicates that this is a region of the genome that has fairly shallow common ancestry (a recent coalescence time). Given that the region has a recent coalescence, we then ask whether there might also be even longer haplotypes in the sample. The simulations show that conditional on the sample having a 715-kb haplotype block, it is not at all unlikely that a 1.4-Mb haplotype block is present, associated with the alternative allele and shared by more than 18% of the chromosomes. In addition, at least 5% of chromosomes share a common haplotype across the 2.1-Mb region in more than 30% of simulated data sets. In summary, long-range haplotype structure is not unlikely to occur by purely neutral processes in this region of the genome, given what we know about the demography of the population and the structure of the genetic variation in this region of the genome, suggesting that the Ashkenazi bottleneck drove this region to have a recent coalescence.

We next performed the same analysis on 1,441 Caucasian, non-Jewish (CNJ) women from the *BRCA2* cohort and 2,186 CNJ women from the *BRCA1* cohort. We found no evidence of an extended haplotype in CNJ samples across either *BRCA2* or *BRCA1*. A haplotype nearly five times shorter (~326 kb) than that observed in the Ashkenazi women is observed in 10% of chromosomes carrying the *BRCA2*-TCCGAAGA allele and a 60-kb haplotype is observed in 18% of chromosomes carrying other core haplotypes (Fig. 2b). Of the chromosomes sharing the major 20-SNP *BRCA1* haplotype (AACAGAGAAGAGAGAAAAGG) in the CNJ women, approximately 5.8% share a 1.27-Mb haplotype and only 3.3% share an identical 2.1-Mb haplotype. While this suggests that some extended haplotype structure exists in CNJ individuals in these two regions, it is to a much lesser extent than that observed in AJ individuals.

There are multiple explanations for the discrepancy between Ashkenazi and CNJ haplotype structure. The core haplotype, which is seen on more than half of the chromosomes studied, may be a surrogate for the 6174delT mutation. This mutation is estimated to have arisen only 29 generations ago (Neuhausen et al. 1998), and hence, there is a higher likelihood for the persistence of the haplotype, resulting in the observed haplotype homozygosity. In light of the recombination landscape, which includes several hotspots (Supplementary Fig. 1), this is unlikely to be the cause of the observed haplotype structure. Alternatively, the haplotype structure observed might be a result of endogamy in the Ashkenazi population (i.e., marrying within one's own ethnic group, social class, etc.). This selective mating practice would cause a loss of genetic variation and an increase in the extent of LD (Gibson et al. 2006). Under endogamy, one would expect that Ashkenazi individuals will exhibit positive values of the inbreeding coefficient F_{IS} (Weir and Cockerham 1984), reflecting a heterozygote deficiency relative to Hardy–Weinberg equilibrium expectations. Contrary to what would be expected under endogamy and inbreeding, a large proportion of SNPs in the

region encompassing *BRCA2* have significantly negative F_{IS} (F_{IS}<mean + 3 SD., Supplementary Fig. 2) compared with the rest of chromosome 13. Of course, the excess of heterozygotes for the long haplotype is to be expected, as most homozygotes at these loci are likely to be prenatally lethal. Natural selection in the AJ population could also result in the observed pattern of haplotype structure. To investigate this possibility, we first looked at the population differentiation (as measured by F_{ST}) between AJ and CNJ individuals. The region surrounding *BRCA2* shows a pronounced level of differentiation compared with the remainder of chromosome 13 (Fig. 3a) and compared to the rest of the genome (data not shown). Although extreme F_{ST} values (compared to a genome-wide distribution) have been used to detect natural selection (Akey et al. 2002), our observation is most likely caused by an inherent bias in the ascertainment of our samples. The individuals in the cohort used in this analysis were specifically chosen because they had a mutation in their *BRCA2* gene. The mutation spectrum observed in AJ women is inherently different from that observed in CNJ and therefore the increased F_{ST} is due to skewed ascertainment rather than true population differentiation. We observe the same effect on chromosome 17 in a cohort of women harboring *BRCA1* mutations (Fig. 3b). However, if we look at F_{ST} among an independent cohort of randomly ascertained Ashkenazi individuals and the HapMap CEU (Olshen et al. 2008), we observe no increase in values in the *BRCA1* or the *BRCA2* region, nor do we see an increase in the *BRCA1* region in the *BRCA2* cohort (data not shown).

A skew in the allele frequency spectrum (quantified with Tajima's *D* statistic) can also be used to find regions under natural selection. Both demographic and selection events can result in non-zero *D* values, but demographic effects should be observed across the whole genome. We used the method implemented by Carlson et al. (Carlson et al. 2005) to calculate *D* values in 100-kb sliding windows. By comparing the value of Tajima's *D* in windows across the *BRCA2* region with the distribution across the entire genome, we can determine whether this region is truly exceptional. Tajima's *D* values across the entire genome range from −1.82 to 6.58 with a mean value of 3.07 (Fig. 4). Of the 201 windows across the region of interest, 0 fall below the 5th percentile ($D = 1.081$) and 3 fall above the 95th percentile ($D = 4.786$) of the genome-wide distribution suggesting that this region is not an outlier. The skew of the genome-wide distribution toward positive values is consistent with a recent population bottleneck. This is an empirical confirmation of the simulations—the past demography of the Ashkenazi has resulted in many genomic regions with a recent coalescence and unusual differentiation from the CNJ population (Olshen et al. 2008; Ostrer 2001).

Discussion

A number of genetic diseases are more frequent in Ashkenazi Jews than in other populations. In 1978, two classic papers independently explored the elevated frequency of Tay Sach's disease in the AJ population using the same statistical analysis, but came to two different conclusions: One group argued that genetic drift is unlikely to have produced the observed gene frequency (Chakravarti and Chakraborty 1978); while the other group suggested that historical bottlenecks and genetic drift could explain the frequency (Wagener et al. 1978). In an accompanying editorial, Ewens (Ewens 1978) concluded that ascertainment bias and different assumptions about effective population size of the AJ affected the calculations. While we did not specifically assign the effective population size, our simulation parameters are based on the best inference of historical Ashkenazi population demography [Slatkin (2004)]. In addition, we explored the impact of ascertainment bias on our conclusions by examining F_{ST} deviations in the unbiased Olshen et al. sample of AJ in the *BRCA1* and *BRCA2* regions, and found no significant increase in F_{ST}. We also examined the *BRCA2* mutation carriers in the instant study for significant F_{ST} deviation in the *BRCA1* region and found none. Finally, our previous work (Olshen et al. 2008) searched for

evidence of long haplotypes throughout the genomes of randomly ascertained AJ and the HapMap CEU and found a similar pattern: extended haplotypes were absent, or present to a much lesser extent, in the CEU just like the CNJ in the present study.

BRCA1 and *BRCA2* founder mutations are observed at a higher frequency in Ashkenazi Jews compared to mutations in other populations. We found long-range haplotype structure in a high percentage of AJ chromosomes carrying one of the three *BRCA1* or *BRCA2* founder mutations. We considered three competing hypotheses for the origin of this extended haplotype. The first, which we quickly excluded, was endogamy and/or positive assortative mating. If there were positive assortative mating for traits conferred by this genomic region, this would inflate the homozygosity, reduce the effect of recombination and potentially give rise to an extended haplotype. This hypothesis was rejected because none of the predictions regarding departures from Hardy–Weinberg frequencies were met.

The second hypothesis was that positive natural selection could have driven up the frequency of a specific haplotype. The fact that simulations under a neutral model with random drift (and a founder effect) can generate haplotype blocks of this size does not prove that natural selection did not occur. However, we do know that the demography of the Ashkenazi was different from CNJs, and that other regions of the genome also show elevated F_{ST}. It would seem that invoking natural selection for every region of the genome that appears differentiated is not the most parsimonious solution.

The third hypothesis was that there was a local bottleneck effect from a population size constriction, and this resulted, by chance in this demographic scenario, in a large haplotype block at high frequency in the *BRCA2* region. This would agree with the global analysis by Atzmon, et al. (Atzmon et al. 2010), who showed that IBD sharing of chromosomal segments across the entire genome in AJ individuals was consistent with a severe population bottleneck followed by expansion. This explanation was also consistent with several objective attributes of our data: (1) Local reductions in variation are expected when a population goes through a bottleneck (Haddrill et al. 2005) and the *BRCA2* region is not at all exceptional with respect to the distortion in its site frequency spectrum, as measured by Tajima's *D*. (2) The haplotype network diagrams of both the Ashkenazi and CNJ demonstrated that the 6174delT mutation arose on the major Caucasian haplotype (Fig. 5). This is consistent with previous reports of very recent ascendancy of the deleterious mutation to high frequency in the AJ (Neuhausen et al. 1998).

One of the more compelling arguments against the hypothesis that genetic drift and founder effects explain all the AJ and CNJ genetic differences is that it is unlikely for common genetic diseases to cluster in so few pathways if drift were the only force acting on allele frequencies. But in the absence of a mechanism by which *BRCA1* and *BRCA2* variants might have been selectively favored in the Ashkenazi, the most satisfactory explanation of the unusual haplotype pattern in this region remains random genetic drift and a founder effect. The hypothesis that disabling homologous recombination positively influences cortical differentiation (Cochran et al. 2006) is without foundation, and a recent report that *BRCA2* mutation in AJ is a consequence of selection, on the basis of iHS tool application (Bray et al. 2010) is, in our view, a misinterpretation. The observation that other recent populations (e.g., *BRCA2* 999del5 mutation in Icelanders) manifest different founder mutations (Thorlacius et al. 1997) is consistent with our simulations and, while the above hypotheses are not mutually exclusive, a population bottleneck remains the most parsimonious explanation for the observed haplotype pattern.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by federal funds from the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. We also acknowledge the support of the Starr Foundation, the Breast Cancer Research Foundation and the Sabin Family Fund. The content of this publication does not necessarily reflect the views of the Department of Health and Human Services nor does its mention of trade names, commercial products or organizations imply endorsement by the US government. The authors wish to thank Dr. Colm O'hUigin, who provided valuable comments on an early version of this manuscript.

UKFOCR was supported by a project grant from CRUK to Paul Pharoah. We thank Paul Pharoah, Simon Gayther, Susan Ramus, Carole Pye and Patricia Harrington for their contributions toward the UKFOCR.

The GEMO study (Cancer Genetics Network “Groupe Génétique et Cancer”, Fédération Nationale des Centres de Lutte Contre le Cancer, France) is supported by the Ligue National Contre le Cancer; Association for International Cancer Research Grant (AICR-07-0454); and the Association “Le cancer du sein, parlons-en!” Award.

We wish to thank all the GEMO collaborating groups for their contribution to this study. GEMO Collaborating Centers are: Coordinating Centres, Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Centre Hospitalier Universitaire de Lyon/Centre Léon Bérard, & UMR5201 CNRS, Université de Lyon, Lyon: Olga Sinilnikova, Laure Barjhoux, Sophie Giraud, Mélanie Léone, Sylvie Mazoyer; and INSERM U509, Service de Génétique Oncologique, Institut Curie, Paris: Dominique Stoppa-Lyonnet, Marion Gauthier-Villars, Claude Houdayer, Virginie Moncoutier, Muriel Belotti, Antoine de Pauw. Institut Gustave Roussy, Villejuif: Brigitte Bressac-de-Paillerets, Audrey Remenieras, Véronique Byrde, Olivier Caron, Gilbert Lenoir. Centre Jean Perrin, Clermont–Ferrand: Yves-Jean Bignon, Nancy Uhrhammer. Centre Léon Bérard, Lyon: Christine Lasset, Valérie Bonadona. Centre François Baclesse, Caen: Agnès Hardouin, Pascaline Berthet. Institut Paoli Calmettes, Marseille: Hagay Sobol, Violaine Bourdon, Tetsuro Noguchi, François Eisinger. Groupe Hospitalier Pitié-Salpétrière, Paris: Florence Coulet, Chrystelle Colas, Florent Soubrier. CHU de Arnaud-de-Villeneuve, Mont-pellier: Isabelle Coupier. Centre Oscar Lambret, Lille: Jean-Philippe Peyrat, Joëlle Fournier, Françoise Révillion, Philippe Vennin, Claude Adenis. Centre René Huguenin, St Cloud: Etienne Rouleau, Rosette Lidereau, Liliane Demange, Catherine Nogues. Centre Paul Strauss, Strasbourg: Danièle Muller, Jean-Pierre Fricker. Institut Bergonié, Bordeaux: Michel Longy, Nicolas Sevenet. Institut Claudius Regaud, Toulouse: Christine Toulas, Rosine Guimbaud, Laurence Gladieff, Viviane Feillel. CHU de Grenoble: Dominique Leroux, Hélène Dreyfus, Christine Rebischung. CHU de Dijon: Cécile Cassini, Laurence Faivre. CHU de St-Etienne: Fabienne Prieur. Hôtel Dieu Centre Hospitalier, Chambéry: Sandra Fert Ferrer. Centre Antoine Lacassagne, Nice: Marc Frénay. CHU de Limoges: Laurence Vénat-Bouvet. Creighton University, Omaha, USA: Henry T. Lynch.

For kConFab, we wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinic and the Clinical Follow Up Study (funded 2001–2009 by NHMRC and currently by the National Breast Cancer Foundation and Cancer Australia #628333) for their contributions to this resource, and the many families who contributed to kConFab. kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia.

The CONSIT TEAM study (Consorzio degli Studi Italiani Tumori Ereditari Alla Mammella), acknowledges Marco Pierotti, Siranoush Manoukian, Daniela Zaffaroni, Carla B. Ripamonti and Paolo Radice of the Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; Monica Barile of the Istituto Europeo di Oncologia, Milan, Italy and Loris Bernard of the Cogentech, Consortium for Genomic Technologies, Milan, Italy. We thank all patients and families who participated in this study. This study was supported by funds from Italian citizens who allocated the 5 × 1,000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects `5 × 1,000').

For the Mayo Study, the authors would like to acknowledge funding from the Breast Cancer Research Foundation, Komen Foundation, and NIHCA116167.

The Cancer Prevention Institute of California was supported by the National Cancer Institute, National Institutes of Health under RFACA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry and P.I.s. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Cancer Family Registry (CFR), nor does mention of trade names, commercial products or organizations imply endorsement by the US government or the CFR.

The SWE-BRCA study acknowledge the study collaborators: Per Karlsson, Margareta Nordling, Annika Bergman and Zakaria Einbeigi, Gothenburg, Sahlgrenska University Hospital; Marie Stenmark-Askmalm and Sigrun Liedgren Linköping University Hospital; Åke Borg, Niklas Loman, Håkan Olsson, Ulf Kristoffersson, Helena Jernström, Katja Harbst and Karin Henriksson, Lund University Hospital; Annika Lindblom, Brita Arver, Anna von Wachenfeldt, Annelie Liljegren, Gisela Barbany-Bustinza and Johanna Rantala, Stockholm, Karolinska University Hospital; Beatrice Melin, Henrik Grönberg, Eva-Lena Stattin and Monica Emanuelsson, Umeå University Hospital; Hans Ehrencrona, Richard Rosenquist Brandell and Niklas Dahl, Uppsala University Hospital. The HCSC study was supported by RD06/0020/0021 from ISCIII.

The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) Collaborating Centers are Coordinating center: Netherlands Cancer Institute, Amsterdam, NL: F.B.L. Hogervorst, S. Verhoef, M. Verheus, L.J. van `t Veer, F.E. van Leeuwen, M.A. Rookus; Erasmus Medical Center, Rotterdam, NL: M. Collée, A.M.W. van den Ouweland, A. Jager, M.J. Hooning, M.M.A. Tilanus-Linthorst, C. Seynaeve; Leiden University Medical Center, NL, Leiden: C.J. van Asperen, J.T. Wijnen, M.P. Vreeswijk, R.A. Toll-enaar, P. Devilee; Radboud University Nijmegen Medical Center, Nijmegen, NL: M.J. Ligtenberg, N. Hoogerbrugge; University Medical Center Utrecht, Utrecht, NL: M.G. Ausems, R.B. van der Luijt; Amsterdam Medical Center, NL: C.M. Aalfs, T.A. van Os; VU University Medical Center, Amsterdam, NL: J.J.P. Gille, Q. Waisfisz, H.E.J. Meijers-Heijboer; University Hospital Maastricht, Maastricht, NL: E.B. Gomez-Garcia, C.E. van Roozendaal, Marinus J. Blok, B. Caanen; University Medical Center Groningen University, NL: J.C. Oosterwijk, A.H. van der Hout, M.J. Mourits; The Netherlands Foundation for the detection of hereditary tumours, Leiden, NL: H.F. Vasen. The HEBON study is supported by the Dutch Cancer Society grants NKI1998-1854, NKI2004-3088 and NKI2007-3756.

The Ohio State University Clinical Cancer Genetics (OSU CCG) study is supported by the OSU Comprehensive Cancer Center. We thank Leigha Senter and Kevin Sweet for patient accrual, sample ascertainment and database management. The Human Genetics Sample bank processed the samples.

This work was supported by Cancer Care Ontario and the US National Cancer Institute, National Institutes of Health under RFA # CA- 06-503 and through cooperative agreements with members of the Breast Cancer Family Registry (BCFR) and principal investigators. The content of this manuscript 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. We wish to thank Teresa Selander, Nayana Weerasooriya and members of the Ontario Cancer Genetics Network for their contributions to the study.

Epidemiological study of BRCA1 and BRCA2 mutation carriers (EMBRACE).

Douglas F. Easton is the PI of the study. EMBRACE Collaborating Centers are Coordinating Centre, Cambridge: Susan Peock, Margaret Cook, Clare T. Oliver, Debra Frost, Radka Platte. North of Scotland Regional Genetics Service, Aberdeen: Zosia Miedzybrodzka, Helen Gregory. Northern Ireland Regional Genetics Service, Belfast: Patrick Morrison, Lisa Jeffers. West Midlands Regional Clinical Genetics Service, Birmingham: Trevor Cole, Kai-ren Ong, Jonathan Hoffman. South West Regional Genetics Service, Bristol: Alan Donaldson, Margaret James. East Anglian Regional Genetics Service, Cambridge: Joan Paterson, Sarah Downing, Amy Taylor. Medical Genetics Services for Wales, Cardiff: Alexandra Murray, Mark T. Rogers, Emma McCann. St James's Hospital, Dublin & National Centre for Medical Genetics, Dublin: M. John Kennedy, David Barton. South East of Scotland Regional Genetics Service, Edinburgh: Mary Porteous, Sarah Drummond. Peninsula Clinical Genetics Service, Exeter: Carole Brewer, Emma Kivuva, Anne Searle, Selina Goodman, Kathryn Hill. West of Scotland Regional Genetics Service, Glasgow: Rosemarie Davidson, Victoria Murday, Nicola Bradshaw, Lesley Snadden, Mark Longmuir, Catherine Watt, Sarah Gibson, Eshika Haque, Ed Tobias, Alexis Duncan. South East Thames Regional Genetics Service, Guy's Hospital London: Louise Izatt, Chris Jacobs, Caroline Langman, Anna Whaite. North West Thames Regional Genetics Service, Harrow: Huw Dorkins, Kashmir Randhawa. Leicestershire Clinical Genetics Service, Leicester: Julian Barwell, Nafisa Patel. Yorkshire Regional Genetics Service, Leeds: Julian Adlard, Carol Chu, Julie Miller. Merseyside & Cheshire Clinical Genetics Service, Liverpool: Ian Ellis, Catherine Houghton. Manchester Regional Genetics Service, Manchester: D Gareth Evans, Fiona Lalloo, Jane Taylor. North East Thames Regional Genetics Service, NE Thames, London: Lucy Side, Alison Male, Cheryl Berlin. Nottingham Centre for Medical Genetics, Nottingham: Jacqueline Eason, Rebecca Collier. Northern Clinical Genetics Service, Newcastle: Fiona Douglas, Oonagh Claber, Irene Jobson. Oxford Regional Genetics Service, Oxford: Lisa Walker, Diane McLeod, Dorothy Halliday, Sarah Durell, Barbara Stayner. The Institute of Cancer Research and Royal Marsden NHS Foundation Trust: Ros Eeles, Susan Shanley, Nazneen Rahman, Richard Houlston, Elizabeth Bancroft, Lucia D'Mello, Elizabeth Page, Audrey Ardern-Jones, Kelly Kohut, Jennifer Wiggins, Elena Castro, Anita Mitra, Lisa Robertson. North Trent Clinical Genetics Service, Sheffield: Jackie Cook, Oliver Quarrell, Cathryn Bardsley. South West Thames Regional Genetics Service, London: Shirley Hodgson, Glen Brice, Lizzie Winchester, Charlotte Eddy, Vishakha Tripathi, Virginia Attard. Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton: Diana Eccles, Anneke Lucassen, Gillian Crawford, Donna McBride, Sarah Smalley. EMBRACE is supported by Cancer Research UK Grants C1287/A10118 and C1287/A11990. D. Gareth Evans and Fiona Lalloo are supported by an NIHR grant to the Biomedical Research Centre, Manchester. 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. Ros Eeles, Elizabeth Bancroft and Lucia D'Mello are also supported by Cancer Research UK Grant C5047/A8385.

CBCS: We thank Bent Ejlertsen, Mette K. Andersen, Anne-Marie Gerdes and Susanne Kjaergaard for clinical data. Moreover, we thank the NEYE foundation for financial support.

Appendix A—msHOT command

/mshot 412 1 -r 1.0 2087508 -s 545 -v 30 46440 55764 92.72 94074 101467 172.87 116262 120708 84.46 158286 166715 60.34 188386 195853 69.08 203508 207586 27.50 408508 419950 23.61 425890 436582 14.86 531508 540447 22.26 661931 673793 120.99 770385 780441 29.47 815508 819627 29.47 824410 832053 71.03 896479 899758 92.66 907490 914508 66.72 927238 930508 5.91 951526 958508 33.76 967928 979794 50.26 996275 999508 107.56 1084508 1088508 28.62 1091504 1098508 126.14 1616508 1619571 340.72 1621508 1627508 15.09 1660329 1663531 623.28 1675309 1679508 35.42 1686508 1706508 5.02 1850331 1854508 65.25 1883508 1887508 237.76 1962482 1967508 38.82 2055476 2087508 53.91 -G 2355557.4 -eG 0.0000018 0.0000571 -eN 0.0000018 0.095.

Abbreviations

AJ	Ashkenazi Jews
CIMBA	Consortium of Investigators of Modifiers of BRCA1 and BRCA2
CNJ	Caucasian non-Jews
LD	Linkage disequilibrium
MAF	Minor allele frequency
MJ	Median joining
PCA	Principal components analysis
SNP	Single nucleotide polymorphism
IBD	Identity by descent

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 2002; 12:1805–1814. doi:10.1101/gr.631202. [PubMed: 12466284]

Antoniou AC, Beesley J, McGuffog L, Sinilnikova OM, Healey S, Neuhausen SL, Ding YC, Rebbeck TR, Weitzel JN, Lynch HT, Isaacs C, Ganz PA, Tomlinson G, Olopade OI, Couch FJ, Wang X, Lindor NM, Pankratz VS, Radice P, Manoukian S, Peissel B, Zaffaroni D, Barile M, Viel A, Allavena A, Dall'olio V, Peterlongo P, Szabo CI, Zikan M, Claes K, Poppe B, Foretova L, Mai PL, Greene MH, Rennert G, Lejbkowicz F, Glendon G, Ozcelik H, Andrulis IL, Thomassen M, Gerdes AM, Sunde L, Cruger D, Birk Jensen U, Caligo M, Friedman E, Kaufman B, Laitman Y, Milgrom R, Dubrovsky M, Cohen S, Borg A, Jernstrom H, Lindblom A, Rantala J, Stenmark-Askmalm M, Melin B, Nathan-son K, Domchek S, Jakubowska A, Lubinski J, Huzarski T, Osorio A, Lasa A, Duran M, Tejada MI, Godino J, Benitez J, Hamann U, Kriege M, Hoogerbrugge N, van der Luijt RB, Asperen CJ, Devilee P, Meijers-Heijboer EJ, Blok MJ, Aalfs CM, Hogervorst F, Rookus M, Cook M, Oliver C, Frost D, Conroy D, Evans DG, Lalloo F, Pichert G, Davidson R, Cole T, Cook J, Paterson J, Hodgson S, Morrison PJ, Porteous ME, Walker L, Kennedy MJ, Dorkins H, Peock S, Godwin AK, Stoppa-Lyonnet D, de Pauw A, et al. Common Breast Cancer Susceptibility Alleles and the Risk of Breast Cancer for BRCA1 and BRCA2 Mutation Carriers: Implications for Risk Prediction. Cancer Res. 2010; 70:9742–9754. doi: 10.1158/0008-5472.CAN-10-1907. [PubMed: 21118973]

Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, Palamara PF, Morrow B, Friedman E, Oddoux C, Burns E, Ostrer H. Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. Am J Hum Genet. 2010; 86:850–859. [PubMed: 20560205]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

Bray SM, Mulle JG, Dodd AF, Pulver AE, Wooding S, Warren ST. Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. Proc Natl Acad Sci USA. 2010; 107:16222–16227. doi:10.1073/pnas.1004381107. [PubMed: 20798349]

Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA. Genomic regions exhibiting positive selection identified from dense genotype data. Genome Res. 2005; 15:1553–1565. [PubMed: 16251465]

Chakravarti A, Chakraborty R. Elevated frequency of Tay-Sachs disease among Ashkenazic Jews unlikely by genetic drift alone. Am J Hum Genet. 1978; 30:256–261. [PubMed: 677122]

Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). Breast Cancer Res. 2007; 9:104. doi: 10.1186/bcr1670. [PubMed: 17466083]

Clayton, D. SNPHAP: a program for estimating frequencies of large haplotypes of SNPs. http://www-gene.cimr.cam.ac.uk/clayton/ software

Cochran G, Hardy J, Harpending H. Natural history of Ashkenazi intelligence. J Biosoc Sci. 2006; 38:659–693. doi:10.1017/ S0021932005027069. [PubMed: 16867211]

Ewens WJ. Tay-Sachs disease and theoretical population genetics. Am J Hum Genet. 1978; 30:328–329. [PubMed: 677129]

Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, Segre AV, McGee K, McGuffog L, Kartsonaki C, Morrison J, Healey S, Sinilnikova OM, Stoppa-Lyonnet D, Mazoyer S, Gauthier-Villars M, Sobol H, Longy M, Frenay M, Collaborators GS, Hogervorst FB, Rookus MA, Collee JM, Hoogerbrugge N, van Roozendaal KE, Piedmonte M, Rubinstein W, Nerenstone S, Van Le L, Blank SV, Caldes T, de la Hoya M, Nevanlinna H, Aittomaki K, Lazaro C, Blanco I, Arason A, Johannsson OT, Barkardottir RB, Devilee P, Olopade OI, Neuhausen SL, Wang X, Fredericksen ZS, Peterlongo P, Manoukian S, Barile M, Viel A, Radice P, Phelan CM, Narod S, Rennert G, Lejbkowicz F, Flugelman A, Andrulis IL, Glendon G, Ozcelik H, Toland AE, Montagna M, D'Andrea E, Friedman E, Laitman Y, Borg A, Beattie M, Ramus SJ, Domchek SM, Nathanson KL, Rebbeck T, Spurdle AB, Chen X, Holland H, John EM, Hopper JL, Buys SS, Daly MB, Southey MC, Terry MB, Tung N, Overeem Hansen TV, Nielsen FC, Greene MI, Mai PL, Osorio A, Duran M, Andres R, Benitez J, Weitzel JN, Garber J, Hamann U, Peock S, Cook M, Oliver C, Frost D, Platte R, Evans DG, Lalloo F, Eeles R, Izatt L, Walker L, Eason J, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. PLoS Genet. 2010; 6:e1001183. doi:10.1371/journal.pgen.1001183. [PubMed: 21060860]

Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet. 2006; 15:789–795. doi:10.1093/hmg/ddi493. [PubMed: 16436455]

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. Genome Res. 2005; 15:790–799. doi:10.1101/gr.3541005. [PubMed: 15930491]

Hartl, DL.; Clark, AG. Principles of Population Genetics. Fourth edn edn.. Sinauer Associates, Inc; Sunderland: 2007.

Hellenthal G, Stephens M. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. Bioinformatics. 2007; 23:520–521. [PubMed: 17150995]

Hudson R. Generating samples under a Wright-Fisher model of genetic variation. Bioinformatics. 2002; 18:337–338. [PubMed: 11847089]

Huttley G, Easteal S, Southey M, Tesoriero A, Giles G, McCredie MRE, Hopper J, Venter D, Study ABCF. Adapative evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Nat Genet. 2000; 25:410–413. [PubMed: 10932184]

King MC, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003; 3002:643–646. [PubMed: 14576434]

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. Science. 2005; 310:247–248. [PubMed: 16224010]

Neuhausen S, Godwin A, Gershoni-Baruch R, Schubert E, Garber J, Stoppa-Lyonnet D, Olah E, Csokay B, Serova O, Lalloo F, Osorio A, Stratton M, Offit K, Boyd J, 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. [PubMed: 9585613]

Olshen AB, Gold B, Lohmueller KE, Struewing JP, Satagopan J, Stefanov SA, Eskin E, Kirchhoff T, Lautenberger JA, Klein RJ, Friedman E, Norton L, Ellis NA, Viale A, Lee CS, Borgen PI, Clark AG, Offit K, Boyd J. Analysis of genetic variation in Ashkenazi Jews by high density SNP genotyping. BMC Genet. 2008; 9:14. [PubMed: 18251999]

Ostrer H. A genetic profile of contemporary Jewish populations. Nat Rev Genet. 2001; 2:891–898. [PubMed: 11715044]

Pereira LHM, Pineda MA, Rowe WH, Fonseca LR, Greene M, Offit K, Ellis NA, Zhang J, Collins A, Struewing J. The BRCA1 Ashkenazi founder mutations occur on common haplotypes and are not highly correlated with anonymous single nucleotide polymorphisms likely to be used in genome-wide case-control association studies. BMC Genet. 2007; 8:68. [PubMed: 17916242]

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. doi:10.1038/ng1847. [PubMed: 16862161]

Sabeti P, Reich D, Higgins J, Levine HZ, Richter D, Schaffner S, Gabriel S, Platko J, Patterson N, McDonald G, Ackerman H, Campbell S, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander E. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002; 419:832–837. [PubMed: 12397357]

Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW. The extent of linkage disequilibrium caused by selection on *G6PD* in humans. Genetics. 2005; 171:1219–1229. [PubMed: 16020776]

Slatkin M. A population-genetic test of founder effects and implications for Ashkenazi Jewish disease. Am J Hum Genet. 2004; 75:282–293. [PubMed: 15208782]

Slatkin M, Rannala B. Estimating allele age. Annu Rev Genomics Hum Genet. 2000; 1:225–249. doi: 10.1146/annurev.genom. 1.1.225. [PubMed: 11701630]

Stefanov S, Lautenberger J, Gold B. An analysis pipeline for genome-wide association studies. Cancer Inform. 2008; 6:455–461. [PubMed: 19096721]

Struewing JP, Hartge P, Wacholder S, Baker S, Berlin M, McAdams M, Timmerman M, Brody L, Tucker M. The risk of cancer associated with specific mutations of BRCA1 and BRCA2. N Engl J Med. 1997; 336:1401–1408. [PubMed: 9145676]

Struewing JP, Coriaty ZM, Ron E, Livoff A, Konichezky M, Cohen P, Resnick MB, Lifzchiz-Mercerl B, Lew S, Iscovich J. Founder BRCA1/2 mutations among male patients with breast cancer in Israel. Am J Hum Genet. 1999; 65:1800–1802. doi:10.1086/302678. [PubMed: 10577940]

Thorlacius S, Sigurdsson S, Bjarnadottir H, Olafsdottir G, Jonasson JG, Tryggvadottir L, Tulinius H, Eyfjord JE. Study of a single BRCA2 mutation with high carrier frequency in a small population. Am J Hum Genet. 1997; 60:1079–1084. [PubMed: 9150155]

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt C, Silverman J, Powell K, Mortensen H, Hirbo J, Osman M, Ibrahim M, Omar S, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet. 2007; 39:31–40. [PubMed: 17159977]

Tonin P, Weber B, Offit K, Couch F, Rebbeck TR, Neuhausen S, Godwin A, Daly M, Wagner-Costalos J, Berman D, Grana G, Fox E, et al. Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi breast cancer families. Nat Med. 1996; 2:1179–1183. [PubMed: 8898735]

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. PLoS Biology. 2006; 4:e72. [PubMed: 16494531]

Wagener D, Cavalli-Sforza LL, Barakat R. Ethnic variation of genetic disease: roles of drift for recessive lethal genes. Am J Hum Genet. 1978; 30:262–270. [PubMed: 677123]

Weir B, Cockerham C. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38:1358–1370.

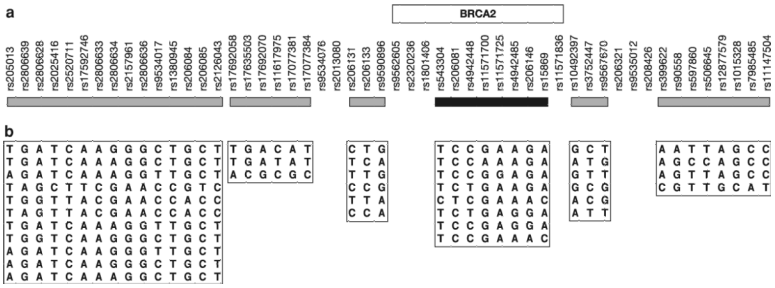

Fig. 1.

Linkage disequilibrium across a 570-kb region in 372 Ashkenazi women. **a** Haploview output showing LD blocks across region. *Black block* represents the core haplotype. **b** Haplotypes estimated using SNPHAP for each of the LD blocks. Haplotypes observed in at least 1% of chromosomes are shown in order of frequency (most frequent first)

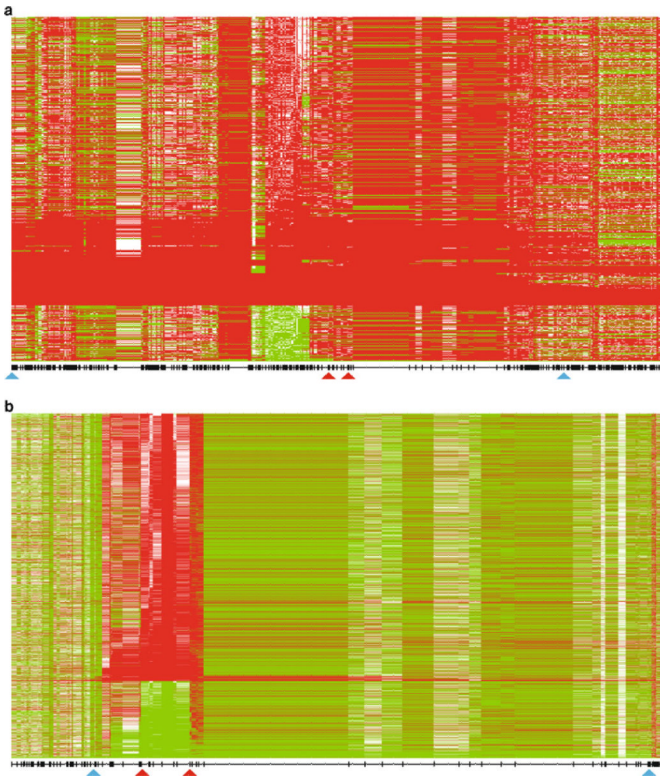

Fig. 2.

Extended haplotype structure observed in (**a**) 372 Ashkenazi women across a 2.1-Mb region and (**b**) 1441 Caucasian, non-Jewish women across an 805-kb region. Each row represents an individual. *Tick marks show* SNP locations, *red triangles* mark the boundaries of the 8-SNP core *BRCA2* haplotype, and *blue triangles* mark the boundaries of the (**a**) 1.4 Mb extended haplotype and (**b**) 715 kb extended haplotype in the AJ and CNJ, respectively. Long horizontal *red lines* represent the core haplotype, while interruptions in that haplotype observed in the data by the presence of frequent alternative alleles are colored in *green*

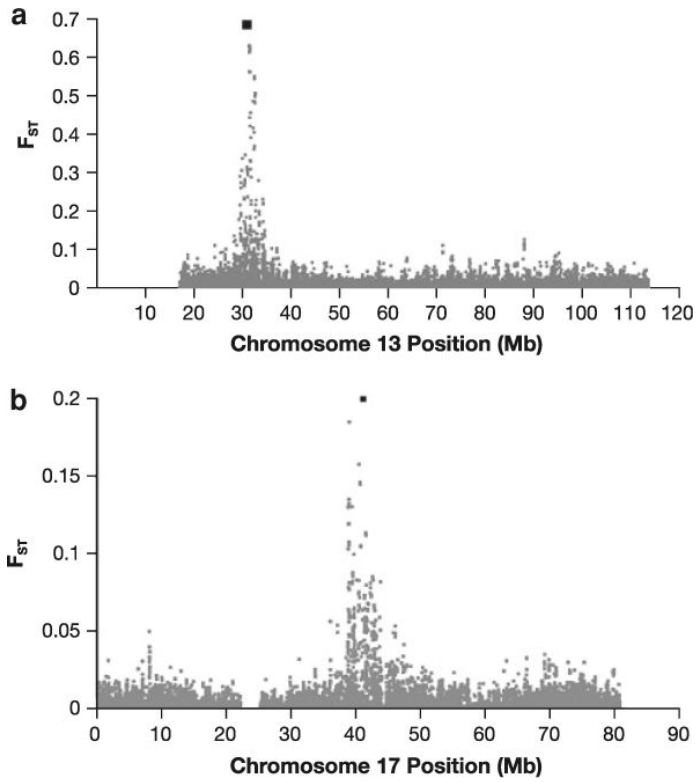

Fig. 3.

Population differentiation among Ashkenazi and CNJ women. **a** F_{ST} between 372 Ashkenazi and 1,441 Caucasian, non-Jewish women across chromosome 13. **b** F_{ST} between 613 Ashkenazi (carriers of the 185delAG and 5382insC founder mutations) and 2,186 Caucasian, non-Jewish (carriers of non-founder mutations) women across chromosome 17. The peaks of F_{ST} occur in the region surrounding *BRCA2* and *BRCA1* in (**a**) and (**b**), respectively, represented by the *black boxes*

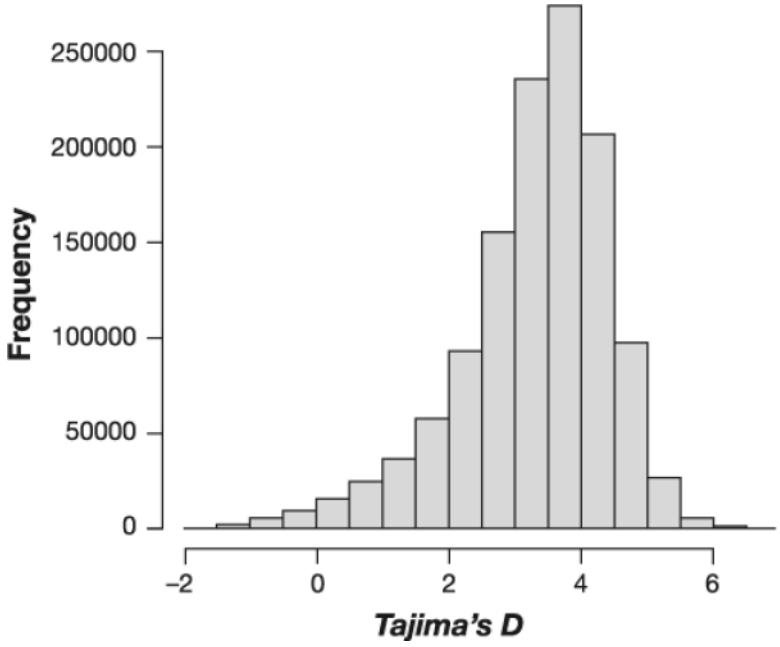

Fig. 4.

Genome-wide distribution of Tajima's *D* in 372 AJ women. *D* statistics were estimated for 100-kb sliding windows

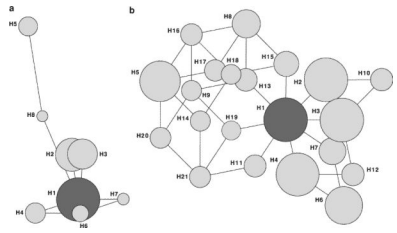

Fig. 5.

Haplotype network results from analyzing the eight SNPs that make up the core *BRCA2* haplotype. Diagrams constructed using haplotypes observed at a frequency of at least 1% are labeled. The *dark gray* sphere (H1) represents the core haplotype. **a** Network for AJ chromosomes. **b** Network for CNJ chromosomes. Frequency of each haplotype can be found in Supplementary Table 1. The size of each sphere is proportional to its frequency in the sample