Human, canine and murine BRCA1 genes: sequence comparison among species

Csilla I. Szabo1,*, Lori A. Wagner2,*, Leigh V. Francisco2, Jared C. Roach3, Rhoda Argonza1, Mary-Claire King1 and Elaine A. Ostrander2, *

1 Division of Medical Genetics, Box 357720, University of Washington, Seattle, WA 98195, USA, 2 Clinical Research Division-M318, 1124 Columbia St., Fred Hutchinson Cancer Research Center, Seattle, WA 98104, USA and 3 Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195, USA

Received April 17, 1996; Revised and Accepted June 28, 1996

Five to ten percent of breast cancer in the western world may be attributed to the inheritance of highly penetrant mutations in the breast and ovarian cancer susceptibility gene, BRCA1. The biological function of BRCA1 and factors affecting expressivity, such as gene–environment and gene–gene interactions, may be more effectively studied in appropriate animal models. We report the cloning and sequencing of the canine and murine BRCA1 genes and contrast the sequences with human BRCA1. The amino terminal 120 residues of the gene are >80% identical among the three species. The C-terminus is also highly conserved, containing an 80 amino acid stretch that is over 80% identical. Motifs of likely functional significance are maintained, including the amino terminal RING finger motif (amino acids 24–64) and the gramin consensus sequence (1214–1223). The distribution of missense mutations and neutral polymorphisms identified in BRCA1-linked breast cancer suggests that disease associated missense mutations occur at highly conserved residues whereas polymorphisms are in regions of lower conservation. Among eighteen missense mutations with unknown consequences, seven occur in amino acids that are identical across species. Four of these seven (E1219D, A1708E, P1749R and M1775R) are also within conserved domains. Taken together, these data predict regions of the gene which may be critical for normal function.

INTRODUCTION

Breast cancer is the most common malignancy among women, with a cumulative lifetime risk of one in eight for a female born in 1990 (1). Five to ten percent of all breast cancer is attributable to inheritance of a highly penetrant mutation in an autosomal dominant susceptibility gene (2,3). Linkage analysis and positional cloning strategies have led to the recent isolation of two novel breast-ovarian cancer susceptibility genes, BRCA1 (4,5) and BRCA2 (6,7). Inherited mutations in either gene confer a lifetime risk of breast cancer greater than 80% and an increased risk of ovarian cancer (2,8,9). BRCA2 mutation carriers appear predisposed to a broader spectrum of cancers, with a striking elevation of male breast cancer risk (6,7).

Several lines of evidence indicate that BRCA1 and BRCA2 act as tumor suppressors. The majority of mutations identified in both BRCA1 (5,10–28) and BRCA2 (7,29–33) are protein truncations, likely representing loss of function alterations characteristic of tumor-suppressor genes. Among breast and ovarian tumors from patients with inherited BRCA1 mutations, >90% have lost the wild-type allele (34–36). Similarly, all chromosome 13q12–13 losses detected in tumors from BRCA2 carriers involve the wild-type allele (37). Although few somatic point mutations or frame-shift alterations have been identified thus far (10,15,16,19), gross somatic mutations at BRCA1 are frequently found in breast and ovarian tumors from patients not selected for family history. LOH in the BRCA1 region ranges from 40% to 80% among sporadic breast carcinomas (9,38,39) and from 30–70% among sporadic ovarian carcinomas (19,38,40–42). LOH frequencies at the BRCA2 locus are more difficult to assess, as analyses are confounded by simultaneous loss of the neighboring RB1 tumor suppressor gene (43,44). A growth inhibitory function for BRCA1 is also suggested by the decreased levels of BRCA1 expression in breast tumors from patients not selected for family history, ranging from none detectable to approximately 50% of that in normal breast epithelial cells (45). Experimental inhibition of BRCA1 expression with antisense oligonucleotides indicates a highly tissue-specific effect, leading to accelerated growth of normal and malignant mammary epithelial cells, but not of epithelial cells derived from non-mammary tissues (45).

Despite evidence of tumor suppressive function, the absence of significant homology between BRCA1 or BRCA2 and known genes provides few clues about the mechanisms by which these genes exert their growth regulatory effects. In the case of BRCA1, the significance of the amino terminal C3H4 RING finger motif (46) was underscored through identification of germline (12,13) and somatic (15) missense mutations of conserved cysteine residues. Protein truncating mutations occur the entire length of BRCA1 (5,10–27), with mutations in the 3′ portion conferring a
lower risk of ovarian cancer than mutations in the 5′ end (22,47). Nevertheless, a mutation truncating the extreme C-terminal 11 amino acids leads to very early onset breast cancer, indicating this region is essential for the normal function, conformation or stability of the protein (13). The apparent lack of mutational clustering has hindered the identification of additional functionally important domains.

Conservation of protein domains in evolutionarily divergent species may reveal regions critical to gene function. In the case of p53, protein domains conserved between human and mouse species may reveal regions critical to gene function. In the case of Rb: (54–57) and led to identification of genetic modifier loci which mitigate the effects of the original mutation (58).

Similarities in biological, epidemiologic, and apparent genetic characteristics of canine and human mammary tumors suggest that dogs may provide a useful model for the study of human breast cancer (59–61). First, as with humans, the most common malignant canine breast tumors (95%) are adenocarcinomas arising from the glandular epithelial tissue (62,63). Second, as with humans, canine breast cancer is a disease that usually occurs late in life; the median age of incidence is about 10 years, and tumors rarely occur before 5 years of age (64). Third, necropsy studies suggest that high rates of metastasis are found in dogs. This is comparable to that seen in humans, but contrasts with the low rates typically found in rodents (65). Fourth, most canine breast tumors metastasize to the regional lymph nodes and lung, as do human breast tumors (66,67). Fifth, canine tumors, like human tumors, have estrogen and progesterone receptors (68,69). Finally, hormonal effects are evident, and early ovariectomy in dogs (before 2.5 years of age) confers some protection against breast cancer (60,70). Inherited cases of breast cancer have been reported in canines, although virtually nothing is known about genes which might be responsible (71,72).

In an effort to identify existing animal models for BRCA1 or to establish genetically engineered models, we have cloned the canine and murine BRCA1 genes and present a cross-species comparison of the human, canine and murine [this manuscript and (73–76)] BRCA1 homologues.

RESULTS

Sequence analysis

A PCR-based strategy was used to clone the canine BRCA1 gene in sequential overlapping fragments. The majority of exon 11, which encompasses about 60% of the gene, was initially cloned from canine genomic DNA amplified using human BRCA1 primers. The remainder of the gene was cloned using PCR-amplified canine cDNA and primers derived from human sequence, or a combination of human primers and primers derived from the canine sequence. To minimize sequence errors due to early cycle misincorporation during PCR, approximately 90% of the 5637 bp coding sequence was independently amplified, cloned and sequenced with a minimum of two-fold redundancy. Because of difficulty in obtaining independent clones, however, two regions were sequenced from only one clone: between base pairs 4945–5431 (bp 5055–5526 in the human sequence) and between base pairs 1–118 (bp 120–237 in the human sequence). A minimum redundancy of two independent cDNA clones and/or PCR fragments was also used to assemble 5593 nucleotides of murine BRCA1 sequence encompassing a 5437 bp open reading frame. A region corresponding to nearly 80% of human BRCA1 exon 11 (HSU14680 nt 742–3480) was PCR amplified in overlapping fragments directly from Balb/C genomic DNA, indicating that this large human exon (5) is conserved in mouse, as it is in the canine genome.

Several cDNA variants were identified among both canine and murine clones spanning the junction between human BRCA1 exons 2 and 3 (HSU14680 nt 199/200). Of six canine clones, three encode an amino acid sequence identical to human and two had a 39 bp in-frame insertion at the boundary of exons 2 and 3, predicted to encode 13 amino acids: HLPMLETQNANTR. Interestingly, comparison of this sequence against the total human genomic DNA sequence showed significant homology to a 39 bp sequence of intron two, as shown below.

**Human intronic DNA:** CAC TTA TCC GTA TTG GAA GCT

**Canine 39 bp insert:** T CAC TTA CCC ATG TTG GAA ACC CAA AAT GCA AAT ACG CG.

In the human genomic sequence, intron two is 8.2 kb in length; the sequence homologous to the insert occurs 3.9 kb 3′ of exon 2 (77). Sequences flanking the human homologue of the canine cDNA insert resemble 3′ and 5′ splice sequences. Although these are not perfect matches to the consensus, they are similar to the corresponding sequences flanking exons 2 and 3, suggesting that they may function to generate alternatively spliced BRCA1 transcripts. In addition to this variant, one canine clone had the same 39 bp insert preceded by an additional 44 base pairs producing an immediate stop codon at the junction of exon 2 and the insert. One of the three mouse 5′ cDNA clones was also found to contain a 75 bp insertion at the splice junction of human BRCA1 exons 2 and 3 which would result in an in-frame addition of 25 amino acids (RGSSFLWLTWSAGTVVKLFEAGS). These variants have not been noted in any human clones reported to date. Hence, they may reflect the occurrence of either alternative splicing or incomplete splicing which occurred in vivo, or they may simply be artefacts of reverse transcription which occurred in vitro. Incomplete splicing is unlikely, at least for the canine variant, since the 39 bp insert does not appear to be immediately adjacent to exons two or three, as noted by the absence of PCR product when primers derived from the insert are tested in combination with exon two or exon three primers on genomic DNA.

Pairwise comparison of the human, dog and mouse homologues of BRCA1 at the nucleotide level using the DNA* alignment program (Madison, WI) revealed high cross-species conservation. The human and canine genes are 84% identical at the nucleotide level, whereas the murine BRCA1 gene retains only 72% identity to human. The murine and canine homologues are 69% identical. The predicted canine BRCA1 protein consists of 1878 amino acids, whereas the murine homologue encodes 1812 amino acids, in agreement with previous reports (73–76). A number of short gaps, including one of 14 amino acids corresponding to human residues 1048–1061, account for the 51 and 66 amino acid difference in length between mouse BRCA1 and the human and canine homologues, respectively (Fig. 1).
Canine BRCA1 contains a 5 amino acid duplication corresponding to human residues 1753–7 as well as a seven amino acid extension beyond the human carboxy terminus. Allowing for length differences, the overall amino acid identity between human and canine BRCA1 is 73.8%, whereas amino acid similarity is 89.6%. Murine BRCA1 is more divergent from both the canine homologue (52.6% identity/78.5% similarity) and the human (55.9% identity/80.1% similarity) than human and canine are from each other, reflecting the more recent evolutionary divergence of humans and dogs (78).

**Cross-species comparison**

To gain insight into the structure-function relationship of BRCA1, a multiple alignment of the human, canine, and mouse amino acid sequence was generated using the Blossum 50 substitution matrix (95). Following alignment, a table of the mean raw structural similarity values at each position was created, with a minimum value of –5 and maximum possible value of 15. The overall mean similarity values at each position was created, with a minimum (95). Following alignment, a table of the mean raw structural sequence was generated using the Blosum 50 substitution matrix a multiple alignment of the human, canine, and mouse amino acid sequences.

To gain insight into the structure-function relationship of BRCA1, a multiple alignment of the human, canine, and mouse amino acid sequence was generated using the Blossum 50 substitution matrix (95). Following alignment, a table of the mean raw structural similarity values at each position was created, with a minimum value of –5 and maximum possible value of 15. The overall mean similarity values at each position was created, with a minimum (95). Following alignment, a table of the mean raw structural sequence was generated using the Blosum 50 substitution matrix a multiple alignment of the human, canine, and mouse amino acid sequences.

To gain insight into the structure-function relationship of BRCA1, a multiple alignment of the human, canine, and mouse amino acid sequence was generated using the Blossum 50 substitution matrix (95). Following alignment, a table of the mean raw structural similarity values at each position was created, with a minimum value of –5 and maximum possible value of 15. The overall mean similarity values at each position was created, with a minimum (95). Following alignment, a table of the mean raw structural sequence was generated using the Blosum 50 substitution matrix a multiple alignment of the human, canine, and mouse amino acid sequences.

The amino and carboxy termini of BRCA1 are highly conserved. A stretch of 125 residues (human 1671–1796) within the conserved carboxy terminus is >85% similar among all three species with 80% identity of residues 1673–1752. This high degree of conservation is maintained between human and canine BRCA1 throughout the remainder of the protein (85% identity for residues 1797–1863), whereas murine BRCA1 diverges somewhat (47% identity) but retains 74% similarity. Conservative substitutions within the C-terminal 200 residues maintain the biochemical features of this region across human, canine, and murine BRCA1. The dog and mouse proteins are both longer than human, with extensions of 7 amino acids. At the amino terminus, the first 200 residues are 90% identical/96% similar between human and dog and 71% identical/90% similar between human and mouse. Within this region, the 40 amino acid RING finger domain differs only at a single site, a conservative amino acid substitution in the canine gene (K55R).

The central portion of BRCA1, largely corresponding to exon 11, is fairly divergent retaining only 70% identity between the human and canine proteins and 53% identity between human and mouse. However, as seen in Figure 2, several short domains are conserved across all three species. One such region (human residues 1335–1420, >80% similarity across all three species) is highly acidic, containing 18 conserved acidic residues, relative to 3 conserved basic residues. For the protein as a whole, the distribution of acidic residues is higher in the carboxy terminal two-thirds of BRCA1 than in the amino terminal one third, reflected by a net excess of 11 conserved basic residues in the first 700 amino acids of the protein, compared to an excess of 32 conserved acidic residues in the remaining two thirds. Human BRCA1 exhibits the lowest predicted pl (5.14), canine an intermediate pl (5.28) and mouse BRCA1 is least acidic (pl = 5.37).

Disease-associated missense mutations and neutral polymorphic variants were compared to determine cross-species conservation of both specific amino acid residues and the locations of the alterations with respect to conserved domains of BRCA1 (Table 1) (5,10–13,15,18,24,27,28,80–82). Missense alterations were classified as disease-associated if they segregate with breast-ovarian cancer in multiple high-risk families, whereas polymorphisms were defined as variants which occur at a frequency >0.05 in control chromosomes, or which were identified in patients carrying a second, truncating mutation of BRCA1. All other missense alterations were classified as having unknown consequences for the purpose of this analysis. The mean structural similarity value of the individual positions of all missense alterations were determined, as were the values for the averaged 21, 51 and 101 residue windows at these positions.

The results are given in Table 1, and the locations of the three classes of missense alterations are indicated with respect to the graph of the 51 residue averages of the mean structural similarity values for human, canine and mouse BRCA1 in Figure 2b. Disease associated missense alterations occur in highly conserved residues within conserved domains. On the other hand, three of the seven neutral polymorphisms (43%) occur in conserved residues. This is consistent with random occurrence of polymorphisms throughout the gene, excepting a few constrained positions. It is interesting to note that the known polymorphisms, including the conserved polymorphic residues, reside in regions of the gene which do not appear to comprise conserved domains of BRCA1 (Table 1: mean structural values of the 21, 51 and 101 residue windows are not significantly above the overall mean structural similarity value for the whole gene, 4.98). The single exception is the 21 window mean structural value for the conserved polymorphism, L871P. Among the eighteen missense alterations with unknown consequences, seven occur in absolutely conserved amino acids, seven in residues that have conservative substitutions across species and four are in residues that are not evolutionarily conserved.

**DISCUSSION**

Sequence analysis of human BRCA1 revealed several possible motifs; the RING finger at amino acids 24–64, two basic motifs at 500–506 and 604–614, the granin site at amino acids 1214–1223, and the leucine zipper at amino acids 1209–1230. The functional importance of a sequence motif is most effectively proven by demonstrating that the gene product has the biochemical features predicted by the sequence [e.g. (79)]. Alternatively, conservation of the motif across species can provide independent support for functional hypotheses. One limitation of argument from conservation, however, is that some consensus sequences are not well-defined. Hence, lack of conservation may reflect the functional irrelevance of the human sequence, different functions in different species, and/or incompletely defined consensus sequences.

Among the possible BRCA1 motifs, the RING finger is the most highly conserved. The 40 amino acid RING finger motif (46) contains only a single amino acid substitution (K55R in the canine sequence) among all three species studied here. This supports the importance of the RING finger to the normal function of BRCA1. Of the two basic motifs previously described
Figure 1. Alignment of the human, canine and murine BRCA1 amino acid sequences. Amino acid identities and similarities were shaded using the Box Shade program (K. Hoffman and M. Barron: http://ulrec3.unil.ch/software/Box_form.html). Identities are highlighted in black. Similarities are highlighted in grey. HS-Homo sapiens; CF-Canis familiaris (GenBank Accession no: U50709); MM-Mus musculus.
one at 500–505 is identical in mouse and human, but differs in dog, and that at 604–614 is perfectly conserved. However, neither motif corresponds completely to the functionally well-defined nuclear localization signal (RTKA)KK(RQNTSG)K (79).

Recent biochemical characterization suggests that human BRCA1 shares functional features, as well as sequence similarity with the granin family of proteins (83). The granin consensus sequence (84) is completely conserved in human BRCA1 (1214–1223), however the canine and murine sequences differ at amino acid 1216. Whether BRCA1 functions as a granin in all three species remains to be determined. A putative leucine zipper motif in human BRCA1 completely overlaps the granin consensus sequence (residues 1209–1230). Although substitution of the second leucine in both canine and murine BRCA1 would not be sufficient to disrupt a leucine zipper (85), the absence of hydrophobic residues at the ‘a’ position within the heptad repeat (abcdefg)4 (86) indicates that the periodicity of leucine residues at this location may not reflect a functional leucine zipper.

One striking feature of all three gene products is the degree of conservation at the carboxy-terminal domain. The region from 1671 to 1796 is more than 85% similar among the three species. However, there are no significant homologies to other gene products or functional motifs. The remaining portion of the protein (human amino acids 1797–1863) is 85% identical between human and dog, and 74% similar between human and mouse. Both dog and mouse sequences contain C-terminal extensions of seven amino acids. All three species conserve the tyrosine at amino acid 1853, which is the site of a truncation mutation in a human family that leads to very early onset breast and ovarian cancer (13). This tyrosine may be important for protein stability, or it may represent a functionally important phosphorylation site.

A small number of mouse and canine clones had insertions at a region corresponding to the junction between human exons two and three. As far as we know, insertions have not been observed at this site in human cDNA clones. In dogs, the inserted sequence is found in testes cDNA and is similar to a 39 bp open reading frame and three. As far as we know, insertions have not been observed at a region corresponding to human nucleotides 216–1210, Where BRCA1 functions as a granin in all three species remains to be determined. A putative leucine zipper motif in human BRCA1 completely overlaps the granin consensus sequence (residues 1209–1230). Although substitution of the second leucine in both canine and murine BRCA1 would not be sufficient to disrupt a leucine zipper (85), the absence of hydrophobic residues at the ‘a’ position within the heptad repeat (abcdefg)4 (86) indicates that the periodicity of leucine residues at this location may not reflect a functional leucine zipper.

One striking feature of all three gene products is the degree of conservation at the carboxy-terminal domain. The region from 1671 to 1796 is more than 85% similar among the three species. However, there are no significant homologies to other gene products or functional motifs. The remaining portion of the protein (human amino acids 1797–1863) is 85% identical between human and dog, and 74% similar between human and mouse. Both dog and mouse sequences contain C-terminal extensions of seven amino acids. All three species conserve the tyrosine at amino acid 1853, which is the site of a truncation mutation in a human family that leads to very early onset breast and ovarian cancer (13). This tyrosine may be important for protein stability, or it may represent a functionally important phosphorylation site.

A small number of mouse and canine clones had insertions at a region corresponding to the junction between human exons two and three. As far as we know, insertions have not been observed at this site in human cDNA clones. In dogs, the inserted sequence is found in testes cDNA and is similar to a 39 bp open reading frame within human intron two (77). The insertion would disrupt loop 1 of the conserved RING finger motif (87) thus altering its function. Testes express high levels of BRCA1 transcript (5,74), however there is no evidence of a phenotypic effect in this tissue among mutation carriers (9). Alternative splicing has been observed for several genes in testicular tissue, conferring tissue-dependent biological function in several instances (88–91), raising the possibility that the observed splicing variants may be of biological significance.

Finally, the location of human BRCA1 missense mutations may ultimately prove useful in suggesting which are disease-associated. The two unquestionably disease-associated mutations in the RING finger are in a completely conserved domain. Of the seven amino acid polymorphisms, two are at the most divergent locale in the sequence; the others are in regions of average or less than average cross-species similarity, excepting L871P. This polymorphism is more prevalent in breast/ovarian cancer patients than in unaffected controls (28) and occurs in a conserved residue at the end of an acidic stretch of 30 amino acids with 90% similarity across species. Of the eighteen missense mutations with unknown consequences, seven occur in amino acids that are identical across species. Four of these seven (E1219D, A1708E, P1749R and M1775R) are also within conserved domains. Taken together, these data predict regions of the gene which may be critical for normal function.

Assessing the functional relevance of these regions may be more effectively accomplished in appropriate animal models. Canine and human breast cancer share a number of critical epidemiologic and histologic features and substantial evidence exists that inherited factors may be important as well (71,72). Cloning of the canine BRCA1 gene thus permits testing the significance of this gene in pedigrees and breeds of dogs with a high incidence of breast cancer. The mouse genome is readily manipulated, enabling construction of strains harboring specific mutations in the murine BRCA1 homologue. Once established, animal models will provide a means to directly study disease etiology, as well as to develop treatment protocols.

**MATERIALS AND METHODS**

**Canine cDNA preparation**

Total RNA was isolated from 100 mg of beagle testis homogenized on ice in 1 ml of RNAsStat-60 (Tel-TestB Inc.). Organic extraction and RNA precipitation were carried out according to manufacturer’s specifications. cDNA was prepared from poly A+ RNA using the SuperScript™ Preamplification System for first strand cDNA synthesis kit (GibcoBRL Life Technologies; Gaithersburg, MD) with both random hexamers and oligo (dT) primers.

**Cloning the canine BRCA1 gene**

Different regions of the canine BRCA1 gene were cloned from PCR amplified DNA using either first strand cDNA or genomic DNA as template. PCR conditions were as described previously (92) using 30 cycles and annealing temperatures of 50–58°C. The 3’ end was the exception and was cloned by Rapid Amplification of cDNA Ends (RACE) using standard protocols (93). All PCR products were cloned into either the Invitrogen pCR II vector, using the Original TA Cloning Kit (Invitrogen; San Diego), or into the Promega pGEM™-T vector (Madison), using the pGEM-T Vector Systems kit. Clones were transformed into either INV°F™ cells (Invitrogen) or XL1Blue cells (Stratagene; La Jolla). Single colonies were selected and processed using prime manufacturer’s suggested protocols.

Primers derived from human BRCA1 cDNA sequence (HSU14680) were used to amplify and clone fragments corresponding to HSU14680 nucleotides 421–527, 1971–2485, 2510–2978 and 2738–3948 from canine genomic DNA and nucleotides 3707–4391, 3707–4886 and 4348–5119 from canine cDNA. A combination of human and canine cDNA derived primers were used to obtain subclones corresponding to nucleotides 464–2197 and 5071–5666 from canine cDNA. Nucleotides 120–477 were cloned from cDNA using a canine primer and a primer with wobble between the human and mouse sequences. The 3’-end (nucleotides 5527–6010) was cloned using a canine primer and an adapter primer to the RACE dT-adapter oligo. Subclones spanning nucleotides 1194–4130, 2161–4130, 2161–2760 and 4833–5054 were obtained from genomic DNA using primers derived entirely from the canine sequence. To verify the specificity of the canine sequence, PCR primers derived from canine sequence were used to amplify canine cDNA for the regions corresponding to human nucleotides 216–1210,
Table 1. Missense alterations in BRCA1

<table>
<thead>
<tr>
<th>Codon</th>
<th>Human BRCA1</th>
<th>Normal</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>396</td>
<td>Cys</td>
<td>Gly</td>
<td>13</td>
</tr>
<tr>
<td>397</td>
<td>Cys</td>
<td>Gly</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Canine BRCA1</th>
<th>Murine BRCA1</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>396</td>
<td>Cys</td>
<td>Gly</td>
<td>13</td>
</tr>
<tr>
<td>397</td>
<td>Cys</td>
<td>Gly</td>
<td>12</td>
</tr>
</tbody>
</table>

**Neutral Polymorphisms**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Human BRCA1</th>
<th>Normal</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>355</td>
<td>Gin</td>
<td>Arg</td>
<td>5</td>
</tr>
<tr>
<td>871</td>
<td>Lut</td>
<td>Pro</td>
<td>5</td>
</tr>
<tr>
<td>1038</td>
<td>Gin</td>
<td>Gin</td>
<td>13</td>
</tr>
<tr>
<td>1040</td>
<td>Ser</td>
<td>Asn</td>
<td>12, 13</td>
</tr>
<tr>
<td>1183</td>
<td>Lys</td>
<td>Arg</td>
<td>5</td>
</tr>
<tr>
<td>1613</td>
<td>Ser</td>
<td>Ser</td>
<td>5</td>
</tr>
<tr>
<td>1628</td>
<td>Met</td>
<td>Val</td>
<td>28, 81</td>
</tr>
</tbody>
</table>

**Mutations with Unknown Consequences**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Human BRCA1</th>
<th>Normal</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>379</td>
<td>Ile</td>
<td>Met</td>
<td>82</td>
</tr>
<tr>
<td>693</td>
<td>Asp</td>
<td>Asp</td>
<td>82</td>
</tr>
<tr>
<td>772</td>
<td>Val</td>
<td>Ala</td>
<td>82</td>
</tr>
<tr>
<td>826</td>
<td>Thr</td>
<td>Lys</td>
<td>11</td>
</tr>
<tr>
<td>841</td>
<td>Arg</td>
<td>Thr</td>
<td>28</td>
</tr>
<tr>
<td>1008</td>
<td>Met</td>
<td>Ser</td>
<td>28</td>
</tr>
<tr>
<td>1143</td>
<td>Arg</td>
<td>Met</td>
<td>28</td>
</tr>
<tr>
<td>1150</td>
<td>Ser</td>
<td>Ser</td>
<td>28</td>
</tr>
<tr>
<td>1219</td>
<td>Gln</td>
<td>Asp</td>
<td>28</td>
</tr>
<tr>
<td>1443</td>
<td>Arg</td>
<td>Met</td>
<td>28</td>
</tr>
<tr>
<td>1565</td>
<td>Thr</td>
<td>Ile</td>
<td>82</td>
</tr>
<tr>
<td>1637</td>
<td>Pro</td>
<td>Leu</td>
<td>12</td>
</tr>
<tr>
<td>1708</td>
<td>Gln</td>
<td>Arg</td>
<td>28</td>
</tr>
<tr>
<td>1713</td>
<td>Val</td>
<td>Gln</td>
<td>28</td>
</tr>
<tr>
<td>1775</td>
<td>Met</td>
<td>Met</td>
<td>5, 10</td>
</tr>
<tr>
<td>1852</td>
<td>Thr</td>
<td>Thr</td>
<td>82</td>
</tr>
</tbody>
</table>

**Figure 2.** Structural similarity of BRCA1 protein in humans, dogs and mice. (A) Graphical representation of mean structural similarity values from the three pairwise comparisons between the human, canine and murine BRCA1 protein sequences averaged over a 21 residue window (thin black line) and a 101 residue window (thick grey line). The overall mean structural similarity value for the whole gene is 3.91. (B) Three classes of missense mutations are graphed with respect to the mean structural similarity values averaged over a 51 residue window. Disease associated mutations are shown as an X, neutral polymorphisms as open circles and missense mutations with unknown consequences as filled circles. The BRCA1 protein coding structure, drawn to scale is shown below the graph, with the locations of the RING finger motif, the putative nuclear localization signals, the granin motif, and the putative leucine zipper indicated. Conserved dibasic cleavage sites at which proteolytic processing of BRCA1 may occur are indicated by filled triangles.

3966–5054, and 5527 to the RACE adapter primer in the untranslated region.

**Murine BRCA1 cDNA isolation**

At the time these experiments were initiated, the mouse BRCA1 gene had not yet been cloned. We screened $1 \times 10^6$ independent clones from an 129/Sv strain embryonic stem cell cDNA library, kindly provided by Philippe Soriano (Division of Molecular Medicine, Fred Hutchinson Cancer Research Center, Seattle, WA), with BRCA1-specific PCR products amplified from human lymphoblast cDNA, using primers and conditions described in (13), or newly isolated cDNA inserts radioactively labelled with the Multiprime labelling system (Amersham). Hybridization conditions, clone isolation and PCR amplification of the cDNA insert have been described previously (21).

Ten BRCA1 clones were isolated: three derived from the 5′ end (corresponding to human BRCA1 cDNA nucleotides 1–1462) and seven from the 3′-end (HSU14680 nucleotides 4474–5711). However, no clones were identified that span the intervening region using either human BRCA1 PCR fragments or the mouse BRCA1 cDNA clones as probes under reduced stringency hybridization ($6 \times$ SSC, 1% SDS, 5% dextran sulfate) and wash conditions ($1 \times$ SSC, 0.1% SDS) at 50°C. This central portion of BRCA1 was thus isolated by PCR amplification from either Balb/C genomic DNA (human nucleotides 742–1985, 1861–2930, 2833–3480) or Balb/C testes cDNA (HSU14680 nt 3102–4532) using relaxed annealing temperatures and long ramp
times with a combination of human \textit{BRCA1} primers and murine primers derived from the newly sequenced cDNA clones.

\textbf{DNA sequencing and analysis}

Canine cDNA and genomic subclones were sequenced using either the Taq DyeDeoxy terminator cycle sequencing kit or the Taq DyePrimer cycle sequencing kit (Applied Biosystems). Samples were analyzed using an ABI 373A sequencer. All Taq Dye Terminator sequencing was done using selected canine primers (17–22-mers) that spanned the gene at approximately 350 bp intervals. Clones were sequenced bidirectionally, with the exception of portions of those clones developed for independent sequence verification only.

Murine cDNA inserts were PCR amplified using vector specific primers (94) and bidirectional sequence obtained manually (USB PCR product sequencing kit) or on the ABI 377 (Perkin Elmer ABI PRISM dye terminator cycle sequencing kit) with either vector specific or sequence derived primers. PCR fragments amplified from Balb/C testes cDNA or genomic DNA were sequenced directly with the amplifying primers and primers derived from the newly obtained sequence.

Sequence assembly of the canine and murine \textit{BRCA1} genes was done using the DNA* (DNA* Inc., Madison, WI) and MacVector 5.0 (Eastman Kodak Co.) sequence analysis programs, respectively. Sequences were aligned by the DNA* alignment program using default penalties for gap initiation (10) and gap extension (10). Minor adjustments were made to optimize the generated amino acid alignments. Analyses of the alignments were performed using the Blosum 50 amino acid substitution matrix (95).

\section*{ACKNOWLEDGEMENTS}

We thank the following individuals for their contributions to the canine \textit{BRCA1} project: Rainer Storb and Amelia A. Langston for helpful comments throughout the experiments, Ed Giniger for assistance with sequence analysis, and Jennifer Thompson for helpful discussions. This work was supported by the National Institutes of Health Grant RO1-CA27632 (M.-C. King), American Cancer Society Grant VM-165 (E.A. Ostrander), NCI Postdoctoral Fellowship F32-CA66293 (C. Szabo), and a grant from the Life & Health Insurance Medical Research Fund (J.C. Roach). M.-C. King is an American Cancer Society Research Professor and E.A.Ostrander is the recipient of an American Cancer Society Junior Faculty Award, JFRA-558.

\section*{REFERENCES}


