Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 86 (2005) 316 – 328

GENOMICS

www.elsevier.com/locate/ygeno

Identification of two evolutionarily conserved and functional regulatory elements in intron 2 of the human *BRCA1* gene

Stacey L. Wardrop[a], kConFab Investigators[b], Melissa A. Brown[a,*]

[a]*School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Australia*
[b]*Peter MacCallum Cancer Centre, Melbourne, Australia*

Received 15 March 2005; accepted 25 May 2005
Available online 28 June 2005

Abstract

Cross-species comparative genomics is a powerful strategy for identifying functional regulatory elements within noncoding DNA. In this paper, comparative analysis of human and mouse intronic sequences in the breast cancer susceptibility gene (*BRCA1*) revealed two evolutionarily conserved noncoding sequences (CNS) in intron 2, 5 kb downstream of the core *BRCA1* promoter. The functionality of these elements was examined using homologous-recombination-based mutagenesis of reporter gene-tagged cosmids incorporating these regions and flanking sequences from the *BRCA1* locus. This showed that CNS-1 and CNS-2 have differential transcriptional regulatory activity in epithelial cell lines. Mutation of CNS-1 significantly reduced reporter gene expression to 30% of control levels. Conversely mutation of CNS-2 increased expression to 200% of control levels. Regulation is at the level of transcription and shows promoter specificity. Both elements also specifically bind nuclear proteins in vitro. These studies demonstrate that the combination of comparative genomics and functional analysis is a successful strategy to identify novel regulatory elements and provide the first direct evidence that conserved noncoding sequences in *BRCA1* regulate gene expression.

Keywords: *BRCA1* gene; Intron; Transcriptional regulatory element; Comparative genomics; Genetic recombination

Enhancer and repressor elements housed in noncoding regions of DNA are known to play an important role in the regulation of many genes [1,2], including tumor suppressor genes and proto-oncogenes [3–6]. Disruption of such sequences or the factors that bind to them can contribute to diseases, including cancer [7]. Identification of critical enhancer and repressor elements has been a challenge, largely due to the size and uncharted nature of noncoding sequences housing these elements and the limitations of available experimental approaches. The recent completion of several eukaryotic genome sequencing projects now provides an opportunity to use comparative genomics to study homologous genes in multiple species. Comparative sequence analysis has been widely used to identify genes and coding boundaries, but recent interest has focused on utilizing this strategy to identify potential regulatory elements in the noncoding fraction of the genome. Indeed, a number of recent studies describe multiple highly conserved noncoding sequences (CNS) in the human genome and suggest that these sequences could represent functional regulatory elements [8–10]. A growing body of experimental evidence supports this hypothesis [11–14]. In this study we have applied both comparative genomics and functional analysis to identify and characterize conserved intronic elements in the human *BRCA1* gene.

BRCA1 is a breast cancer susceptibility gene that is implicated in both familial and sporadic breast cancers ([15], reviewed in [16,17]). It encodes a nuclear phosphoprotein involved in the regulation of DNA repair, transcription, cell-cycle control, ubiquitination, and apoptosis. The *BRCA1*

* Corresponding author. Fax: +61 7 3365 4699.
E-mail address: melissa.brown@uq.edu.au (M.A. Brown).
URL: www.kconfab.org.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.05.006

gene was originally thought to account for approximately 45% of hereditary breast cancers [18,19]; however, mutations affecting the coding region or the promoter have been detected in less than 20% of familial cases [20]. In sporadic breast cancer, somatic mutations of *BRCA1* are rarely found; however, reduced BRCA1 protein levels and mRNA expression are often observed [21–23], suggesting that *BRCA1* repression rather than mutation is involved. Taken together, these data support the hypothesis that transcriptional and/or posttranscriptional downregulation of *BRCA1* may contribute to the genesis of some breast cancers.

The molecular mechanisms controlling *BRCA1* expression are only partly understood. The proximal promoter of *BRCA1* is bidirectional, separated by only 218 nucleotides from an unrelated gene called *NBR2* [24]. There is some evidence to suggest that *BRCA1* and *NBR2* are reciprocally regulated [25,26], thus activation of the *NBR2* promoter could lead to decreased *BRCA1* expression. Several sequence elements that can regulate the *BRCA1* promoter have been identified. These include a CREB/ATF-1 site that appears to be a strong positive transcriptional element [27,28], a RIBS element that binds the multisubunit *ets* protein GA-binding protein-α/β and acts as an activator in MCF7 cells [29], and an E2F site that may mediate estrogen-dependent *BRCA1* expression [30]. One mechanism of transcriptional repression of *BRCA1* is promoter hypermethylation. Disruption of histone acetylation and hypermethylation via aberrant methylation of CpG islands significantly reduces *BRCA1* mRNA expression [31–34]; however, this appears to be significant in only 10–20% of breast tumors.

To understand fully the normal and abnormal expression of the *BRCA1* gene, it is critical to include noncoding regions in the search for novel regulatory regions. Interestingly, only one regulatory element outside of the *BRCA1* promoter region has been described to date. Suen and Goss used deletion studies to localize a 36-bp repressor element in intron 1 of *BRCA1* [35]. This element is not conserved between humans and mice. The other noncoding regions that could potentially regulate *BRCA1* expression are large and poorly characterized. The *BRCA1* gene comprises 24 exons and spans over 80 kb of genomic DNA on chromosome 17q21 [36,37]. *BRCA1* intronic sequences vary in length from 0.4 to 9.2 kb and total 73 kb, equating to over 90% of the *BRCA1* gene [37]. Using a cross-species comparative sequence analysis, we have identified several regions of strong evolutionary conservation in the *BRCA1* 3′UTR. Gene reporter experiments and protein binding assays further indicate that these regions of the *BRCA1* 3′UTR are functionally important (J.M. Saunus et al., submitted for publication).

In this study we have used a similar bioinformatic approach to identify conserved regions between human and mouse *BRCA1* intronic sequences. Interestingly, we found two regions in intron 2 with striking homology to sequences in mouse *Brca1* intron 2. The first conserved noncoding sequence (CNS-1) is 105 bp and shows 81% identity, the second sequence (CNS-2) is 195 bp and has 91% identity to the mouse sequence. To examine the abilities of the two candidate regulatory elements to regulate *BRCA1* gene expression, we tagged a cosmid with reporter genes and then examined the effect of mutating these regions on reporter gene expression in HeLa and SVCT cell-lines. Our results show that mutation of CNS-1 significantly reduced reporter gene expression to 30% of the control, whereas mutation of CNS-2 increased expression to 200%. Regulation by both regions is at the level of transcription and shows promoter specificity. Further analysis revealed that part of CNS-2 is a rarely used alternative splice variant of the human *BRCA1* gene and that it is located in a region subject to recombination-mediated deletion.

Results

Intronic sequence conservation between human and mouse BRCA1 identifies nine potential regulatory regions

As an initial approach to identifying novel sequence elements involved in regulation of *BRCA1* expression, we performed a comparative analysis of the human and mouse *BRCA1* genomic sequences. Based on the observation that these elements are usually composed of long sequences (≥100 bp in length) and are highly conserved (≥70% identity) [14], nine CNS matching these criteria were identified in six different introns of the *BRCA1* gene (unpublished data). A BLAST search of the conserved sequences showed that three of these elements were also present in the human *BRCA1* pseudogene 1 (*BRCA1P1*) [36], but showed no significant homology to any other regions in the human genome. To determine whether these conserved regions contain functional *BRCA1* regulatory elements, we initially focused on two elements located in intron 2 (Fig. 1, CNS-1, CNS-2), because they were present on a well-characterized and reporter-tagged cosmid clone (Fig. 3A), thus enabling immediate analysis. Although the remaining intronic CNS and sequences flanking the *BRCA1* gene are also likely to be important in the regulation of *BRCA1* expression, functional analysis of these sequences was beyond the scope of the present study.

BRCA1 intron 2 is approximately 8.2 kb in humans and contains a high density (~57%) of interspersed repetitive elements (Fig. 1). CNS-1 and CNS-2 are located in the middle of the intron, in the longest repeat-free region, approximately 5 kb downstream of the core *BRCA1* promoter (Fig. 1).

Intron 2 of BRCA1 contains evolutionarily conserved subsequences and putative transcription factor binding sites

Evolutionary conservation of noncoding sequences between relatively closely related species can be indicative

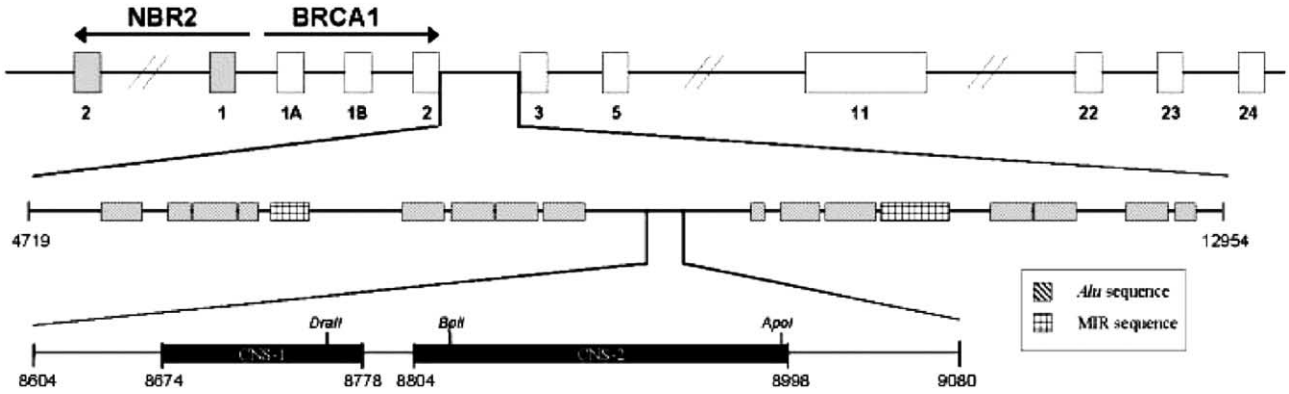

Fig. 1. Identification of intron 2 conservation between human and mouse *BRCA1*. Schematic diagram of the human *BRCA1* gene region showing the relative positions of repetitive elements and two conserved noncoding sequences (CNS-1 and CNS-2) within the *BRCA1* intron 2 region. Nucleotide positions correspond to GenBank Accession No. L78833 [37].

of functional regulatory elements [10]; therefore we extended our evolutionary sequence comparison using intron 2 of *BRCA1* homologues (Fig. 2). The *BRCA1* intron 2 genomic sequences for human, chimpanzee, gorilla, orangutan, rhesus macaque, dog, mouse, rat, and chicken were obtained from the NCBI (http://www.ncbi.nlm.nih.gov) or UCSC (http://genome.ucsc.edu/) sequence databases. Cross-species comparative analysis revealed that CNS-1 and CNS-2 are highly conserved among the mammalian species studied (Fig. 2), showing sequence identities of 65 and 75%, respectively. This degree of conservation is significantly higher than that of surrounding sequences, showing a 20–30% identity, which is likely to be random and thus insignificant. The alignment also revealed that there was generally more conservation between the higher order mammals and that the chicken sequence was not homologous to any of the mammalian *BRCA1* intron 2 sequences (Fig. 2).

Further analysis of the DNA sequence of both regions using the TRANSFAC [38] and rVISTA [39] transcription factor databases indicates the presence of many potential binding sites for known transcription factors (TF). In CNS–2, these include sites for members of the activator protein (AP); hepatocyte nuclear factor (HNF); Forkhead box-containing protein, O subfamily (FoxO); and nuclear factor erythroid 2 (NFE-2) families. CNS-1 contained one potential HNF-1 binding site. Interestingly, putative binding sites for all four above-mentioned TF were clustered around nt 8924–8950 in CNS-2 (Fig. 2), a region that had a high evolutionary conservation. Functionally relevant regions, such as promoters and enhancers, often have clustered and evolutionarily conserved TF binding sites [40], supporting a functional role for the sequences identified in the study.

Identification of two cis-acting functional regulatory elements in intron 2 of the BRCA1 gene

To determine the functional significance of CNS-1 and CNS-2, we chose to examine the effect of disrupting these regions in the context of the *BRCA1* promoter and surrounding sequences. Such a genetic approach has proved to be a very powerful and sensitive means of identifying critical noncoding regulatory elements in several other genes, the activity of which is not always detectable by cloning them next to heterologous promoters and away from their normal genomic environment (reviewed in [1]). A homologous recombination approach [41] was used first to tag *BRCA1* genomic clones with the reporter gene luciferase or Venus (modified yellow fluorescent protein YFP [42]), followed by targeted replacement of the conserved regions with a tetracycline resistance cassette (TetR; Fig. 3A). Insertion of the luciferase or Venus coding sequence into exon 2 of cosA11100 places reporter gene expression under the normal transcriptional regulatory elements of the *BRCA1* locus (Fig. 3A). This was confirmed by transient transfection of HeLa cells with the parental constructs followed by RT-PCR with vector-specific primers. The results showed that two transcripts, which contained either the *BRCA1* 1A or the *BRCA1* 1B 5′UTR, as previously described [43], *BRCA1* exon 2, and the reporter cDNA in frame, were produced (unpublished data). 3′RACE was also used to confirm that the 3′ end of each mRNA ended after the polyadenylation signal of the reporter gene (unpublished data). Together these results confirm that the cosA11100 parental constructs contain all the elements necessary to initiate synthesis of *BRCA1* mRNA.

Constructs were transiently transfected into HeLa and SVCT cells and reporter gene expression was analyzed. In the case of luciferase-tagged constructs, luciferase activity was measured relative to a cotransfected *Renilla* control (Fig. 3B). For comparison of *BRCA1* expression between parental and mutated constructs, it was essential to include appropriate control constructs. In the positive control construct, designated Δα-promoter, the *BRCA1* proximal and major promoter was replaced with TetR. This resulted in decreased luciferase activity to 30% of parental levels in both cell lines (Fig. 3B). These results indicate that the α

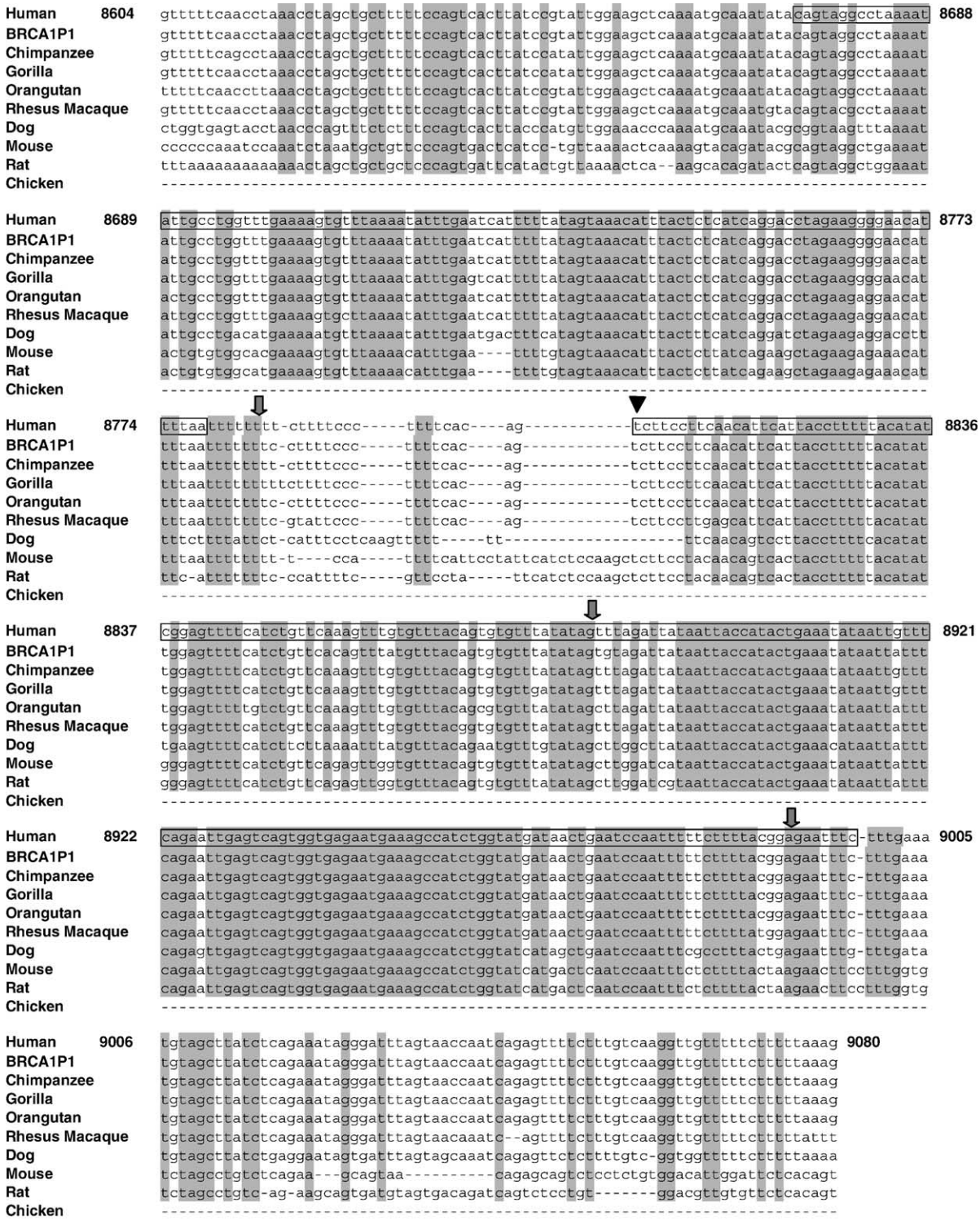

Fig. 2. Two CNS in intron 2 of the *BRCA1* gene are conserved in multiple species. VISTA multiple sequence alignment comparing nt 8604–9080 of the human *BRCA1* intron 2 with homologous sequences from the human *BRCA1* pseudogene 1 (*BRCA1P1*), chimpanzee, gorilla, orangutan, rhesus macaque, dog, mouse, rat, and chicken. Conserved nucleotides between five of the six species are shaded. CNS-1 (nt 8674–8778) and CNS-2 (nt 8804–8998) are boxed. Arrowheads denote the splice donor site and arrows show genomic breakpoints.

promoter is stronger than the β promoter in this system, contributing to 70% of the assayed luciferase activity, consistent with previous findings [43]. Two negative controls, the 5′Tet and 3′Tet, had the TetR gene inserted into nonconserved regions at the 5′ and 3′ ends of the reporter constructs, respectively. As shown in Fig. 3B,

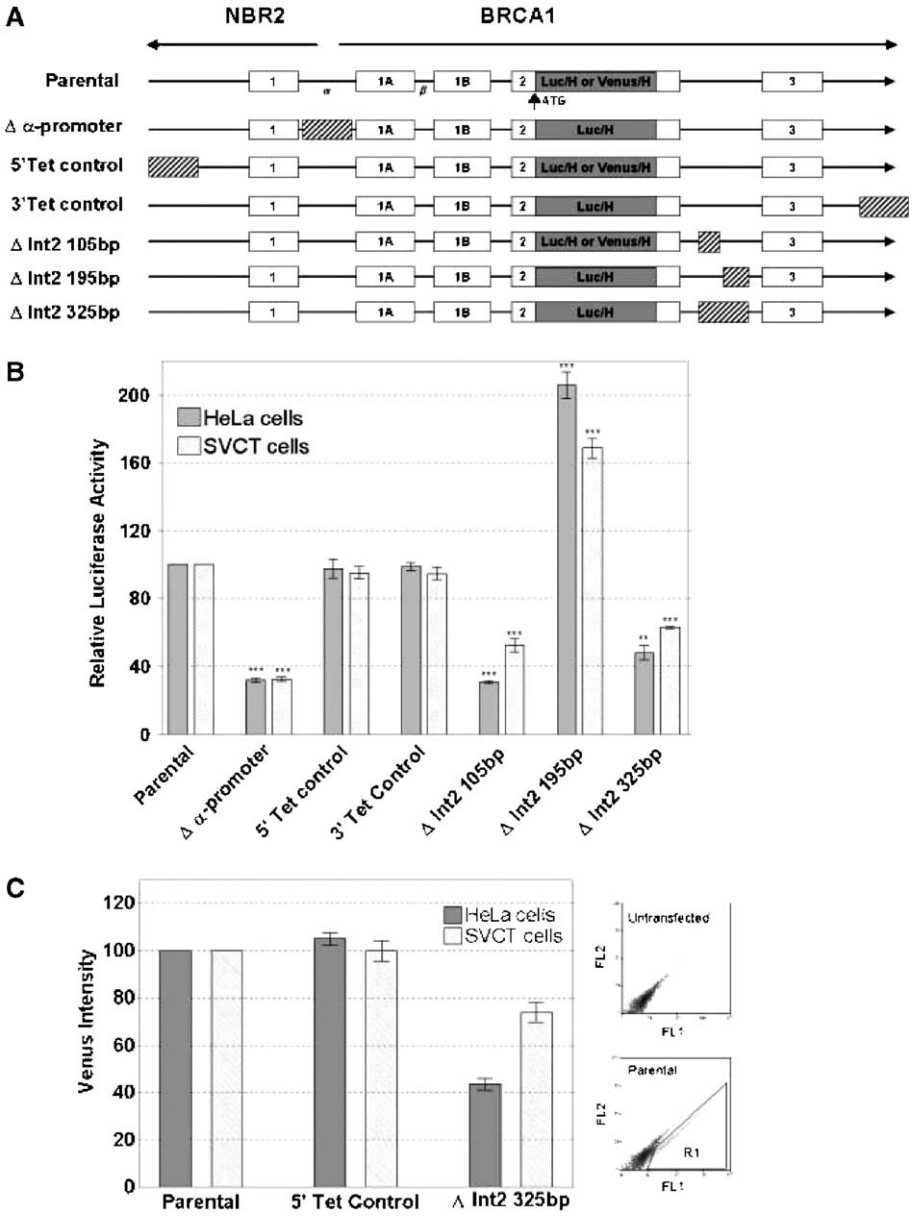

Fig. 3. Disruption of CNS-1 and/or CNS-2 in *BRCA1* intron 2 alters reporter gene activity. (A) Schematic representation of the luciferase and Venus cosmid reporter vectors constructed using two-stage GET recombination (refer to Materials and methods). Exons are indicated by white boxes and numbered, α and β symbols denote *BRCA1* alternative promoters. Gray boxes indicate the insertion of the luciferase (Luc/H) or Venus (Venus/H) reporter gene in-frame into *BRCA1* exon 2, striped boxes indicate insertion points of the TetR gene. (B) Luciferase reporter activities relative to the *Renilla* internal control ± the standard error of the mean (SEM) are shown as percentages of the parental vector control. Data were generated from at least three independent experiments. $^{**}p < 0.001$ and $^{***}p < 0.0001$. (C) Right: Representative FACS analysis of transiently transfected cells sorted for YFP expression (FL1, *x* axis) and orange autofluorescence (FL2, *y* axis), cells in R1 represent Venus YFP-expressing cells. Left: Venus YFP intensity ± SEM is shown as a percentage of the parental vector control. Data were generated from three independent transfections.

these constructs did not significantly change luciferase expression, indicating that sequence length or conformational changes due to insertion of the TetR gene do not contribute to *BRCA1* transcriptional regulation in this system.

Mutation of CNS-1 resulted in a significant decrease in luciferase activity compared to the parental construct (two- to threefold), raising the possibility that positive regulatory elements map to this region (Fig. 3B). This decrease is comparable to mutation of the α promoter alone, suggesting that this region is critical. Interestingly, mutation of CNS-2 resulted in a twofold increase in luciferase activity, suggesting that negative regulatory elements exist in this region of intron 2 (Fig. 3B). Finally, mutation of the entire 325-bp region (CNS-1 and CNS-2 together) decreased luciferase activity (Fig. 3B), but not quite to the same

extent as mutating CNS-1 alone. These results indicate a possible interaction between elements, with the positive regulatory activity of CNS-1 being dominant in the context of this reporter assay.

To complement the luciferase assays, the same constructs were analyzed using a Venus reporter instead of a luciferase reporter. Transiently transfected cells were subjected to flow cytometry analysis (Fig. 3C). Venus intensity levels were determined in individual cells (20,000 cells/construct) and untransfected cells were discarded from calculations, thus removing any bias attributable to transfection efficiencies. Transient transfection of cell lines with the 5′ Tet control construct showed no significant change in Venus intensity levels compared to the parental control; however, the ΔInt2 325-bp (CNS-1 + 2) construct decreased intensity to levels equivalent to relative luciferase activities (Figs. 3B and 3C). These results confirm that the change in the reporter activities observed between constructs is not attributable to the reporter system used.

CNS-1 and CNS-2 function as transcriptional regulators of the BRCA1 promoter in vitro

In the context of the *BRCA1* reporter constructs, CNS-1 and CNS-2 are located outside of the transcribed mRNA, suggesting that the observed changes in luciferase activity are most likely due to changes in transcription, rather than mRNA stability or translational efficiency. To investigate the proposed transcriptional activity of regions CNS-1, CNS-2, and combined 325-bp fragment (Fig. 4A), these sequences were cloned downstream of the luciferase reporter gene in pGL3-B and pGL3-C vectors, the former containing the *BRCA1* α promoter and the latter the SV40 promoter. The human *BRCA1* α promoter was used to determine the abilities of the intron elements to regulate the transcriptional activity of their own promoter. The SV40 promoter was used to address the ability of these fragments to regulate a heterologous promoter. Luciferase activity and mRNA

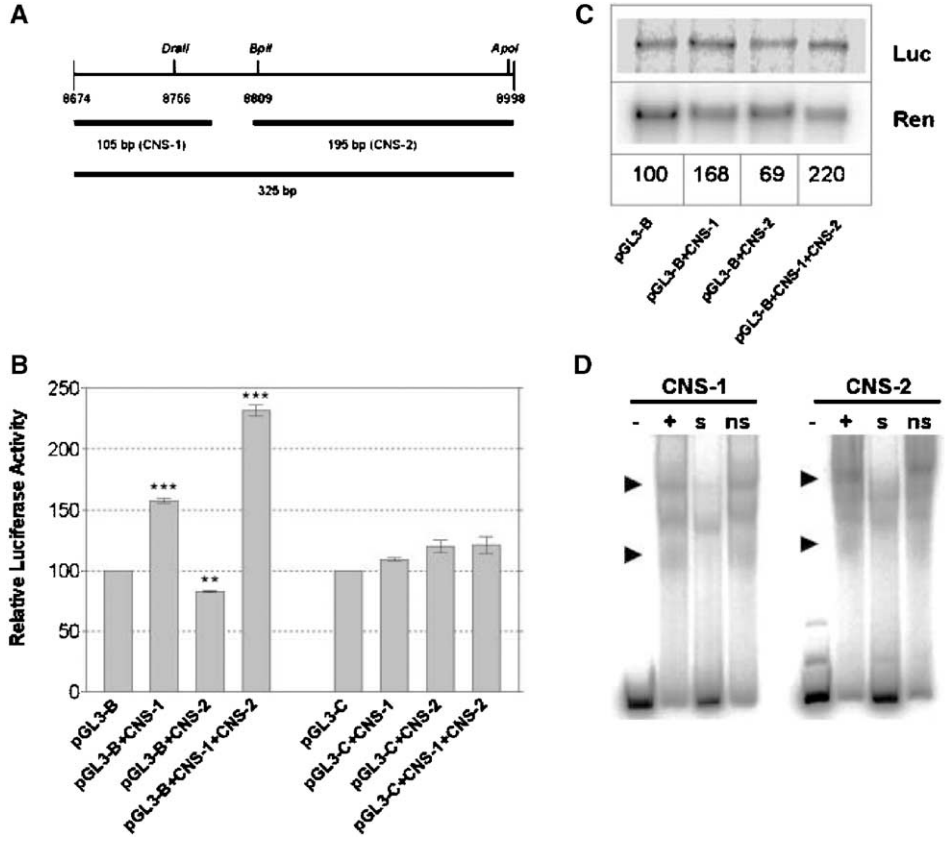

Fig. 4. CNS-1 and CNS-2 demonstrate differential transcriptional regulation of the *BRCA1* promoter. (A) Schematic representation of the human *BRCA1* intron 2 CNS-1 (105 bp) and CNS-2 (195 bp) and the combined 325-bp fragment used in pGL3 reporter gene experiments. Nucleotide positions correspond to GenBank Accession No. L78833 [37]. (B) Luciferase reporter activities relative to the *Renilla* internal control ± SEM are shown as percentages of the pGL3-B or pGL3-C vector controls, respectively. Data were generated from three independent transfections. $^{**}p < 0.001$ and $^{***}p < 0.0001$. (C) Top and middle: Representative blots of luciferase (Luc) and *Renilla* (Ren) mRNA levels as determined by RNase protection assays. Bottom: Luciferase mRNA levels relative to the *Renilla* internal controls are shown as percentages of the levels in cells transfected with the pGL3-B vector control. These data are typical of two independent experiments performed. (D) Representative EMSA with CNS-1 and CNS-2 fragments. Radiolabeled DNA probes were incubated with either no protein (−) or 10 μg of HeLa nuclear protein (+). Shifted DNA–protein complexes show a slower migration compared to the probe alone. A 100-fold excess of unlabeled self-fragment (s) or nonspecific DNA (ns) was added as competitor to determine the specificity of the DNA–protein complexes. Arrowheads indicate specific complexes.

levels, relative to a cotransfected *Renilla* control, were then determined in HeLa cells. Results from transient transfections of the pGL3-B constructs demonstrated that insertion of CNS-1 increased luciferase activity and mRNA levels by 1.6-fold compared to the vector control (Figs. 4B and 4C). Conversely, addition of CNS-2 decreased luciferase activity, with a concomitant decrease in luciferase mRNA level (Figs. 4B and 4C). Insertion of the 325-bp fragment increased luciferase activity and mRNA levels by 2.4-fold, suggesting that CNS-1 is the dominant activator. Interestingly, addition of the intronic fragments to the pGL3-C vector (SV40 promoter) had no significant effect on luciferase activity (Fig. 4B), demonstrating that transcriptional regulation has some promoter specificity.

Conserved enhancer or repressor elements situated in noncoding regions often interact with DNA binding proteins [11,44]. To determine whether this was true for CNS-1 and CNS-2, electrophoretic mobility shift assays (EMSAs) were performed using radiolabeled double-stranded DNA probes and HeLa nuclear extracts. As shown in Fig. 4D, several slowly migrating bands representing DNA–protein complexes were observed with both probes in the presence of HeLa nuclear extract. Two specific complexes were identified using either CNS-1 or CNS-2 DNA probe (Fig. 4D, arrowheads). These complexes were competed by the addition of a 100-fold excess of unlabeled self-competitor (s) but not by a nonspecific competitor (ns).

In summary, these data indicate that intron 2 of the *BRCA1* gene contains two evolutionarily conserved noncoding elements that can differentially regulate *BRCA1* mRNA in vitro. CNS-1 appears to be essential for promoter activation, while CNS-2 is a potential repressor element. Regulation by both regions is at the level of transcription and shows specificity for the *BRCA1* promoter. Both regions also contain a number of potential TF binding motifs and specifically bind nuclear proteins in vitro, although the identity of the proteins is yet to be determined.

BRCA1 intron 2 mutation analysis

Given that intronic variants in several genes have been associated with cancer [7,45,46] we sought to determine whether there was any sequence variation in CNS-1 or CNS-2 in the normal population or in familial breast cancer cases. The NCBI single nucleotide polymorphisms (SNPs) database was initially searched to locate any previously identified SNPs in *BRCA1* intron 2. A number of putative SNPs have been characterized in intron 2 but none were located in CNS-1 or CNS-2. Indeed this region of intron 2 is uncharacteristically lacking in any variation (see www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9606&QSTR=brca1&QUERY=uid(7811760)&CHR=17&MAPS=ugHs,genes,morbid,snp,pheno[38516465.26%3A38524540.63]-r&CMD=DN).

To determine whether *BRCA1* intron 2 mutations are present in familial breast cancer, 79 patients with a strong family history of breast cancer but no detectable *BRCA1* or *BRCA2* mutation were screened using DHPLC. Both CNS-1 and CNS-2 were analyzed; however, no sequence variations were observed (data not shown). This suggests that if mutations exist in this region of *BRCA1* intron 2 they are rare. This result also suggests that disease-associated disruption of intron 2-mediated regulation of *BRCA1* expression is more likely to involve altered expression

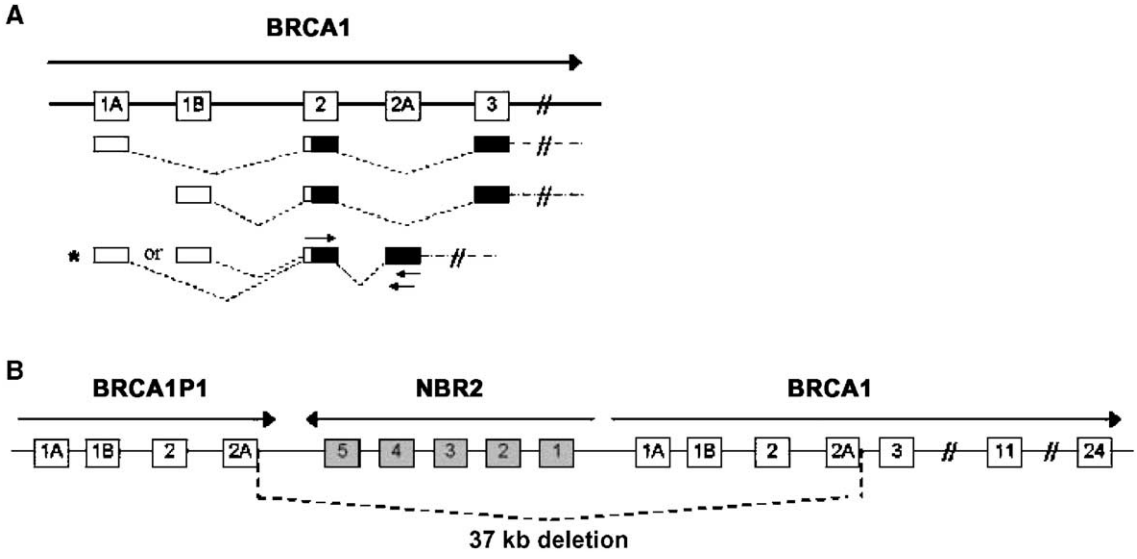

Fig. 5. CNS-2 is a novel *BRCA1* alternative splice variant and located in a genomic nonhomologous recombination hot spot. (A) Schematic representation of the two *BRCA1* transcripts generated by the use of alternative first exons [49] and the novel (*) *BRCA1* alternative splice variant identified in this study. White boxes represent untranslated mRNA sequences and black boxes indicate open reading frames. Arrows represent the direction of the primers used in nested RT-PCR (B) Genomic organization of the human *BRCA1* gene, human *BRCA1* pseudogene 1 (*BRCA1P1*), and its neighboring gene, *NBR2*. *BRCA1* and *BRCA1P1* exons and *NBR2* exons are indicated by white and gray boxes, respectively. The dotted line is representative of the 37-kb deletions previously identified in two separate studies [63,64].

and/or activity of *trans*-acting factors that associate with and act on this region of the genome.

Identification of a novel BRCA1 splice variant located in a region subject to recombination-mediated deletion

Because of the unusually high evolutionary conservation between human and mouse CNS-2 sequences (91%, Fig. 1), we hypothesized that this region may be an alternatively spliced exon of *BRCA1*. To test this hypothesis we performed nested RT-PCR using a *BRCA1* exon 2 primer and specific CNS-2 primers (see Materials and methods). Sequencing of the resulting PCR products revealed that the first 138 nt of CNS-2 are included as an alternative exon (designated exon 2A, Figs. 2 and 5A). It was of interest to note that the consensus splicing acceptor site was not conserved between humans and rodents, suggesting this variant may be present only in higher order mammals (Fig. 2). Numerous *BRCA1* splice variants have been reported in the literature (reviewed in [47]); however, their regulation and function are poorly understood. It was beyond the scope of this study to investigate a role for this variant; however, it is unlikely to produce a functional protein as inclusion of the new exon (exon 2a) introduces a premature termination codon (unpublished data). Also of interest is the mapping of several disease-associated gene deletion events to this region (e.g., Figs. 2 and 5B and see Discussion), suggesting that this region of the genome is also subject to recombination-mediated gene deletion.

Discussion

Comparative genomics offers a potential strategy for identifying important regulatory elements on the basis that evolutionary conservation implies functional constraints. In this study, we utilized human–mouse genomic comparisons to identify potential noncoding regulatory elements in the *BRCA1* gene. The evolutionary distance been humans and mice is reported to be sufficient for identification of functionally important elements [48]. This premise is supported by studies identifying functional sequences exclusively through the use of human–mouse comparisons [12,13,49].

In this paper we report the identification of two evolutionarily conserved regions (CNS-1 and CNS-2) in intron 2 of the breast cancer susceptibility gene, *BRCA1*. *BRCA1* expression is regulated during the cell cycle [50], in response to DNA-damaging agents [51] and by hormones that regulate proliferation and differentiation of the mammary gland, including progesterone [52] and prolactin [53]. The *cis*-elements that mediate regulation by these biological events are currently unknown and could include the new elements CNS-1 and CNS-2 described in this paper.

Analysis of CNS-1 and CNS-2 revealed that both regions are also conserved in other mammalian species (Fig. 2). Consistent with this, recent studies have demonstrated that noncoding sequences conserved between mammals are frequently functional [11,54,55]. Together, these findings suggest a functional role for CNS-1 and CNS-2 in regulating *BRCA1* gene expression. It was of interest to note that the chicken sequence was not homologous to any of the mammalian *BRCA1* intron 2 sequences. This was not unexpected, considering that human and chicken *BRCA1* cDNA sequences show relatively low identity (~30% [56]). Overall, chicken *BRCA1* intron lengths are much shorter than human introns (e.g., intron 2, 1.7 kb vs 8.2 kb), which could reflect differences in the regulatory mechanisms controlling this gene in different species.

To determine the functional significance of CNS-1 and CNS-2, we mutated reporter constructs generated using *BRCA1*-derived genomic clones. This approach allowed potential functional noncoding sequences to be analyzed in the context of the *BRCA1* genomic region and, in particular, in the context of *BRCA1's* own promoter. Mutation of either CNS-1 or CNS-2 or the entire region resulted in significant changes in reporter gene activity (Figs. 3B and 3C). Such changes in reporter activity (30–200%) are considered to be biologically significant. Functional noncoding elements identified for the human *APOE* and *TNF*-α genes, for example, confer changes in reporter activity in the range of 40–140%, relative to controls [57,58]. Our results indicate the presence of significant regulatory elements that mediate both activation and repression of the *BRCA1* gene. This is the first direct demonstration that *BRCA1* is regulated by evolutionarily conserved intronic elements.

Although CNS-1 and CNS-2 are over 5 kb downstream of the human *BRCA1* proximal promoter, our data indicate that these elements regulate *BRCA1* expression at the transcriptional level. Transcriptional control elements located in introns have been reported for various genes. For example, two AP-1 sites located in intron 1 of the breast cancer-specific gene 1 (*BGSG1*) are essential for transcriptional activation in breast carcinomas [59]. A conserved enhancer located in intron 3 of the *BCL3* proto-oncogene is also critical for transcriptional upregulation in T cells [4].

As transcription can be regulated by proteins that bind to sequence-specific transcription factor-binding sites (TFBS), a combination of TFBS predictions, sequence conservation, and cluster analysis was used to identify potential TFBS in CNS-1 and CNS-2. Significantly, a number of these putative sites (e.g., AP-1, NFE-2, and HNF-1) were clustered together in regions of high evolutionary conservation, supporting their functional importance. As mentioned previously, AP-1 has recently been implicated in transcriptional regulation of the *BGSG1* gene [59]. Furthermore, AP-1 and NFE-2 binding sites can both contribute to long-range α-globin gene activation [60]. Consistent with these

bioinformatic data, EMSAs show that both elements do specifically bind proteins in vitro; however, whether the candidate proteins involved are known TF or novel *BRCA1* regulatory proteins has yet to be determined. Candidate proteins include members of the AP, HNF, FoxO, and NFE-2 families, binding sites for which were identified in these regions of *BRCA1* intron 2. A protein that exclusively regulates *BRCA1* has not yet been identified, but would be an obvious therapeutic target.

Two recent studies have detected the presence of distinct homologous recombination events that have occurred in three different families between intron 2 of *BRCA1* and intron 2 of *BRCA1P1,* leading to 37-kb deletions [61,62]. Puget et al. mapped two breakpoint junctions to intron 2 and interestingly these junctions are located within CNS-1 and CNS-2 identified in this study (Fig. 2). Brown et al. also described a third patient with a similar germ-line deletion; however, the exact breakpoint junction was not mapped at the time. Interestingly, the breakpoint in this patient occurs in CNS-2 (nt 34221–34232 in *BRCA1* and nt 71156–71167 in *BRCA1P1*, referring to GenBank Accession No. AC060780), close to but at a distinct site from the previously identified breakpoints (S. Mayozer, personal communication, and Fig. 5B). Three separate patients demonstrating germ-line deletions within a 220-bp region are consistent with the hypothesis that this conserved noncoding region is a target for recombination-mediated deletions.

In this paper we show that two evolutionarily conserved regions in intron 2 of the *BRCA1* gene transcriptionally regulate *BRCA1* expression in vitro. Furthermore, we provide evidence that this regulation shows specificity for the *BRCA1* promoter and predict that disruption of these elements or the associated *trans*-acting factors may contribute to *BRCA1* repression. A priority now is to assess the functionality of other conserved regions identified in the *BRCA1* gene and the factors that bind to them and to determine the contribution of these elements and factors to *BRCA1* regulation. These regions may include the other seven evolutionarily conserved sequences identified in the *BRCA1* introns noted above, as well as sequences outside of the *BRCA1* gene. The elucidation of regulatory pathways involving noncoding DNA sequences in the human *BRCA1* gene is important not only for understanding the control of normal gene expression, but also for understanding how expression may be compromised in some breast cancer patients.

Materials and methods

Cell culture

Cervical adenocarcinoma HeLa cells (ATCC CCL-2) and SV40 T-antigen-immortalized human mammary epithelial SVCT cells (ECACC 94122105) were cultured in DMEM (Invitrogen, Carlsbad, CA, USA) containing 10% fetal calf serum (Invitrogen). SVCT growth medium was also supplemented with insulin (Sigma–Aldrich Pty. Ltd., Sydney, Australia; 10 μg/ml) and hydrocortisone (Sigma–Aldrich; 0.5 μg/ml). Cells were maintained at 37°C in a 5% CO_2 humidified incubator.

Comparative sequence analysis

BRCA1 genomic DNA from the following species was analyzed in this study: human (GenBank Accession No. L78833), human *BRCA1* pseudogene 1 *BRCA1P1* (GenBank Accession No. AC060780), chimpanzee (GenBank Accession No. AY365046), gorilla (GenBank Accession No. AY589042), orangutan (GenBank Accession No. AY589040), rhesus macaque (GenBank Accession No. AY589041), dog (http://genome.ucsc.edu), mouse (GenBank Accession No. NT_039521), rat (GenBank Accession No. NW_047336), and chicken (http://genome.ucsc.edu). Blocks of high similarity were mapped according to those considered significant by VISTA [63]. Percentage sequence identity within each block was calculated as the number of nucleotides equivalent to the human sequence at each position as a fraction of the total number of nucleotides in the block. Putative transcription factor binding sites were predicted using the TRANSFAC [38] and rVISTA [39] databases.

Cosmid and plasmid vectors used in GET homologous recombination (GHR)

Cosmid A11100 (cosA11100) was used as the primary recombination template [36,64]. This construct contains a 22-kb genomic insert that includes *NBR2* exon 1 and *BRCA1* exons 1A, 1B, 2, and 3 and their surrounding introns (Fig. 3A). Plasmid pGETrec, containing the arabinose-inducible recombination genes *recE, recT,* and *gam,* was kindly provided by P. Ioannou (The Murdoch Children's Research Institute, Melbourne, Australia) [41]. To generate the luciferase (Luc)/hygromycin (Hygro) plasmid, which was used as a PCR template in GHR, the Luc gene was first PCR amplified from pGL3 Basic (Promega, Madison, WI, USA) using XhoHindLuc primers (Supplementary Table 1) and cloned into the *Xho*I/*Hin*dIII sites of pGEM-7Zf (Promega). The Hygro gene was then PCR amplified from pMG-H2 (InvivoGen, San Diego, CA, USA) using HindNsiHygro primers (Supplementary Table 1) and cloned next to the Luc gene into the *Hin*dIII/*Nsi*I sites. To generate the Venus/Hygro template, the pCS2-Venus plasmid [42] (kindly provided by T. Nagai, Brain Science Institute, RIKEN, Japan) was digested and the Venus coding sequence cloned into the *Xho*I/*Hin*dIII of pGEM-7Zf; the Hygro gene was then cloned downstream as described above. Plasmid pBR322 was used as the template for PCR amplification of the tetracycline resistance (TetR) cassette.

Preparation and purification of PCR products for GHR

The Luc/Hygro cassette derived from the pGEM-7zf-Luc-Hygro construct was amplified using the primers GHRLucHygroF and GHRHygroR (Supplementary Table 1). The resultant 3.3-kb PCR product was designated Luc/H (Fig. 3A). The Venus/Hygro cassette derived from the pGEM-7zf-Venus-Hygro construct was amplified using the primers GHRVenusHygroF and GHRHygroR (Supplementary Table 1). The resultant 2.3-kb PCR product was designated Venus/H (Fig. 3A). Both the Luc/H and the Venus/H cassettes were inserted by GHR in-frame into exon 2 of the *BRCA1* gene directly after the ATG translation start codon. The TetR cassettes were amplified using various primers depending on the insertion point into the reporter cosA11100 constructs (Supplementary Table 1). All PCRs were performed in 50-μl reactions for 35 cycles (94°C, 30 s; 55°C, 30 s; 72°C, 4 min) with the Expand High Fidelity PCR system (Roche Diagnostics, Basel, Switzerland). PCR products were gel purified using the QIAquick gel extraction kit (Qiagen, Inc., Valencia, CA, USA) and resuspended in 10 mM Tris–HCl, pH 8.5.

Construction of cosmid reporter vectors using GHR

To generate the reporter cosmids, GHR was performed as previously described [41]. Five hundred nanograms of purified Luc/H or Venus/H PCR product was electroporated (Gene-Pulser II; Bio-Rad, Hercules, CA, USA) into 40 μl of electrocompetent *Escherichia coli* DH10B cells carrying both cosA11100 and pGETrec. Electrocompetent cells were prepared by inoculating 250 ml of LB containing kanamycin (30 μg/ml; Sigma–Aldrich) and ampicillin (100 μg/ml; Amresco, Solon, OH, USA) to an OD_{600} of 0.4. Expression of the pGETrec plasmid was induced by the addition of L-arabinose to a final concentration of 0.2% for 40 min. The cells were then harvested and made electrocompetent by standard procedures. Following electroporation with Luc/H or Venus/H PCR products, cells were incubated in LB medium for 1 h at 37°C and then spread onto LB-agar plates containing Hygro (100 μg/ml; Invitrogen). Single colonies were randomly picked, grown in 1 ml LB containing Hygro, and 1-μl aliquots of culture were used for PCR screening of recombinant clones. For the second round of TetR GET recombination, 500 ng of the TetR PCR product was electroporated into electrocompetent *E. coli* DH10B cells carrying the parental cosA11100 + Luc/H or Venus/H and pGETrec. Electrocompetent cells were prepared and harvested as described above, except modified clones were selected using LB-agar plates containing Tet (100 μg/ml; Sigma–Aldrich). To rescue cosmid clones from the pGETrec plasmid, cells were grown on plates in the absence of Amp, and positive clones were identified by replica plating on antibiotic plates ± Amp. DNA was then extracted using a Qiagen DNA extraction kit (Qiagen, Inc.) and mutagenesis confirmed by restriction mapping and nucleotide sequence analysis.

Construction of pGL3 luciferase reporter vectors

The *BRCA1* 1A promoter was PCR amplified from HeLa genomic DNA with *Xho*IBRCA1promF and *Xho*IBRCA1promR primers (Supplementary Table 1) and then cloned into the *Xho*I site of pGL3-Basic (Promega). This construct is referred to as pGL3-B. Three fragments from *BRCA1* intron 2 were PCR amplified from HeLa genomic DNA: CNS-1, 105 bp (primers *Bam*HIInt2 325F and *Bam*HIInt2 105R, Supplementary Table 1, Fig. 5A); CNS-2, 195 bp (primers *Bam*HIInt2 195F and *Bam*HIInt2 325R, Supplementary Table 1, Fig. 5A); and CNS-1 + 2, 325 bp (primers *Bam*HIInt2 325F and *Bam*HIInt2 325R, Supplementary Table 1, Fig. 5A). The fragments were cloned separately into the *Bam*HI site located downstream of the luciferase gene and SV40 late poly(A) signal of both the pGL3-B and the pGL3 Control (pGL3-C) vectors. The sequence of all reporter gene constructs was confirmed by the Australian Genome Research Facility (AGRF; University of Queensland, St Lucia, QLD, Australia).

Transient transfections and luciferase assays

HeLa and/or SVCT cells were transiently transfected with 1 μg of cosA11100 + Luc/H recombinant clones or pGL3 constructs and 20 ng pRL-TK in a 24-well plate, using Lipofectamine 2000 (Invitrogen). After 24 h, relative luciferase activity was determined using a dual luciferase reporter assay kit (Promega) and a Wallac Microbeta Trilux Luminometer counter (EG&G Wallac, MD, USA) according to the manufacturer's instructions. To correct for any differences in transfection efficiency or cell lysate preparation, firefly luciferase activity was normalized to that of *Renilla* luciferase. Statistical analysis was performed using unpaired, two-tailed *t* tests, with p values <0.01 considered significant.

Fluorescence-activated cell sorting (FACS)

Cells transiently transfected with cosA11100 + Venus/H recombinant clones were trypsinized 24 h after transfection, washed with phosphate-buffered saline, and resuspended in Cell Dissociation Buffer (Invitrogen). The intensity of Venus fluorescence was analyzed by flow cytometry (MoFlo; Cytomation, Fort Collins, CO, USA) with Summit software (v3.1; Cytomation).

RT-PCR

To determine if cosA11100 parental mRNA was expressed, 5 μg of total RNA was extracted from HeLa cells using TRIzol (Invitrogen) and DNase I treated, and cDNA was synthesized using an oligo(dT) primer and Superscript III reverse transcriptase according to the

manufacturer's instructions (Invitrogen). PCR was then performed using *BRCA1* exon 1A- or 1B-specific primers (Supplementary Table 1, BRCA1Exon1A or 1BFor) with a luciferase- or Venus-specific reverse primer (Supplementary Table 1). This cDNA was also used to determine if CNS-2 was an alternative *BRCA1* splice variant. Nested PCR was performed using the BRCA1Exon1AF primer (Supplementary Table 1) with CNS-2-specific downstream (Supplementary Table 1, Int2aRev) and internal (second round PCR, Supplementary Table 1, Int2aRevInt) primers. All PCR products were sequenced by AGRF.

3′RACE

The 3′ end of the cosA11100 parental mRNA was determined using 3′RACE. Five micrograms of total RNA was extracted from transiently transfected HeLa cells using TRIzol (Invitrogen) and DNase I. cDNA was synthesized using 500 ng RACE-1 primer [65] and Superscript III reverse transcriptase (Invitrogen). Nested PCR was performed using RACE-2 primer [65] and luciferase-specific upstream (Supplementary Table 1, 3′RACE Luc For) and internal (second round PCR, Supplementary Table 1, 3′RACE Luc Int) primers. PCR products were sequenced by AGRF.

RNase protection assays

All experiments were performed using the Ribonuclease Protection Assay kit (RPA III; Ambion, Inc., Austin, TX, USA) as specified by the manufacturer. Briefly, ^{32}P-labeled RNA probes were synthesized complementary to the target RNA. The luciferase (400 bp) and Renilla (200 bp) probes used were amplified by PCR from pGL3-Basic and pRL-TK, respectively, using specific primers (Supplementary Table 1) and Expand High Fidelity Taq (Roche). The fragments were then cloned into the pGEM-T Easy vector system (Promega) and linearized with *Sal*I. T7 polymerase was used to generate antisense RNA transcripts using an in vitro transcription kit (Promega). Total RNA (30 μg) and excess labeled probes ($\sim 2 \times 10^5$ cpm) were incubated at 42°C overnight to hybridize the probe to its complement in each sample RNA. After hybridization, the mixtures were treated with a 1:100 dilution of a ribonuclease A/T_1 mix at 37°C for 30 min to degrade unhybridized probe. The protected fragments were run on 8 M urea 5% polyacrylamide gels at 250 V for 2 h, imaged with a phosphoimager (Typhoon 9400; Molecular Dynamics, CA, USA) and quantified using ImageQuant software (v5.0; Molecular Dynamics).

Electrophoretic mobility shift assays (EMSA)

EMSAs were performed as previously described [66]. Nuclear extracts were prepared from HeLa cells according to the method of Dignam et al. [67]. DNA fragments were obtained by digesting pGEM-T plasmids containing either CNS-1 (105 bp) or CNS-2 (195 bp) inserts with *Bam*HI and end-labeling with [γ-^{32}P]ATP using T4 polynucleotide kinase (MBI Fermentas, Hanover, MD, USA). For binding reactions, 20,000 cpm of labeled probe was incubated with 10 μg of nuclear protein in a 30-μl reaction mixture containing 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 1 mM $MgCl_2$, 5% glycerol, and 3 μg poly(dI–dC). For competition experiments, 100-fold excess unlabeled self-fragment or nonspecific DNA was added to the binding reaction before the addition of labeled probe. Reaction mixtures were incubated at room temperature for 25 min and then separated on 6% native polyacrylamide gels. DNA–protein complexes were visualized by exposure of dried gels to Kodak XAR film at −80°C.

Mutation analysis

DNA samples were obtained from the Kathleen Cuningham Consortium for Research into Familial Breast Cancer (kConFab). Seventy-nine samples were from index cases (defined as the youngest affected from whom a blood sample was available) of families meeting the kConFab criteria for high-risk familial breast cancer (www.kconfab.org), but who had previously tested negative for mutations in the coding regions of *BRCA1* and *BRCA2*. Mutational screening for *BRCA1* intron 2 mutations was carried out by prescreening with DHPLC analysis [68]. Briefly, the 325-bp fragment from intron 2 of the *BRCA1* gene was PCR amplified from 20 ng of genomic DNA using primers *Bam*HIInt2 325F and *Bam*HIInt2 325R (Supplementary Table 1). PCR amplification was carried out using a touchdown program: 94°C (30 s), 65°C (30 s), and 72°C (1 min) with the annealing temperature being lowered by 0.5°C each cycle for 20 cycles, followed by 20 cycles at 94°C (30 s), 55°C (30 s), and 72°C (1 min). PCR products were then subjected to a slow annealing step from 95 to 60°C over 40 min (1°C/min) to promote heteroduplex formation prior to analysis on the Varian Helix System (Varian, Walnut Creek, CA, USA). DHPLC was carried out at the recommended melt temperatures for all amplicons as determined by the Stanford melt algorithm (http://insertion.stanford.edu/melt.html). Analysis was performed using the Star Workstation version 5 (Varian).

Acknowledgments

The authors are grateful to Panayiotis Ioannou (Murdoch Institute, Melbourne) for the pGETrec plasmid, the ICRF (Cancer Research UK) for the use of the SVCT cell line, Matthew Wakefield (ANU, Canberra) for the marsupial genomic DNA, Sylvie Mayozer (IARC, Lyon) for primers and help with mapping *BRCA1* deletions, Georgia Chenevix-Trench and Aaron Lewis (QIMR, Brisbane) for help with mutation analysis, and Ibtissam Abdul Jabbar (UQ, Brisbane) for MoFlo FACS. The authors thank the kConFab research nurses and staff for data collection; Heather

Thorne, Lynda Williams, and Danni Surace (Peter MacCallum Cancer Institute (PMCI), Melbourne) for DNA preparation; Eveline Niedermayr (PMCI) for supplying data; the staff of the Familial Cancer Clinics for their support of kConFab; and the families for their participation. The authors also thank Ross McAree and Kevin Austin (UQ, Brisbane) and Sean Tavtigian (IARC, Lyon) for valuable discussions and Juliet French and Jodi Saunus (UQ, Brisbane) and Georgia Chenevix-Trench (QIMR, Brisbane) for critically reading the manuscript. This work was supported by grants from The National Health and Medical Research Council of Australia, the UQ Foundation, and UQ Research Development Grant Scheme. kConFab is supported by the National Health and Medical Research Council of Australia; the National Breast Cancer Foundation; and the Cancer Councils of Victoria, Western Australia, Queensland, New South Wales, and South Australia; Perpetual Philanthropic Foundations NSW; and Breast Cancer Research Association, Inc.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2005.05.006.

References

[1] Q. Li, S. Harju, K.R. Peterson, Locus control regions: coming of age at a decade plus, Trends Genet. 15 (1999) 403–408.
[2] E.S. Hwang, A. Choi, I.-C. Ho, Transcriptional regulation of GATA-3 by an intronic regulatory region and fetal liver zinc finger protein 1, J. Immunol. 169 (2002) 248–253.
[3] M.A. Wilson, S.A. Chrysogelos, Identification and characterization of a negative regulatory element within the epidermal growth factor receptor gene first intron in hormone-dependent breast cancer cells, J. Cell. Biochem. 85 (2002) 601–614.
[4] B. Ge, O. Li, P. Wilder, A. Rizzino, T.W. McKeithan, NF-κB regulates *BCL3* transcription in T lymphocytes through an intronic enhancer, J. Immunol. 171 (2003) 4210–4218.
[5] A.C. Gao, et al., Defining regulatory elements in the human *KAI1* (CD 82) metastasis suppressor gene, Prostate 57 (2003) 256–260.
[6] R.L. Brake, P.K. Chatterjee, U.R. Kees, P.M. Watt, The functional mapping of long-range transcription control elements of the *HOX11* proto-oncogene, Biochem. Biophys. Res. Commun. 313 (2004) 327–335.
[7] C. Pena, et al., Intronic deletion affecting a negative regulatory region of TP73 is related to breast and colorectal carcinoma, Genes Chromosomes Cancer 39 (2004) 257–262.
[8] G. Bejerano, et al., Ultraconserved elements in the human genome, Science 304 (2004) 1321–1325.
[9] J.W. Thomas, et al., Comparative analyses of multi-species sequences from targeted genomic regions, Nature 424 (2003) 788–793.
[10] S.H. Williams, N. Mouchel, A. Harris, A comparative genomic analysis of the cow, pig, and human CFTR genes identifies potential intronic regulatory elements, Genomics 81 (2003) 628–639.
[11] K.A. Frazer, et al., Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional, Genome Res. 14 (2004) 367–372.
[12] G.G. Loots, et al., Identification of a coordinate regulator of interleukins 4, 13 and 5 by cross-species sequence comparisons, Science 288 (2000) 136–140.
[13] M. Shnyreva, et al., Evolutionarily conserved sequence elements that positively regulate IFN-γ expression in T cells, Proc. Natl. Acad. Sci. USA 101 (2004) 12622–12627.
[14] M.A. Nobrega, I. Ovcharenko, V. Afzal, E.M. Rubin, Scanning human gene deserts for long-range enhancers, Science 302 (2003) 413.
[15] Y. Miki, et al., A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, Science 266 (1994) 66–71.
[16] S.A. Narod, W.D. Foulkes, BRCA1 and BRCA2: 1994 and beyond, Nat. Rev. Cancer 4 (2004) 665–676.
[17] A.R. Venkitaraman, Cancer susceptibility and the functions of BRCA1 and BRCA2, Cell 108 (2002) 171–182.
[18] L.H. Castilla, et al., Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer, Nat. Genet. 8 (1994) 387–391.
[19] D.F. Easton, D.T. Bishop, D. Ford, G.P. Crockford, Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium, Am. J. Hum. Genet. 52 (1993) 678–701.
[20] J. Peto, et al., Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer, J. Nat. Cancer Inst. 91 (1999) 943–949.
[21] J. Taylor, et al., An important role for BRCA1 in breast cancer progression is indicated by its loss in a large proportion of non-familial breast cancers, Int. J. Cancer 79 (1998) 334–342.
[22] M.E. Thompson, R.A. Jensen, P.S. Obermiller, D.L. Page, J.T. Holt, Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression, Nat. Genet. 9 (1995) 444–450.
[23] C.A. Wilson, et al., Localization of human BRCA1 and its loss in high-grade, non-inherited breast carcinomas, Nat. Genet. 21 (1999) 236–240.
[24] C.F. Xu, et al., Isolation and characterisation of the NBR2 gene which lies head to head with the human BRCA1 gene, Hum. Mol. Genet. 6 (1997) 1057–1062.
[25] S.D. Dimitrov, E. Matouskova, J. Forejt, Expression of BRCA1, NBR1 and NBR2 genes in human breast cancer cells, Folia Biol. (Prague) 47 (2001) 120–127.
[26] T.C. Suen, P.E. Goss, Transcription of BRCA1 is dependent on the formation of a specific protein–DNA complex on the minimal BRCA1 bi-directional promoter, J. Biol. Chem. 274 (1999) 31297–31304.
[27] D.N. Mancini, et al., CpG methylation within the 5′ regulatory region of the BRCA1 gene is tumor specific and includes a putative CREB binding site, Oncogene 16 (1998) 1161–1169.
[28] S. Thakur, C.M. Croce, Positive regulation of the BRCA1 promoter, J. Biol. Chem. 274 (1999) 8837–8843.
[29] E. Atlas, M. Stramwasser, K. Whiskin, C.R. Mueller, GA-binding protein alpha/beta is a critical regulator of the BRCA1 promoter, Oncogene 19 (2000) 1933–1940.
[30] A. Wang, R. Schneider-Broussard, A.P. Kumar, M.C. MacLeod, D.G. Johnson, Regulation of BRCA1 expression by the Rb–E2F pathway, J. Biol. Chem. 275 (2000) 4532–4536.
[31] A. Catteau, W.H. Harris, C.F. Xu, E. Solomon, Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics, Oncogene 18 (1999) 1957–1965.
[32] A. Dobrovic, D. Simpfendorfer, Methylation of the BRCA1 gene in sporadic breast cancer, Cancer Res. 57 (1997) 3347–3350.
[33] F. Magdinier, S. Ribieras, G.M. Lenoir, L. Frappart, R. Dante, Downregulation of BRCA1 in human sporadic breast cancer; analysis of DNA methylation patterns of the putative promoter region, Oncogene 17 (1998) 3169–3176.
[34] J.C. Rice, K.S. Massey-Brown, B.W. Futscher, Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells, Oncogene 17 (1998) 1807–1812.

[35] T.C. Suen, P.E. Goss, Identification of a novel transcriptional repressor element located in the first intron of the human BRCA1 gene, Oncogene 20 (2001) 440–450.
[36] M.A. Brown, et al., The 5′ end of the BRCA1 gene lies within a duplicated region of human chromosome 17q21, Oncogene 12 (1996) 2507–2513.
[37] T.M. Smith, et al., Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1, Genome Res. 6 (1996) 1029–1049.
[38] E. Wingender, et al., Transfac: an integrated system for gene expression regulation, Nucleic Acids Res. 28 (2000) 316–319.
[39] G.G. Loots, I. Ovcharenko, rVISTA 2.0: evolutionary analysis of transcription factor binding sites, Nucleic Acids Res. 32 (2004) W217–W221.
[40] J.W. Fickett, W.W. Wasserman, Discovery and modeling of transcriptional regulatory elements, Curr. Opin. Biotechnol. 11 (2000) 19–24.
[41] K. Narayanan, R. Williamson, Y. Zhang, A.F. Stewart, P.A. Ioannou, Efficient and precise engineering of a 200 kb beta-globin human/-bacterial artificial chromosome in E. coli DH10B using an inducible homologous recombination system, Gene Ther. 6 (1999) 442–447.
[42] T. Nagai, et al., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications, Nat. Biotechnol. 20 (2002) 87–90.
[43] C.F. Xu, et al., Distinct transcription start sites generate two forms of BRCA1 mRNA, Hum. Mol. Genet. 4 (1995) 2259–2264.
[44] D. Boffelli, et al., Phylogenetic shadowing of primate sequences to find functional regions of the human genome, Science 299 (2003) 1391–1394.
[45] T. Chen, et al., An intronic variant of the TGFBR1 gene is associated with carcinomas of the kidney and bladder, Int. J. Cancer 112 (2004) 420–425.
[46] M. Rusin, H. Zientek, M. Krzesniak, E. Malusecka, A. Zborek, Intronic polymorphism (1541–1542delGT) of the constitutive heat shock protein 70 gene has functional significance and shows evidence of association with lung cancer risk, Mol. Carcinog. 39 (2004) 155–163.
[47] T.I. Orban, E. Olah, Emerging roles of BRCA1 alternative splicing, Mol. Pathol. 56 (2003) 191–197.
[48] D. Tautz, Evolution of transcriptional regulation, Curr. Opin. Genet. Dev. 10 (2000) 575–579.
[49] L.A. Pennacchio, et al., An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing, Science 294 (2001) 169–173.
[50] J.P. Vaughn, et al., *BRCA1* expression is induced before DNA synthesis in both normal and tumor-derived breast cells, Cell Growth Differ. 7 (1996) 711–715.
[51] B. Aunoble, D. Bernard-Gallon, Y.J. Bignon, Regulation of BRCA1 and BRCA2 transcript in response to cisplatin, adriamycin, taxol and ionising radiation is correlated to p53 functional status in ovarian cancer cell lines, Oncol. Rep. 8 (2001) 663–668.
[52] S.T. Marquis, et al., The developmental pattern of BRCA1 expression implies a role in differentiation of the breast and other tissues, Nat. Genet. 11 (1995) 17–26.
[53] D.A. Favy, et al., Prolactin-dependent up-regulation ofBRCA1 expression in human breast cancer cell lines, Biochem. Biophys. Res. Commun. 258 (1999) 284–291.
[54] E.T. Dermitzakis, et al., Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs), Science 302 (2003) 1033–1035.
[55] M.P. Hare, S.R. Palumbi, High intron sequence conservation across three mammalian orders suggests functional constraints, Mol. Biol. Evol. 20 (2003) 969–978.
[56] B.J. Orelli, J.M. Logsdon, D.K. Bishop, Nine novel conserved motifs in BRCA1 identified by the chicken orthologue, Oncogene 20 (2001) 4433–4438.
[57] R. Barthel, A.E. Goldfeld, T cell-specific expression of the human TNF-α gene involves a functional and highly conserved chromatin signature in intron 3, J. Immunol. 171 (2003) 3612–3619.
[58] P. Zheng, L.A. Pennacchio, W.L. Goff, E.M. Rubin, J.D. Smith, Identification of a novel enhancer of brain expression near the apoE gene cluster by comparative genomics, Biochim. Biophys. Acta 1676 (2003) 41–50.
[59] A. Lu, F. Zhang, A. Gupta, J. Liu, Blockade of AP1 transactivation abrogates the abnormal expression of breast cancer-specific gene 1 in breast cancer cells, J. Biol. Chem. 277 (2002) 31364–31372.
[60] M.R. Loyd, Y. Okamoto, M.S. Randall, P.A. Ney, Role of AP1/NFE2 binding sites in endogenous alpha-globin gene transcription, Blood 102 (2003) 4223–4228.
[61] M.A. Brown, et al., Germline BRCA1 promoter deletions in UK and Australian familial breast cancer patients: identification of a novel deletion consistent with BRCA1:psiBRCA1 recombination, Hum. Mutat. 19 (2002) 435–442.
[62] N. Puget, et al., Distinct *BRCA1* rearrangements involving the *BRCA1* pseudogene suggest the existence of a recombination hot spot, Am. J. Hum. Genet. 70 (2002) 858–865.
[63] G.G. Loots, I. Ovcharenko, L. Pachter, I. Dubchak, E.M. Rubin, rVISTA for comparative sequence-based discovery of functional transcription factor binding sites, Genome Res. 12 (2002) 832–839.
[64] M.A. Brown, et al., Physical mapping, cloning, and identification of genes within a 500-kb region containing BRCA1, Proc. Natl. Acad. Sci. USA 92 (1995) 4362–4366.
[65] S.L. Wardrop, C. Wells, T. Ravasi, D.A. Hume, D.R. Richardson, Induction of Nramp2 in activated mouse macrophages is dissociated from regulation of the Nramp1, classical inflammatory genes, and genes involved in iron metabolism, J. Leukocyte Biol. 71 (2002) 99–106.
[66] T.C. Suen, P.E. Goss, Characterization of a repressor element and a juxtaposed tissue-restricted activator element located on the distal neu gene promoter, J. Biol. Chem. 275 (2000) 6600–6607.
[67] J.D. Dignam, R.M. Lebovitz, R.G. Roeder, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, Nucleic Acids Res. 11 (1983) 1475–1489.
[68] A.C. Jones, et al., Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis, Clin. Chem. 45 (1999) 1133–1140.