Am. J. Hum. Genet. 64:1427–1439, 1999

Evidence for Effective Suppression of Recombination in the Chromosome 17q21 Segment Spanning *RNU2–BRCA1*

Xudong Liu and David F. Barker

Department of Physiology, University of Utah Health Sciences Center, Salt Lake City

Summary

Characterization of associations between polymorphic sites located throughout the ~200–400-kb variable-length region spanning *RNU2–BRCA1* reveals nearly complete linkage disequilibrium. This segment spans the *RNU2* array, which includes 6–30 tandem copies of the *U2 snRNA* gene, and an adjacent region containing *NBR1,* the *LBRCA1* pseudogene, *NBR2,* and *BRCA1* in a tandemly duplicated structure. A series of biallelic polymorphisms define two common haplotypes that do not vary significantly, in structure or frequency, between populations of primarily European (n = 275) or Asian (n = 34) ancestry. Lower-frequency variants occurring at distantly located sites within this region also show very strong associations. The rarer haplotype classes appear to be distinguished by mutational alteration and are not recombination products of the two major classes. The two major haplotypes also exhibit significantly different allele-length distributions for local simple tandem-repeat markers. The conservation of extensive distinct chromosomal haplotypes during a long period of human population expansion and divergence indicates that selective forces or specific chromosomal mechanisms result in effective recombination suppression. The extreme degree of long-range linkage disequilibrium at this locus may be exceeded only by that reported for the human *MHC* locus, where allele-specific functional interactions are believed to be significant. These findings have implications for the estimation of the time of origin of *BRCA1* mutations having a founder effect, the interpretation of the significance of rare allelic variants, and the study of the origins of modern populations.

Received November 4, 1997; accepted for publication February 19, 1999; electronically published April 6, 1999.

Address for correspondence and reprints: Dr. David F. Barker, Department of Physiology, Room 155, 410 Chipeta Way, University of Utah Research Park, Salt Lake City, UT 84108. E-mail: david.f.barker@m.cc.utah.edu

0002-9297/99/6405-0024$02.00

Introduction

Factors governing the frequency and location of recombination events on the chromosomes of higher eukaryotes are not well understood. In humans, differences in male versus female rates in pericentromeric, telomeric, or subtelomeric regions (Fain et al. 1989; Buetow et al. 1991; Blouin et al. 1995) imply the existence of global governing mechanisms. More-localized variation, including "hot" and "cold" spots located at the human *MHC* locus (Klitz et al. 1995; Cullen et al. 1997; Malfroy et al. 1997) and other genomic regions (Chakravarti et al. 1984), suggest the existence of localized structures or functional elements that affect crossover occurrence or product viability by means similar to those demonstrated in lower eukaryotes, as reviewed by Lichten and Goldman (1995). More highly detailed maps of human chromosomes are now providing the opportunity to make precise comparisons of genetically and physically defined distances, revealing the overall patterns of variation in their correspondence. The recent completion of a high-resolution X-chromosome physical map revealed a large chromosomal segment with a below-average rate of recombination (Nagaraja et al. 1997).

The chromosome 17q region containing *BRCA1* has been characterized extensively. Physical maps are derived from cosmid, YAC, P1, and BAC clones (Albertsen et al. 1994; Neuhausen et al. 1994; Couch et al. 1995), as well as from pulsed-field gel electrophoresis (Jones et al. 1994) and full-sequence analysis (Smith et al. 1996). Numerous simple tandem-repeat (STR) and base-pair polymorphisms (Couch et al. 1996) are known to exist throughout the region. More-recent studies have detected the tandem genomic duplication that includes the promoter and 5′ portion of *BRCA1* (Barker et al. 1996*b;* Brown et al. 1996) and have elucidated the structure of the *NBR2* gene, which is transcribed from a bidirectional promoter shared with *BRCA1* (Xu et al. 1997*a,* 1997*b*). The *NBR1* gene is located head-to-head with the *LBRCA1* pseudogene, separated by the promoter region paralogous to that between *NBR2* and *BRCA1*.

The *RNU2* array, lying downstream from the *NBR1* gene, has also been well characterized and contains a variable number, from 6 to >30, of tandem 6-kb ele-

ments, each of which includes a transcribed copy of the *U2 snRNA*. There is a high degree of sequence homogeneity for elements within a single array, even at sites that have been demonstrated to be polymorphic in the population (Liao et al. 1997). The mechanism for maintaining array homogeneity operates in situ, because the cytogenetic location of the *RNU2* array has remained stable during primate evolution whereas the tandem elements have undergone extensive remodeling (Pavelitz et al. 1995). The basis of the array-homogenizing mechanism is not understood, but it is thought to involve some form of controlled recombination.

Previous studies of linkage disequilibrium in this portion of chromosome 17 have revealed strong associations between sites located within *BRCA1* exon 11 (Durocher et al. 1996; Dunning et al. 1997), as well as between sites located at opposite ends of the *RNU2* array (Liao et al. 1997) and within the *EDH17B2* gene (Normand et al. 1993). The data described herein demonstrate that a strong, nearly absolute, association between polymorphic sites extends throughout a genomic region that contains the human *RNU2* array and the *NBR1, NBR2,* and *BRCA1* genes but that does not include *EDH17B2*.

Biallelic sites at positions spanning *RNU2–BRCA1* were genotyped in 275 members of a primarily European population and in 34 Asians, revealing two common haplotype patterns that have been conserved since the divergence of these groups ~100,000 years ago (> 4,000 generations). Subsequent characterization of the STR allele patterns in a subset of individuals homozygous for the common biallelic haplotypes revealed evidence for founder alleles specific to each haplotype. For the most part it appears that the two major haplotypes "diverged" at some point in human evolution and that there is no evident means for the exchange of accumulated genetic variation from one to the other. This, along with the apparent similar divergence of the rarer haplotypes, indicates effective suppression of recombination within this region. These findings imply that the significance of haplotype sharing by carriers of specific *BRCA1* mutations will require reevaluation, as will the ages of *BRCA1* mutations with founder effects. This region may also prove to be of great value for the characterization of human population dispersion. Since the DNA segments that are behaving as meiotic alleles are all ≥200 kb in length, there is a vast potential for distinctive DNA structural variation, some of which may serve for molecular timekeeping.

Subjects and Methods

Population Samples

A control group of unrelated individuals, unselected for any disease phenotype, including primarily residents of the United States who are of European ancestry has been described by Barker et al. (1996*a*). Samples of DNA purified from blood were used for genotyping a subset comprising 285 subjects. Asian control sample populations were made up primarily of native-born Chinese and Japanese individuals who are either students at the University of Utah or local residents. Buccal cells collected with rayon swabs were extracted as described by Balnaves et al. (1990) or with use of a Puregene kit (Gentra Systems). The studies reported here are institutional-review-board approved, with the provision that no individual identifiers are retained.

Haplotype Inferences and Disequilibrium Calculations

Allele-specific oligonucleotide (ASO) hybridization was used for genotyping (Barker et al. 1996*a*). Partial data from 10 individuals with a poor DNA sample quality were disregarded. In 12 instances in which ASO resulted in weak or ambiguous hybridization signals for individuals with otherwise complete data, PCR products were sequenced directly.

Biallelic sites with frequencies (*f*) close to .66 and .34 are referred to here as "common." All these sites are in Hardy-Weinberg equilibrium, in both Asian and U.S./European samples, on the basis of χ^2 evaluation. Thus, the high frequency of individuals homozygous for all the 13 common markers tested—117/275 for the .66-frequency alleles and 32/275 for the .34-frequency alleles—allowed the inference that only two haplotypes account for ~98% of the chromosomes present in the U.S./European group. One of these haplotypes consists entirely of the .66-frequency allele at each site; the second bears the .34-frequency allele at each site. For those individual samples that were not evidently homozygous or heterozygous for the common marker-haplotype patterns, inference of the haplotypes was made on the assumption that the observed allele pattern represented the combination of one of the common haplotypes with a second novel haplotype.

For other dimorphic markers, with one allele with a frequency ≤.07, we inferred a unique association of the less-frequent allele with one of the two common marker haplotypes, by examining individuals with the rare allele who were also homozygous for a common marker haplotype. The strong associations found between distantly located sites with the same rare-allele frequency, as well as the failure to observe any instance of the rare allele in a homozygote for the alternate common haplotype, strongly supported the conclusion that each rare site was exclusively associated with one of the two common haplotypes. We interpreted two individuals who were found to carry rare alleles that are not apparently associated with each other, by introducing the fewest number (one) of additional new haplotypes necessary to explain the

observation. These interpretations represent the most parsimonious inferences from the genotypes of unrelated individuals. Because they have not been confirmed by inheritance studies, the precise features of very rare or unique haplotype patterns are tentative, although this uncertainty has little effect on overall data evaluation.

Using haplotype inferences, we calculated the disequilibrium measure D for any pair of biallelic loci, as $D = P11 - p1q1$, where $P11$ is the frequency of the most common haplotype of two loci and $p1$ and $q1$ are the frequencies of the alleles present on that haplotype (Lewontin and Kojima 1960). The correlation coefficient is calculated as $r = D/(p1q1p2q2)^{\frac{1}{2}}$, where r denotes correlation and p and q are the frequencies of each allele at the two loci (Hill and Robertson 1968). As an alternate approach, the estimation-maximization algorithm described by Weir (1990) was implemented in an EXCEL file format. Each pair of sites was analyzed. We selected appropriate initial values for the iterative calculation method using counts of explicitly known haplotypes from homozygous individuals. The calculated values confirmed those obtained with haplotypes inferred by inspection.

STR and RFLP Analysis

We determined STR genotypes as described by Barker and Fain (1993). The STR markers *D17S1333, D17S902, D17S1325, D17S1332, D17S1326, D17S1327, D17S1323, D17S1322, D17S855, BRCA1 90565, D17S1328, D17S1340, D17S1321, D17S1146, D17S1320,* and *D17S1185* have been reported elsewhere (Weissenbach et al. 1992; Couch et al. 1994; Neuhausen et al. 1994; Bennett-Baker et al. 1996) and/or are included in the Genome Database, except for *BRCA 190565,* for which the primers (5′ TAA GTT TGT GAC TAC AGA GTG TTT 3′ and 5′ CCC TCT CTT CCC TCA TCT A 3′) spanning the site of a GT repeat located in the intergenic region between the 3′ end of *BRCA1* and the *RHO7* gene (Smith et al. 1996) were derived from GenBank entry L78833. For *D17S1332* (Bennett-Baker et al. 1996), an alternate reverse primer (5′ CAG AAG GCC GCA GTG AGG 3′) was used and PCR was done with an annealing temperature of 67°C for 26 cycles. During this work, we noted that markers *D17S1146* and *D17S1321* appeared to represent the same locus, as indicated by similarity of allele patterns, partial overlap of primer sequences, and successful amplification, from genomic clone templates, of appropriately sized products with pairs *D17S1146* F plus *D17S1321* F or *D17S1146* R plus *D17S1321* R. Site E within the *EDH17B2* gene was assayed by *Bfa*I digestion of PCR products spanning this site (Normand et al. 1993).

Results

Haplotypes of Biallelic Sites

Table 1 summarizes 19 polymorphisms examined in this study. Eight of these were recently detected in our laboratory, and details of methods used for their discovery and characterization are described elsewhere. The positions of sequence polymorphisms and STRs with respect to the known gene structures in this region, including the *RNU2* array, *NBR1, NBR2,* and *BRCA1,* are illustrated in figure 1.

Figure 2*A* summarizes haplotypes inferred from genotypes of biallelic markers A–R in the U.S./European group of 275 individuals, as well as in a separate group of 34 Asians. Two major haplotypes of the 13 common dimorphic sites were detected. The class I haplotype includes the more frequent allele at sites A, C, E–I, K, L, N, O, Q, and R. The class II haplotype carries the alternate allele at each of these sites. Of 550 chromosomes in the U.S./European set, 353 are type I, ($f = .64$), and 183 are type II ($f = .33$). Only 14 chromosomes ($f = .025$) are not consistent with a type I or type II pattern. Eleven of these may be grouped with type I, and two others with type II, because they are different at just one or two of the common sites. One chromosome is unlike type I and type II, at seven and six of the common sites, respectively. This chromosome, labeled "type III," represents an apparent interspersion of three segments that are similar to type I and three segments similar to type II (fig. 2*A*), a pattern that is not consistent with derivation from the common types by a simple crossover but that could represent an ancestral structure. The individual with this singular haplotype appeared to carry a second chromosome with haplotype IIa.

The two major haplotype classes found in Europeans are found at a very similar frequency in Asians (fig. 2*A*). Because of sample limitations, fewer sites were tested for the Asian group; however, these sites represent the full length of the segment spanning *RNU2–BRCA1*. Among the 34 Asians tested, 16 are homozygous for haplotype I, 4 are homozygous for haplotype II, and 13 are type I/II heterozygotes, closely matching expectations for Hardy-Weinberg equilibrium. One II-like haplotype was found, which is possibly similar to the IIb form found in U.S./European subjects. The haplotype frequencies in the Asian group are .66 for type I and .34 for type II (or II like). Similar frequencies for individual common sites within the *BRCA1* locus have been reported for a Japanese population, by Inoue et al. (1995).

Five additional biallelic sites with rare-allele frequencies $\leq .07$, detected in the U.S./European group, provide further insight into the nature of haplotype diversity in this region. All the variation in these low-frequency sites was found on I or I-like chromosomes. A major hap-

Table 1

Locations and Characteristics of Biallelic Sites Assayed

Site	Genbank Accession No. and Location	Gene	A1	A2	f(1)[a]	f(2)	Reference
A	None	*RNU2* (JL+5)	G	C	.65	.35	Pavelitz et. al. 1995; Liao et al. 1997
B	X76952 3866-68	*NBR1*	AAA	AAAA	.99	.01	Liu and Barker, unpublished data
C	X76952 2860	*NBR1*	A	G	.66	.34	Campbell et al. 1994
D	U72483 1659	LBRCA1	A	G	.99	.01	Liu and Barker, unpublished data
E	U72483 1723	LBRCA1	G	A	.66	.34	Liu and Barker, unpublished data
F	U72483 2182	LBRCA1	G	A	.66	.34	Liu and Barker, unpublished data
G	U37574 612	*NBR2*	C	T	.66	.34	Barker et al. 1996*b*
H	U37574 873	*NBR2*	A	G	.65	.35	Liu and Barker, unpublished data
I	U37574 980-982	*NBR2*	$(AAC)_1$	$(AAC)_2$	.66	.34	Barker et al. 1996*b*
J	U37574 993	*NBR2*	A	G	.94	.06	Liu and Barker, unpublished data
K	U37574 2642	*BRCA1*	A	G	.66	.34	Liu and Barker, unpublished data
L	U37574 2743	*BRCA1*	T	C	.66	.34	Liu and Barker, unpublished data
M	U14680 1186	*BRCA1*	A	G	.93	.07	Miki et al. 1994; Durocher et al. 1996
N	U14680 2430	*BRCA1*	T	C	.66	.34	Friedman et al 1994
O	U14680 2731	*BRCA1*	C	T	.65	.35	Miki et al. 1994; Dunning et al. 1997
P	U14680 3238	*BRCA1*	G	A	.98	.02	Castilla et al. 1994; Durocher et al 1996
Q	U14680 4427	*BRCA1*	T	C	.67	.33	Miki et al. 1994
R	U14680 4956	*BRCA1*	A	G	.66	.34	Miki et al. 1994
	M27138 2019	*EDH17B2*	T (+)	C (-)	.47	.53	Normand et al. 1993

[a] Allele frequencies are for the 275 U.S./European individuals studied in this report, except for *EDH17B2,* for which frequencies are as reported in the cited reference.

lotype subclass, Id, included substitution A→G at both site J, within *NBR2* intron 1, and at site M, within *BRCA1* exon 11. These sites (upward-pointing arrows in fig. 2*A*) are separated by 31.5 kb (GenBank L78833; Smith et al. 1996). Of 36 individuals with the rare G allele at site M, 35 also carried it at site J, indicating cooccurrence on these 35 chromosomes. Haplotype Ih represents an additional example of strong association between distantly located low-frequency variants. This I-like chromosome carries the type I allele at 11 of the 13 common polymorphic sites. At sites H and O (fig. 1), one located within *NBR2* intron 1 and another within the 3′ portion of *BRCA1* exon 11, Ih chromosomes carry the type II allele. Ih also includes very rare variants that are located at site D, within the *LBRCA1* region that is similar to *BRCA1* exon 1a, and at site B, located in the 3′ UTR of the *NBR1* gene. In the Ih type, a rare exception to the association between the common sites is associated exclusively with rare alleles at sites that are located 50–80 kb distant, as indicated by the diamond-shaped markings in figure 2*A*.

A summary of the observed sequence relations between all the haplotypes in the U.S./European sample is presented in figure 2*B* and represents one possible evolutionary pathway. It is evident from figure 2 that mutation, not recombination, is likely to be responsible for the rarer haplotypes, since only haplotypes If and Ig might have arisen by crossover between chromosomes bearing the two major haplotypes. All other rare haplotypes appear to represent the serial accumulation of mutational difference, and the strong associations between distant rare sites appear to result from the occurrence of mutations at the same evolutionary branch point.

Allele-Length Distributions of STR Markers in the RNU2–BRCA1 *Region*

To determine whether STR marker alleles reflect a similar haplotype specificity and to further define the borders of the region of recombination suppression, we investigated marker-allele patterns for 14 STRs and one RFLP located in a broader region in the vicinity of *RNU2–BRCA1*. For each marker, ~30 individuals homozygous for haplotype I and a similar number homozygous for haplotype II were genotyped. Figure 3 summarizes the observed allele-length distributions in each group. The patterns are consistent with the absence of significant genetic interchange in a region including the entire *RNU2* locus and extending beyond the 3′ end of *BRCA1*. Two markers located ⩾300 kb beyond each end of this region, *D17S1333* and *D17S1185* (fig. 1),

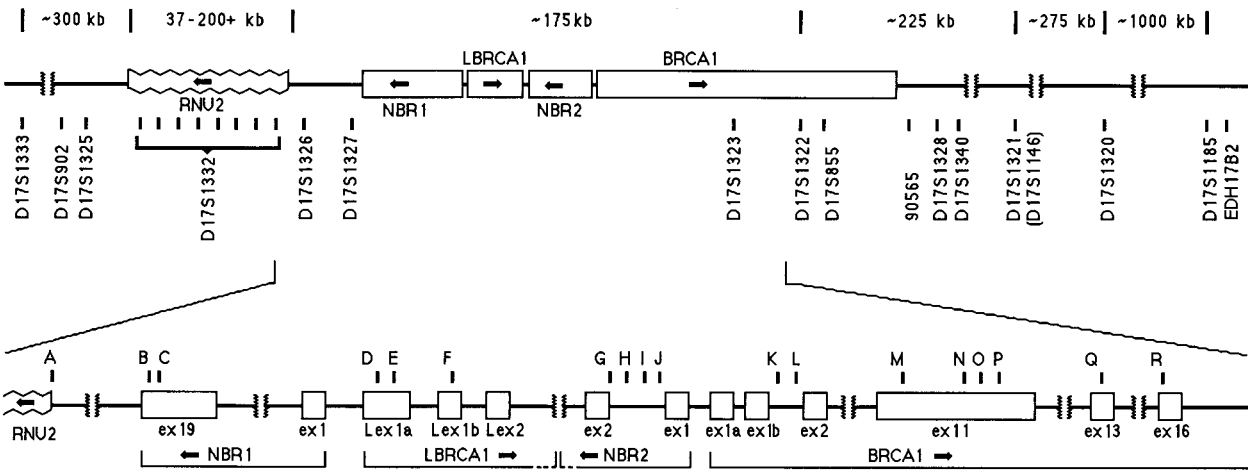

Figure 1 Summary of physical map distances, gene locations, STR sites, and other polymorphisms in the vicinity of the *RNU2–BRCA1* segment of chromosome 17. The indicated distances, as well as the nature and relative locations of other map structural features, are taken from published information (Albertsen et al. 1994; Couch et al. 1994, 1995; Jones et al. 1994; Neuhausen et al. 1994, 1996; Pavelitz et al. 1995; Barker et al. 1996*b;* Bennett-Baker et al. 1996; Brown et al. 1996; Smith et al. 1996; Xu et al. 1997*a*). The relative positions of two pairs of sites, *D17S902/D17S1325* and *D17S1185/EDH17B2,* are uncertain.

show similar patterns of allele distribution in each tested group. In contrast, all the STR markers located within the more immediate vicinity of *RNU2* and *BRCA1* have distinct allele length–frequency distributions for type I and type II chromosomes. For the multicopy intra-*RNU2* marker *D17S1332,* as well as *D17S1326, D17S1327, D17S1322,* and *BRCA1 90565,* the I- and II-specific distributions have little or no overlap. For *D17S1327,* nearly all I/I individuals are homozygous for the same allele, and II/II individuals carry 10 different frequent alleles, the smallest of which is 20 bp larger than the allele present on most type I chromosomes. For D17S1326, nearly all II/II individuals are homozygous, whereas I/I individuals carry five different alleles, with the smallest being 12 bp larger than the most common II-associated allele (fig. 3). In these samples, the *RNU2* tetranucleotide-repeat locus *D171332* is concordant with the biallelic marker pattern, with all type I chromosomes carrying the larger tetranucleotide allele and all type II chromosomes carrying the smaller allele. For *D17S1325, D17S1323, D17S855,* and *D17S1328,* the most common allele detected in I/I or II/II individuals is either different in size (*D1751325, D1751323,* and *D175855*), or is notably different in frequency (*D1751328*), and there are varying degrees of distinction in the overall shape and overlap of the allele-length distributions. Overall, the distinctive STR allele-distribution patterns occur in a region that extends at least from *D17S1325* through *D17S1328.* Small effects of uncertain significance may be evident at much greater distances, as indicated by the slightly skewed pattern observed for the biallelic polymorphism at *EDH17B2* (fig. 3).

RNU2 (D17S1332) *Repeat Polymorphism Patterns and Associations*

Acrylamide-gel analysis of all 275 haplotyped individuals from the U.S./European group shows that *D17S1332* conforms mainly to the homogeneous intra-array pattern, with some detectable variation, including forms that show striking associations with rare haplotypes of biallelic markers. The *RNU2* locus is composed of an array of 6 to >30 tandemly repeated 6.1-kb units, with each unit containing one copy of the 188-bp *U2 snRNA* gene (Westin et al. 1984; Pavelitz et al. 1995). The *D17S1324* STR corresponds to the dinucleotide array $(CT)_{\sim 70}$ that lies within each *RNU2* unit between nucleotides 5380 and 5521 as numbered in GenBank U57614. Intra-array variation of the *D17S1324* STR is frequent in humans, apparently because the mechanism that maintains near-homogeneity of the unit elements in each *RNU2* array is outpaced by the very high rate of mutation of the *D17S1324* STR (Liao and Weiner 1995). In contrast, all the elements of a single array are generally homogeneous for base-pair or small insertion/deletion (indel) polymorphisms (Liao et al. 1997). The *D17S1332* STR polymorphism (Bennett-Baker et al. 1996) corresponds to a variable tetranucleotide array, $(CAAA)_4$ or $(CAAA)_6$, at a location within each *RNU2* repeat bounded by nucleotides 490 and 507.

Examples of the different *D17S1332* allele patterns

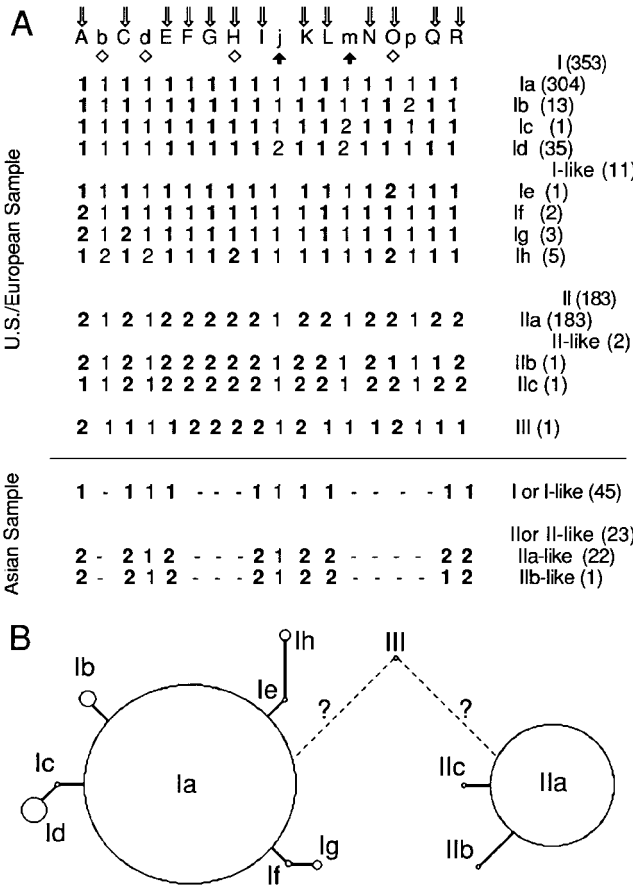

Figure 2 *A*, Summary of *RNU2–BRCA1* region haplotype structures observed in a U.S./European sample of 550 chromosomes and an Asian sample of 68 chromosomes. The number of chromosomes of each specific type and general class is shown (*right*). The top line shows each biallelic site as denoted in table 1 and figure 1. The commonly associated sites are indicated by capital letters and downward-pointing arrows. Diamond shapes and upward-pointing arrows indicate two sets of sites associated in rare haplotypes as described in the text. In each haplotype string, a "1" indicates the more frequent allele. Alleles at the common sites are in boldface. *B*, Summary of sequence and possible evolutionary relation of haplotypes observed in the U.S./European sample. Circles representing each haplotype class are not to exact scale.

are shown in figure 4. The correlation of haplotype I with $(CAAA)_6$ and of haplotype II with $(CAAA)_4$ predicts the major allelic PCR product bands for *D17S1332* in 268 of the 275 tested individuals and in 543 (99%) of 550 chromosomes. In the majority of these, the observed band patterns correspond to simple $(CAAA)_4$ or $(CAAA)_6$ arrays (fig. 4*A*). However, 48 of the 268 concordant samples showed one additional minor band at a position 4 bp larger than the $(CAAA)_6$ allelic band (fig. 4*B* and *C*). The presence and relative intensity of the minor band were confirmed in a duplicate experiment, showing that the variation reflects intra-array heterogeneity and is not a result of the random appearance of artifactual satellite bands of the type often observed with PCR of STRs (Hauge and Litt 1993). Because the additional bands always appeared in marker patterns that included at least one $(CAAA)_6$ allele, they are likely derived from *RNU2* arrays that contain mainly $(CAAA)_6$, with a small proportion of $(CAAA)_7$ derived from $(CAAA)_6$ by gain of a single CAAA. Of these 48 cases, 14 appeared to show greater intensity of the $(CAAA)_7$ band compared with the other 34 (compare fig. 4*C* with 4*B*), indicating a relatively higher $(CAAA)_7$:$(CAAA)_6$ unit ratio. Notably, these 14 cases included all 13 individuals with the rare form of the polymorphism at site P in exon 11 of *BRCA1*. This demonstrates a strong correlation between an *RNU2*-array variant and a site polymorphism located ~150 kb distant.

The seven chromosomes with *D17S1332* alleles that are not predicted by the haplotype of major biallelic sites in the adjacent region include all three occurrences of haplotype Ig and one of haplotype If. These four chromosomes carry the type I alleles at all sites except for one or two sites immediately adjacent to the *RNU2* array, where they carry a type II allele (fig. 2*A*). In these four cases, the adjacent *RNU2* array is also type II, possibly consistent with an origin by recombination between a type I and a type II chromosome. The precise location of the crossover point is undetermined, because some observed differences could also result from mutation. The significance of the latter possibility is highlighted by the second If chromosome as well as the single IIc chromosome, in which the status of site A, which is immediately adjacent to *RNU2*, indicates a possible recombination breakpoint (fig. 2*A*). However, in these two instances, the distal *RNU2*-array pattern is consistent with the status of biallelic sites B–R, suggesting that the site A difference is due to mutational variation.

The remaining three cases with distinctive *D17S1332* alleles provide further examples of associations of rare *RNU2*-array patterns with variation at distantly located sites. Individuals with haplotypes IIb and Ie, both of which are unique with respect to the pattern of biallelic sites in the 3′ portion of the *BRCA1* gene, also carry unique *RNU2* arrays. The Ie haplotype array (fig. 4*D*) is predominantly $(CAAA)_7$, with a small proportion of $(CAAA)_6$ and, possibly, $(CAAA)_5$; the IIb haplotype carries an *RNU2* array (fig. 4*E*) consisting primarily of $(CAAA)_8$ units. Finally, the type III chromosome, which is unlike either type I or type II and appears to be a mixed aggregate and possible ancestor of these types, carries an *RNU2* array that is unique in that it includes a near-equal mixture of $(CAAA)_8$ and $(CAAA)_9$, with small amounts of $(CAAA)_{10}$ and $(CAAA)_2$ (fig. 4*F*). In each of the individuals bearing these rare haplotypes and *RNU2* alleles, the alternate *RNU2* allele and the inferred

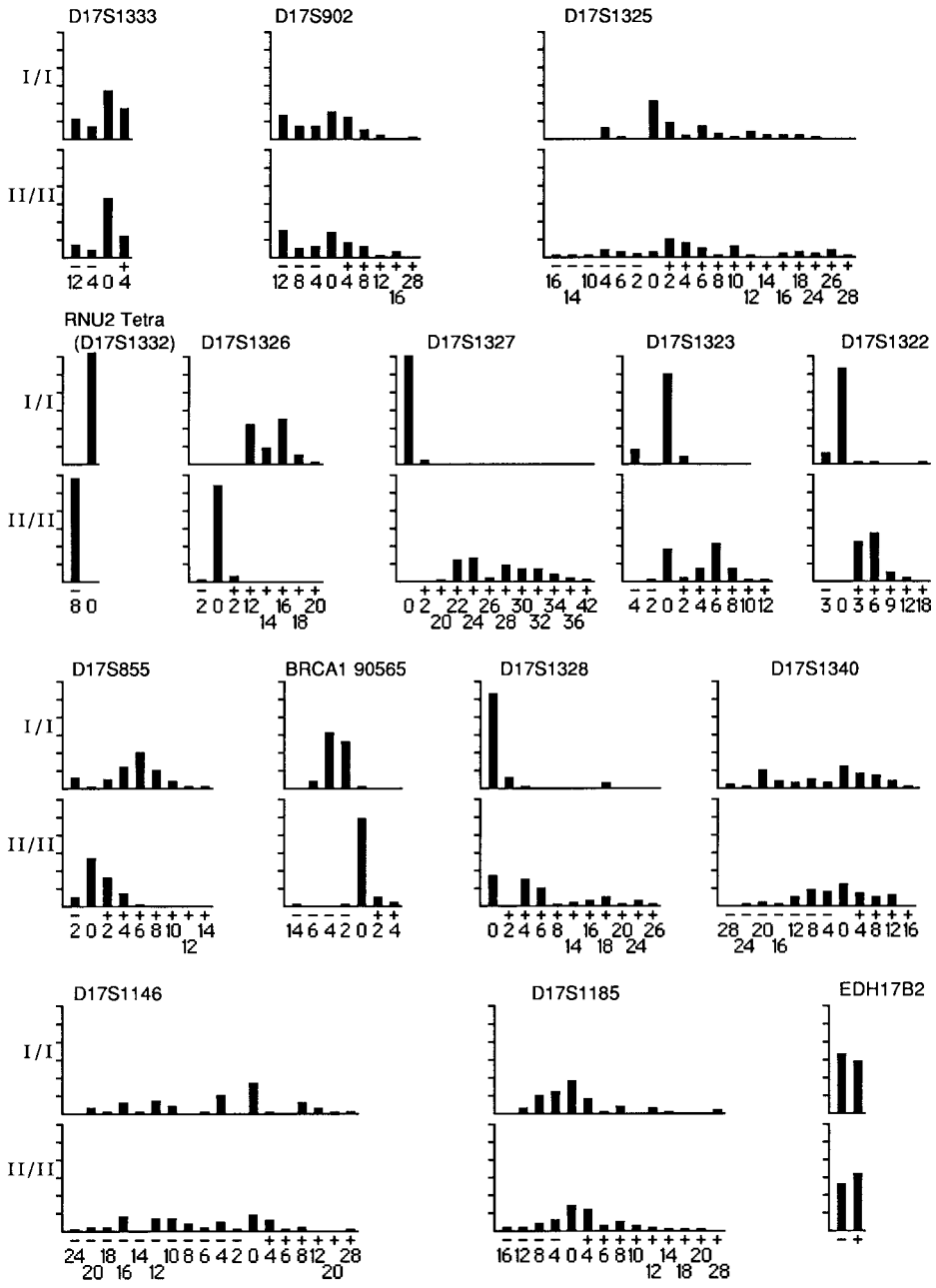

Figure 3 Haplotype-specific allele-length distributions for markers in the vicinity of *RNU2–BRCA1* as determined in type I and type II homozygous individuals. The markers are presented in the physical order described in figure 1. For each STR, the most common allele is designated as "0," and the nucleotide size difference of other alleles is as indicated. Allele lengths not observed are omitted from the length axis. For *EDH17B2*, the allelic forms are the presence (+) or absence (−) of the variable *Bfa*I site (Normand et al. 1993). Each distribution includes data from ~60 (range 56–62) chromosomes, and the marked units of the vertical axis represent 10 chromosomes.

alternate haplotype appeared to conform to the more prevalent associations.

Discussion

The unusually strong associations between biallelic polymorphisms in the region containing the *RNU2* array and the *BRCA1* gene define two major haplotypes, with frequencies of .64 and .33. The two haplotypes are highly conserved and account for nearly all chromosomes in European and Asian populations, which diverged ~100,000 years ago (4,000–5,000 generations). This high degree of conservation could be due to a mechanism specifically suppressing genetic exchange in this

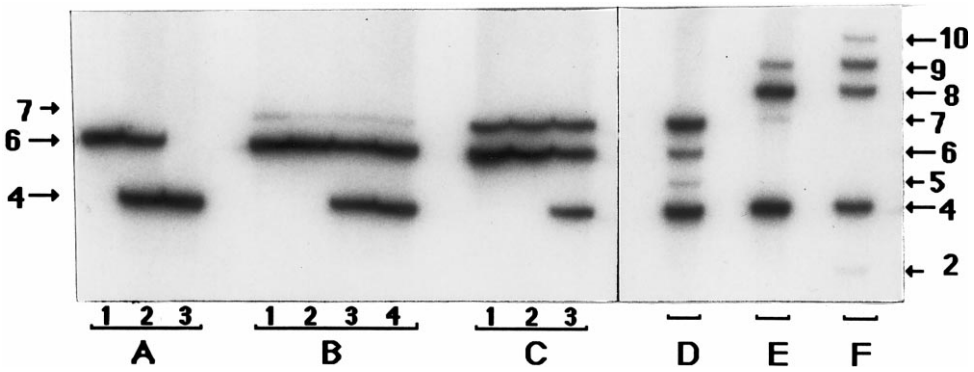

Figure 4 Representative patterns of *RNU2* tetranucleotide (*D17S1332*) polymorphism. Group A represents the most common homozygous and heterozygous patterns. Groups B and C contain, respectively, the additional weak or strong $(CAAA)_7$ band. The length of the CAAA array in both common allelic forms and in the strong additional band was determined by direct sequencing of genomic PCR products or products reamplified from acrylamide gel–purified bands. Lanes D–F represent individuals with rare *RNU2* alleles who also have rare haplotypes of biallelic sites as described in the text. The number of CAAA repeats corresponding to each observed product band in the very rare alleles is indicated, made on the basis of a sequence-ladder standard comparison. Direct sequencing of the rare pattern bands has not been done to rule out changes in the number of other tetranucleotide elements.

region and/or the nonviability of exchange products. The possibility that viable recombination occurs only between similar haplotypes seems unlikely, because there are also strong associations between lower-frequency variants located at physically distant sites. Examples of this include haplotypes Id and Ih (fig. 2) and the association of a rare *RNU2* array–structure variant with a rare site in distal *BRCA1*. If we assume that all these rare associations are not a result of recent admixture from isolated homogeneous populations, then their existence suggests that free recombination does not occur, even between chromosomes of the same haplotype class.

Complete or nearly complete concordance of alleles at polymorphic STR or RFLP sites separated by <10 kb is not unusual (Barker et al. 1984; Lifton et al. 1990; Sherrington et al. 1991; Normand et al. 1993; McGinnis and Spielman 1994; Castiglione et al. 1995; Charmley et al. 1995); however, for sites separated by ⩾100 kb, conservation of strong linkage disequilibrium is highly atypical. Studies of large segments in or near human genes *NF1* (Jorde et al. 1993; Valero et al. 1996), *APC* (Jorde et al. 1994), and *vWF* (Watkins et al. 1994) have detected significant associations, but in most cases these involve either a lesser degree of correlation than observed in the present study or only a portion of the examined sites or pairs of sites within a region, or the associations were not highly conserved in diverse population groups. An extensive review of published disequilibrium reports (Jorde et al. 1994) concluded that linkage disequilibrium is inversely correlated with physical distance for distances >50–60 kb and is greatly attenuated as the distance between sites approaches 500 kb. Aside from the *MHC* locus, which is excluded from the Jorde et al. (1994) analysis, because of the likelihood that disequilibrium there is influenced by selective forces (Begovich et al. 1992; Gyllensten and Erlich 1993), only the *NF1* locus (Jorde et al. 1993, 1994), located proximal to *BRCA1* on chromosome 17q, has properties that approach those of the region described here. For all pairings of RFLP sites within a region ~80 kb intragenic to *NF1*, *r* values of .98–1.0 were found in both Utah CEPH (northern European) and non-Utah CEPH (Venezualan) populations. Also at *NF1*, an *r* value of .82 was observed between sites located 340 kb apart in the Utah CEPH sample (168 chromosomes). However, the *r* value for the same pair of sites in the non-Utah CEPH sample (48 chromosomes) was −.83, indicating the predominance of an alternate haplotype in a different population group. For the *RNU2–BRCA1* region, the *r* value for the most distantly located (~175 kb) non-STR sites A and R (fig. 1) is .97 in the U.S./European group (550 chromosomes) and 1.0 in the Asian sample (68 chromosomes). Within this region, associations with *r* values in the range of .95–1 are found for all pairs of common sites in both the U.S./European sample and the Asian sample, consistent with the inference of nearly complete haplotype conservation, even in highly divergent populations. STR data for *D17S1332* in the U.S./European group indicate that the region of strong association includes the entire variable-length *RNU2* locus, with a range of ~37 to >200 kb. This is supported by the report by Liao et al. (1997), demonstrating strong association, in non-African populations, of sites located at the extremities of the *RNU2* array, including site A (table 1and fig. 1). The *BRCA1 90565* data indicate that the 33-kb segment between site R and the location of this STR in the intergenic region between the 3′ end of *BRCA1* and the neighboring *RHO7* gene (Smith et al. 1996) also has

significantly reduced recombination, so that the total length of the involved region is ⩾250 kb and completely spans at least four nonoverlapping transcriptional units (fig. 1).

Direct evidence that regions of high disequilibrium are maintained by recombination suppression has been obtained at the *MHC* locus that represents the extreme case of linkage disequilibrium in the human genome. For the highly polymorphic HLA *DQ1* and *DR1* sites, which are separated by 80 kb (Cullen et al. 1997), only a small proportion of the potential haplotypes have been found in any population, although the relative prevalence of haplotypes is variable in different ethnic groups (Gao et al. 1991; Doherty et al. 1992). If selective forces are involved in maintaining this high disequilibrium (Begovich et al. 1992; Gyllensten and Erlich 1993), they may also have resulted in the evolution of mechanisms that affect recombination rates in specific regions. For the *DQ1–DR1* region and adjacent segments of the *MHC* locus, CEPH reference families have been studied to determine the locations and frequencies of observable recombination events, and the results are consistent with the regions of relatively high and low disequilibrium (Martin et al. 1995; Cullen et al. 1997; Malfroy et al. 1997). Although similarly detailed analyses are not yet available for the *RNU2–BRCA1* region, it is notable that, in the extensive search for recombinant chromosomes to refine the location of *BRCA1,* the meiotic breaks reported generally involved markers no closer than 0.5–1 Mb (Albertsen et al. 1994; Tonin et al. 1994) and that the closest documented meiotic breaks (Neuhausen et al. 1994) involved markers *D17S1325* and *D17S1321* (*D17S1146*), which appear to lie outside the region where recombination suppression is most extreme.

Overall, the STR marker-allele distributions associated with each of the two major *RNU2–BRCA1* haplotypes support the hypothesis that several of these loci have evolved within a zone of complete recombination suppression. For these loci, the haplotype-specific allele-length distributions are best explained by founder effects and mutational histories specific to each class of haplotype. STR markers are believed to evolve by relatively frequent mutation most often involving gain or loss of single-repeat units, caused by "replication slippage" (Weber and Wong 1993), and the probability of mutation increases with the length of the tandem array (Weber 1990; Brinkmann et al. 1998). For markers for which the allele-length distributions for types I and II show little or no overlap, it is likely that the ancestral chromosomes carried different founder alleles at the time of divergence and that the initial differences and length-dependent mutation rates resulted in the observed distinct distributions. This is most apparent for *D17S1326* and *D17S1327,* in which haplotypes II and I, respectively, show a predominant single allele that is much smaller than the variety of alleles on the alternate haplotype (fig. 3). For markers such as *D17S1333, D17S902, D17S1146,* and *D17S1185,* which show great allele diversity for type I and type II chromosomes, normal recombination between these loci and the conserved segment would explain the similar pattern of alleles occurring on both types of chromosomes; it is also possible, however, that there was little difference between alleles on the founder chromosomes or that very fast STR mutation rates have masked any founder-allele effect. For STRs for which there is an intermediate degree of overlap in the distributions, it is not yet possible to assess the contributions of founder alleles, mutation frequency, or the rate of meiotic exchange between the STR site and the region of suppressed recombination. Because two STRs displaying haplotype-specific allele distributions, *D17S1332* and *BRCA 190565,* are located outside the region containing the biallelic sites assayed by ASO, these loci appear to define the least-outer bounds of the region of recombination suppression. However, the possibility cannot be excluded that STR sites that are located beyond these points and that do not have extreme haplotype-specific distributions also fall within the region of recombination suppression.

These findings have an important impact on the interpretation of identity of STR marker haplotypes found on independently ascertained chromosomes carrying the same *BRCA1* mutation. Figure 3 shows that the potential variety of different haplotypes of the STR markers closest to *BRCA1* is much less than would be expected if associations between alleles at different STR sites were random. Specific associations not evident in figure 3, between individual alleles of different STR loci, may also exist, further reducing potential haplotype variability. In cases in which inferences about the likelihood of common ancestry are made by STR haplotype analysis, it is necessary to assess the actual population frequency of the involved haplotype, to avoid overestimation of the significance of its cooccurrence with a specific mutation.

Significant revisions are also necessary for attempts to compute both the time of origin of *BRCA1* mutations found in multiple, apparently unrelated families and explicit values for the likelihood of common ancestral origin. One such effort (Neuhausen et al. 1996) assumed a model in which recombination across the *BRCA1* region is similar to the average of the human genome. The error involved in this assumption results in the underestimation of the age of any mutation with a founder effect, because the deficit of haplotype diversity is interpreted as reflecting a more recent origin. It is notable that although Neuhausen et al. (1996) proposed 460–1,600 years ago as the time of origin of the *BRCA1* 185delAG mutation, further studies of additional Jewish populations (Abeliovich et al. 1997; Levy-Lahad et al.

1997; Bar-Sade et al. 1998) establish the origin of the mutation prior to the geographical dispersion of Jews that occurred >1,900 years ago—and a true age of >2,000 years. Attempts to calculate the likelihood that two chromosomes bearing identical mutations have the same ancestral origin also involve assumptions regarding mutation rates at the STR loci used for haplotype analysis, as well as at the site of the mutation, and are sensitive to changes in any of these rates (Neuhausen et al. 1996). For any STR locus, the rate of mutation is strongly dependent on the initial allele size (Weber 1990; Weber and Wong 1993; Brinkmann et al. 1998). The small founder alleles of *D17S1327* and *D17S1326* apparently have remained quite stable, with little allelic diversification evident after a long period of human population expansion, and the presumed larger founder alleles on the alternate haplotype have engendered nearly all the variety evident at these loci. Any attempt to use STR mutation rates as an accurate molecular clock will require close attention to locus-specific and allele-specific mutation rates. Similar constraints apply to the estimates of rates of indel and base-pair mutations affecting gene function, because DNA structural and sequence context have an important role in determining the likelihood of specific mutations (Krawzcak and Cooper 1991, 1996; Rodenhiser et al. 1996; Krawczak et al. 1998).

The absence of recombinational reassortment at this locus also raises questions about the evaluation of the functional significance of rare variants of *BRCA1*. Although the two common haplotypes differ at all of the common sites described here, and probably at many other known (Couch et al. 1996) or as yet undetected sites, there does not appear to be any major distinction in cancer susceptibility associated with the totality of this class of differences (Dunning et al. 1997). However, numerous rare sequence variants are very likely to exist, and epidemiological studies could establish associations with disease susceptibility. For example, Newman et al. (1998) found a low-frequency variant in the *BRCA1* 3′ UTR significantly more often in black women with breast cancer than in matched controls. In cases such as this, there is a possibility that one or more additional sequence differences in strong linkage disequilibrium with the assayed site and located anywhere within the region of strong disequilibrium could be involved in causing the observed effect. This possibility is illustrated by our finding of rare variants within the *RNU2* locus that are very strongly associated with rare sites or constellations of sites ~150 kb away, within exon 11 of *BRCA1*. Similarly, in cases in which functional evaluation of any specific detected variant by artificial gene constructs (Humphrey et al. 1997) leads to the conclusion that it has no effect on gene activity, the possibilities must be considered that the variant is in strong linkage disequilibrium with a second distantly located sequence change that does have a functional effect or that the combination of the tested variant with one or more associated but undetected sequence differences is etiologically significant. Such situations could cause apparent conflicts in the conclusions of epidemiological versus functional evaluation approaches.

The distinctive properties of the *RNU2–BRCA1* region may provide a significant new tool for studies of the origin of modern human populations. The theory of recent human origin in Africa predicts that greater diversity of haplotypes from the *RNU2–BRCA1* region will be found in African populations, with a subset of these being similar to those found in other regional populations (Armour et al. 1996; Mountain and Cavalli-Sforza 1997). However, if non-African populations have ancestors from more than one founding lineage (Harding et al. 1997), then the strongly conserved haplotypes in this region may help to detect those secondary lineages, serving as one of the required "nuclear loci that allow us to quantitate subdivisions in the time range of 800,000–150,000 years ago" (Cann 1997, p. 757). Indeed, for this locus, the extremely low rate of recombination, the traceable accretion of point mutations, and "slow" and "fast" STR mutation rates offer the possibility of distinct molecular clocks that, with appropriate characterization, may prove useful for the study of genetic relations between population groups with recent or ancient common ancestors.

The mechanism and/or functional basis for the apparent recombination suppression at *RNU2–BRCA1* remains obscure, although the structural features of this region, the transcriptional arrangement, and the function of the *BRCA1* gene in meiosis suggest several hypotheses. The *RNU2–BRCA1* region includes the variable-length *RNU2* tandem array, as well as the genomic duplication that includes the *NBR1* and *BRCA1* promoter regions. Recombination suppression in and near large tandem arrays, particularly polymorphic ones, may be a general phenomenon in complex genomes. Alternatively, because the relative sequence homogeneity of the individual tandem elements of the *RNU2* locus is maintained by a mechanism that operates over all the elements of an existing array (Pavelitz et al. 1995), it is possible that genomic elements involved in this mechanism may influence the sites of initiation and/or resolution of chiasma in this region. The involvement of the *BRCA1* gene product in genomic recombination and/or repair (Callebaut and Mornon 1997; Scully et al. 1997), as well as expression of the *BRCA1* gene near the time of meiosis, as found in mice (Zabludoff et al. 1996; Blackshear et al. 1998), may be important. Transcriptional activity at the paralogous bidirectional promoters could prevent recombination, or specific mechanisms may exist to regulate recombination in regions encoding proteins directly involved in this process. If so, regions

with similar properties might occur at the sites of either genes expressed at the time of meiosis or other repair/recombination genes. An additional possibility is that maintenance of the observed associations preserves a significant, as yet unrecognized function that requires the interaction of allele-specific elements that are distantly located on the same chromosome, as postulated for HLA. Finally, the observation that the *NF1* locus that is also located on proximal 17q has disequilibrium properties nearly as extreme as those described here may indicate the existence of recombination control mechanisms operating at a yet higher level of genome organization. Further insights into the recombination properties, genetic functions, and genomic structures in this and other regions with extreme disequilibrium are needed to resolve these issues.

Acknowledgments

We thank Joyce C. Denison, Eduardo Almeida, and Matthew Donaldson for help with human sample collection and organization, DNA preparation, and other technical support. Financial support was provided by National Cancer Institute grant CA63689 and NCI CA77453 (both to D.B.) and National Institutes of Health grant CA42014 (to the University of Utah Cancer Center core peptide and DNA synthesis facility).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank/(for L78833, U57614, X76952, U72483, U37574, U14680, and M27138 DNA and cDNA sequences)

Genome Database, http://www.gdb.org/ (for markers and primers)

References

Abeliovich D, Kaduri L, Lerer I, Weinberg N, Amir G, Sagi M, Zlotogora J, et al (1997) The founder mutations 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashnkenazi women. Am J Hum Genet 60:505–514

Albertsen HM, Smith SA, Mazoyer S, Fujimoto E, Stevens J, Williams B, Rodriquez P, et al (1994) A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21. Nat Genet 7:472–479

Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, et al (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13: 154–160

Balnaves ME, Nasioulas S, Dahl H-H M, Forrest S (1990) Direct PCR from CVS and blood lysates for detection of cystic fibrosis and Duchenne muscular dystrophy deletions. Nucleic Acids Res 19:1155

Barker DF, Almeida ERA, Casey G, Fain PR, Liao S-Y, Masunaka I, Noble B, et al (1996*a*) BRCA1 R841W: a strong candidate for a common mutation with moderate phenotype. Genet Epidemiol 13:595–604

Barker DF, Fain PR (1993) Definition and mapping of STSs at STR and RFLP loci in Xp11-Xq22. Genomics 18: 712–716

Barker DF, Holm T, White R (1984) A locus on chromosome 11p with multiple restriction site polymorphisms. Am J Hum Genet 36:1159–1171

Barker DF, Liu X, Almeida ERA (1996*b*) The BRCA1 and 1A1.3B promoters are parallel elements of a genomic duplication at 17q21. Genomics 38: 215–222

Bar-Sade RB, Kruglikova A, Modan B, Gak E, Hirsh-Yechezkel G, Theodor L, Novikov I, et al (1998) The 185delAG BRCA1 mutation originated before the dispersion of Jews in the diaspora and is not limited to Ashkenazim. Hum Mol Genet 7:801–805

Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, Erlich HA, et al (1992) Polymorphism, recombination and linkage disequilibrium with the HLA class II region. J Immunol 148:249–258

Bennett-Baker PE, Kiousis S, Chandrasekharappa SC, King SE, Abel KJ, Collins FS, Weber BL, et al (1996) Isolation of tetranucleotide repeat polymorphisms flanking the BRCA1 gene. Genomics 32:163–167

Blackshear PE, Goldsworthy SM, Foley JF, McAllister KA, Bennett LM, Collins NK, Bunch DO, et al (1998) BRCA1 and BRCA2 expression patterns in mitotic and meiotic cells of mice. Oncogene 16:61–68

Blouin JL, Christie DH, Gos A, Lynn A, Morris MA, Ledbetter DH, Chakravarti A, et al (1995) A new dinucleotide repeat polymorphism at the telomere of chromosome 21q reveals a significant difference between male and female rates of recombination. Am J Hum Genet 57:388–394

Brinkmann B, Klintschar M, Neubuhr F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Brown MA, Xu C-F, Nicolai H, Griffiths B, Chambers JA, Black D, Solomon E (1996) The 5′ end of the BRCA1 gene lies within a duplicated region of chromosome 17q21. Oncogene 12:2507–2513

Buetow KH, Shiang R, Yang P, Nakamura Y, Lathrop GM, White R, Wasmuth JJ, et al (1991) A detailed multipoint map of human chromosome 4 provides evidence for linkage heterogeneity and position-specific recombination rates. Am J Hum Genet 48:911–925

Callebaut I, Mornon JP (1997) From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. FEBS Lett 400:25–30

Campbell IG, Nicolai HM, Foulkes WD, Senger G, Stamp GW, Allan G, Boyer C, et al (1994) A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. Hum Mol Genet 3:589–594

Cann RL (1997) Phylogenetic estimation in humans and neck riddles. Am J Hum Genet 60:755–757

Castiglione CM, Deinard AS, Speed WC, Sirugo G, Rosen-

baum HC, Zhang Y, Grandy DK, et al (1995) Evolution of haplotypes at the DRD2 locus. Am J Hum Genet 57: 1445–1456

Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, et al (1994) Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. Nat Genet 8:387–391

Chakravarti A, Buetow KH, Anonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination with the human β-globin gene cluster. Am J Hum Genet 36: 1239–1258

Charmley P, Concannon P, Hood L, Rowen L (1995) Frequency and polymorphism of simple sequence repeats in a contiguous 685-kb DNA sequence containing the human T-cell receptor β-chain gene complex. Genomics 29:760–765

Couch FJ, Castilla LH, Xu J, Abel KJ, Welcsh P, King SE, Wong L, et al (1995) A YAC-, P1- and cosmid-based map of the BRCA1 region on chromosome 17q21. Genomics 25: 264–273

Couch FJ, Kiousis S, Castilla LH, Xu J, Chandrasekharappa SC, Chamberlain JS, Collins FS, et al (1994) Characterization of 10 new polymorphic dinucleotide repeats and generation of a high-density microsatellite-based physical map of the BRCA1 region of chromosome 17q21. Genomics 24: 419–424

Couch FJ, Weber BL, Breast Cancer Information Core (1996) Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. Hum Mutat 8:8–16

Cullen M, Noble J, Erlich H, Thorpe D, Beck S, Klitz W, Trowsdale J, et al (1997) Characterization of recombination in the HLA class II region. Am J Hum Genet 60:397–407

Doherty DG, Vaughan RW, Donaldson PT, Mowat AP (1992) HLA DQA, DQB, and DRB genotyping by oligonucleotide analysis: distribution of alleles and haplotypes in British Caucasoids. Hum Immunol 34:53–63.

Dunning AM, Chiano M, Smith NR, Dearden J, Gore M, Oakes S, Wilson C, et al (1997) Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population. Hum Mol Genet 6:285–289

Durocher F, Shattuck-Eidens D, McClure M, Labrie F, Skolnick MH, Goldgar DE, Simard J (1996) Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations. Hum Mol Genet 5:835–842

Fain PR, Goldgar DE, Wallace MR, Collins FS, Wright E, Nguyen K, Barker DF (1989) Refined physical and genetic mapping of the NF1 region on chromosome 17. Am J Hum Genet 45:721–728

Friedman LS, Ostermeyer EA, Szabo CI, Dowd P, Lynch ED, Rowell SE, King M-C (1994) Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nat Genet 8:399–404

Gao XJ, Sun YP, An JB, Fernandez-Vina M, Qou JN, Lin L, Stastny P (1991) DNA typing for HLA-DR and -DP alleles in a Chinese population using the polymerase chain reaction (PCR) and oligonucleotide probes. Tissue Antigens 38: 24–30

Gyllensten UB, Erlich HA (1993) MHC class II haplotypes and linkage disequilibrium in primates. Hum Immunol 36:1–10

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African *and* Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Hauge XY, Litt M (1993) A study of the origin of "shadow bands" seen when typing dinucleotide repeat polymorphisms by the PCR. Hum Mol Genet 2:411–415

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Popul Biol 33:54–78

Humphrey JS, Salim A, Erdos MR, Collins FS, Brody LC, Klausner RD (1997) Human BRCA1 inhibits growth in yeast: potential use in diagnostic testing. Proc Natl Acad Sci USA 94:820–5825

Inoue R, Fukutomi T, Ushijima T, Matsumoto Y, Sugimura T, Nagao M (1995) Germline mutation of BRCA1 in Japanese breast cancer families. Cancer Res 55:3521–3524

Jones KA, Black DM, Brown MA, Griffiths BL, Nicolai HM, Chambers JA, Bonjardim M, et al (1994) The detailed characterisation of a 400 kb cosmid walk in the BRCA1 region: identification and localisation of 10 genes including a dual-specificity phosphatase. Hum Mol Genet 3:1927–1934

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 54:884–898

Jorde LB, Watkins WS, Viskochil D, O'Connell P, Ward K (1993) Linkage disequilibrium in the neurofibromatosis 1 (NF1) region: implications for gene mapping. Am J Hum Genet 53:1038–1050

Klitz W, Stephens JC, Grote M, Carrington M (1995) Discordant patterns of linkage disequilibrium of the peptide-transporter loci within the HLA class II region. Am J Hum Genet 57:1436–1444

Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitutions in human genes. Am J Hum Genet 63:474–488

Krawczak M, Cooper DN (1991) Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. Hum Genet 86: 425–441

———(1996) Single base-pair substitutions in pathology and evolution: two sides to the same coin. Hum Mutat 8:23–31

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky E, Shohat M, et al (1997) Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. Am J Hum Genet 60: 1059–1067

Lewontin RC, Kojima KI (1960) The evolutionary dynamics of complex polymorphisms. Evolution 14:458–472

Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J 16:588–598

Liao D, Weiner AM (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the $(CT)_n \cdot (GA)_n$ microsatellite embedded within the U2 repeat unit. Genomics 30:583–593

Lichten M, Goldman ASH (1995) Meiotic recombination hotspots. Annu Rev Genet 29:423–444

Lifton RP, Sardet C, Pouyssegur J, Lalouel J-M (1990) Cloning of the human genomic amiloride-sensitive Na+/H+ antiporter gene, identification of genetic polymorphisms, and localization on the genetic map of chromosome 1p. Genomics 7:131–135

Malfroy L, Roth MP, Carrington M, Borot N, Volz A, Ziegler A, Coppin H (1997) Heterogeneity in rates of recombination in the 6-Mb region telomeric to the human major histocompatibility complex. Genomics 43:226–231

Martin M, Mann D, Carrington M (1995) Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. Hum Mol Genet 4:423–428

McGinnis RE, Speilman RS (1994) Linkage disequilibrium in the insulin gene region: size variation at the 5′ flanking polymorphism and bimodality among "class I" alleles. Am J Hum Genet 55:526–532

Miki Y, Swensen J, Shattuck-Eidens D, Futreal P A, Harshman K, Tavtigian S, Liu Q, et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66–71

Mountain JL, Cavalli-Sforza LL (1997) Multilocus genotypes, a tree of individuals and human evolutionary history. Am J Hum Genet 61:705–718

Nagaraja R, MacMillan S, Kere J, Jones C, Griffin S, Schmatz M, Terrell J, et al (1997) X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. Genome Res 7:210–222

Neuhausen SL, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlinson G, et al (1996) Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. Am J Hum Genet 58:271–280

Neuhausen SL, Swensen J, Miki Y, Liu Q, Tavtigian S, Shattuck-Eidens D, Kamb A, et al (1994) A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene BRCA1. Hum Mol Genet 3:1919–1926

Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King M-C (1998) Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. JAMA 279:915–921.

Normand T, Narod S, Labrie F, Simard J (1993) Detection of polymorphisms in the estradiol 17β-hydroxysteroid dehydrogenase II gene at the EDH17B2 locus on 17q11-q21. Hum Mol Genet 2:479–483

Pavelitz T, Rusche L, Matera AG, Scharf JM, Weiner AM (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. EMBO J 14:169–177

Rodenhiser D, Chakraborty P, Andrews J, Ainsworth P, Mancini D, Lopes E, Singh S (1996) Heterogeneous point mutations in the BRCA1 breast cancer susceptibility gene occur in high frequency at the site of homonucleotide tracts, short repeats and methylatable CpG/CpNpG motifs. Oncogene 12:2623-2629

Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, et al (1997) Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell 88:265–275

Sherrington R, Melmer G, Dixon M, Curtis D, Mankoo B, Kalsi G, Gurling H (1991) Linkage disequilibrium between two highly polymorphic microsatellites. Am J Hum Genet 49:966–971

Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, et al (1996) Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. Genome Res 6:1029–1049

Tonin P, Serova O, Simard J, Lenoir G, Feunteun J, Morgan K, Lynch H, et al (1994) The gene for hereditary breast-ovarian cancer, BRCA1, maps distal to EDH17B2 in chromosome region 17q12–q21. Hum Mol Genet 3:1679–1682

Valero MC, Velasco E, Valero A, Moreno F, Hernandez-Chico C (1996) Linkage disequilibrium between four intragenic polymorphic microsatellites of the NF1 gene and its implications for genetic counselling. J Med Genet 33:590–593

Watkins WS, Zenger R, O'Brien E, Nyman D, Eriksson AW, Renlund M, Jorde LB (1994) Linkage disequilibrium patterns vary with chromosomal location: a case study from the von Willebrand factor region. Am J Hum Genet 55: 348–355

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Weber JL, Wong C (1993) Mutations of human short tandem repeats. Hum Mol Genet 2:1123–1128

Weir BS (1990) Genetic analysis. Sinauer Associates, Sunderland, MA

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, et al (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Westin G, Zabielski J, Hammarström K, Monstein H-J, Bark C, Pettersson U (1984) Clustered genes for human U2 RNA. Proc Natl Acad Sci USA 81:3811–3815

Xu C-F, Brown MA, Nicolai H, Chambers J, Griffiths B, Solomon E (1997*a*) Isolation and characterisation of the NBR2 gene which lies head to head with the human BRCA1 gene. Hum Mol Genet 6:1057–1062

Xu C-F, Chambers JA, Solomon E (1997*b*) Complex regulation of the BRCA1 gene. J Biol Chem 272:20994–20997

Zabludoff SD, Wright WW, Harshman K, Wold BJ (1996) BRCA1 mRNA is expressed highly during meiosis and spermiogenesis but not during mitosis of male germ cells. Oncogene 13:649–653