Conservation of the BRCA1 Gene

Approved: ________________________________

Gary R. Skuse, Ph.D.
Director of Bioinformatics

____

Richard L. Doolittle, Ph.D.
Head, Department of Biological Sciences

Submitted in partial fulfillment of the requirements for the Master of Science degree in Bioinformatics at the Rochester Institute of Technology.

Matthew Wronkowski

April 2007

Committee Members

Anne Haake, Ph.D. is professor of information technology and bioinformatics courses at the Rochester Institute of Technology.

Dina Newman, Ph.D. is a professor of human genomics and biology courses in the College of Science at the Rochester Institute of Technology.

Maureen Ferran, Ph.D. is a professor of biology and virology in the College of Science at the Rochester Institute of Technology.

Gary Skuse, Ph.D. is director of the Bioinformatics program at the Rochester Institute of Technology.

Copyright Release Form

Title of thesis: Conservation of the BRCA1 Gene

Name of author: Matthew Gary Wronkowski
Degree: Masters of Science
Program: Bioinformatics
College: Science

I understand that I must submit a print copy of my thesis to the RIT Archives, per current RIT guidelines for the completion of my degree. I hereby grant the Rochester Institute of Technology and its agents the non-exclusive license to archive and make accessible my thesis in whole or in part in all forms of media perpetuity. I retain all other ownership rights to the copyright of the thesis. I also retain the right to use in future work (such as articles and books) all or part of the thesis.

Print Reproduction Permission Granted:
I, Matthew Gary Wronkowski, hereby grant permission to the Rochester Institute of Technology to reproduce my print thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature of Author: ____ Date: ____________

Inclusion in the RIT Digital Media Library (ETD)
I, Matthew Gary Wronkowski, additionally grant to the Rochester Institute of Technology Digital Media Library (RIT DML) the non-exclusive license to archive and provide electronic access to my thesis in whole or in part in all forms of media in perpetuity.

I understand that my work, in addition to its bibliographic record and abstract, will be available to the world-wide community of scholars and researchers through the RIT DML. I retain all other ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles and books) all or part of the thesis. I am aware that the Rochester Institute of Technology does not require registration of copyright for ETDs.

I hereby certify that, if appropriate, I have obtained and attached written permission statements from the owners of each third party copyrighted matter to be included in my thesis. I certify that the version I submitted is the same as that approved by my committee.

Signature of Author: ____ Date: ____________

Abstract

Most research has been conducted on the structure and function of many individual genes as well as complementary proteins. Many of these studies look at specific population groups' country of origin, geographic location, or ethnicity but do not correlate findings between these groups. In this project I attempted to associate the structural and functional consequences of individual mutations in Human *BRCA1* to other organisms. My findings suggest a strong correlation between human geographic isolation and mutation exclusivity. Furthermore strong sequence and mutation conservation as well as structural homology exist between many species.

Acknowledgments

First, I would like to extend my appreciation to the professors who dedicated countless hours and effort preparing material as we ventured into the first year of this master's program.

Second, I want to recognize my sister for proofreading and assistance formatting this document.

List of Figures

Figure 1 – Map of BRCA1 3
Figure 2 – BRCA1/2 Related Diseases 8
Figure 3 – BRCA1Pathway 9
Figure 4 – BRCA1-MRE11/Rad50 10
Figure 5 – NMR Predictions of 1JM7 13
Figure 6 – ClustalW Alignment 20
Figure 7 – BRCA1 Phase Checkpoints 24
Figure 8 – ATM-ATR Damage Response Pathway 25
Figure 9 – BRCA1 Pathway 30
Figure 10 – BRCA1-FANCD2 Pathway 32
Figure 11 – BRCA1-FANCD2 Pathway 33
Figure 12 – BRCA1 Binding Domains 34
Figure 13 – Single BRCT Domain and BRCT Dimer 36
Figure 14 – C3HC4 Representation 37
Figure 15 – BRCA1 RING Domain bound to BARD1 38
Figure 16 – Mouse Comparative Map 61

List of Tables

Table 1 – MolScript sample input for 1JM7 14
Table 2 – SQL Queries for BIC Database 16
Table 3 – Most Frequent Mutations in BIC Database 17
Table 4 - Frequent Mutations in C-terminus 17
Table 5 – Populations with High BRCA1 Mutations 18
Table 6 – BRCA1 Related Proteins 23
Table 7 – Frequent BRCA1 Mutations 41
Table 8 – BRCT Secondary Structure Effects 43
Table 9 – BRCN Secondary Structure Effects 46
Table 10 – BRCA1 Central Region Mutations 47
Table 11 – 1JM7 48
Table 12 – Conservation of RING Domain 48
Table 13 – 1JNX 49
Table 14 – Conservation of BRCT Domain 50
Table 15 – BRCT vs. XRCC1 51
Table 16 – Cumulative Breast Cancer Risk 52
Table 17 – Founder / Population Specific Mutations 55
Table 18 – Cancer Causing Mutations 56
Table 19 – BRCA1 N-terminal Homology 57
Table 20 – BRCA1 C-terminal Homology 58
Table 21 – BRCA1 Exon 11 59
Table 22 – Human-Canine duplication 62
Table 23 – Human-Drosophila 62
Table 24 – Percent Identity to Human BRCA1 63

Table of Contents

Committee Members ii
Copyright Release Form iii
Abstract iv
Acknowledgments v
List of Figures vi
List of Tables vii
Table of Contents viii
Introduction 1
BRCA1 – The Breast Cancer Gene *2*
Materials and Results 4
Gene Function *6*
Protein Structure *10*
Structure/Functional Mutations *14*
Conservation *19*
Conservation in Other Species *20*
Gene Function 21
G1 Phase *24*
S Phase *26*
G2 phase *27*
G2 repair *31*
M phase *33*
Structure of BRCA1 Proteins 33
BRCT – BRCA1 C-terminal *34*
RING – BRCA1 N-terminal *36*
Other BRCA1 Binding Sites *38*
BRCA1 Mutations 39
Most Frequent Mutations Reported by BIC *39*
BRCT Mutations *40*
RING Mutations *44*
Central Region / Other Domain mutations *46*
BRCA1 Conservation 47
Risk of Breast Cancer *51*
BRCA1 Cancer Causing Mutations *53*
BRCA1 Conservation in Other Species *56*
BRCA1 Central Region Conservation *59*
BRCA1 Mouse Model *60*
Rattus norvegicus *61*
Canis familiaris *62*

Drosophila 62
Overall Conservation 63
Discussion 64
Bibliography 66
Appendix A ..... A
Molscript input file for 1JNX terminus: ..... A
Molscript input file for 1JNX terminus: ..... B

Introduction

Breast cancer is the second leading cause of cancer death in women after lung cancer. Every year 200,000 women and 1,000 men will be diagnosed with this disease in the US of which approximately 40,000 will die (American Cancer Society, 2005). However these rates continue to decline due to a combination of improved early detection and treatment techniques.

Although the exact cause of breast cancer is unknown, all cancers begin when a single cell containing altered genetic material has disrupted normal cellular regulation. The result may be lack of gene expression, increased cell death, or uncontrolled cell growth (cancer). There are dozens of factors that may alter an individual's risk of having these mutations. Though not all-inclusive, these may include age, race, smoking, family history / genetics, personal history, and number of pregnancies. This project will focus primarily on heritage and genetic background factors.

Symptoms of breast cancer can range from swelling and skin irritation to physically noticeable lumps. Detection consists of routine clinical MRIs or mammograms accompanied by self-examination. If a tumor is suspected, further x-rays may be taken and a biopsy performed. The doctor may also perform staging tests to determine if the cancer has spread to other tissues. Tests include PET, bone, and CT scans.

Breast cancer proliferates through direct invasion, lymphatic spreading or the circulatory system. Direct invasion is characterized by when a tumor invades and destroys surrounding tissue. Lymphatic spreading occurs when cells breaking away from the tumor are collected by the lymphatic system and are deposited throughout the body. Similarly, the cancer may spread to other organs via the circulatory system as tiny clumps or as individual cells break away. It is unclear how a tumor grows, but it is known that surrounding blood vessels must infiltrate the tumor for it to reach sizes beyond a few millimeters in diameter.

Treatment options are classified into either local or systemic categories. In the past, surgery or localized radiation treatments were administered with broad based chemotherapy.

Systemic treatments such as chemotherapy target any rapidly reproducing cells. Unfortunately this leads to the destruction of both the cancer cells and other rapidly reproducing cells such as blood producing bone marrow.

Recently, combinations of localized and ER-targeted systemic methods have become more attractive as they inhibit carcinomas throughout the body. This allows simultaneous treatment of all three potential spreading methods. Hormone supplements and immunotherapy are two common forms of combined localized and systemic methods.

Cell growth and division is controlled by cell surface receptors and intracellular steroidal hormone receptors in addition other growth factors (Osborne, 2004). Endocrine therapies block estrogen-signaling pathways by lowering the amount of estrogen available for cell binding. A systemic endocrine hormone treatment known as Tamoxifen (Tam) is the most effective in its class. Tamoxifen targets ER (estrogen receptor) positive breast cancer and is administered to patients in conjunction with other therapeutic methods. However, many patients develop resistance to this drug after prolonged use (Schafer, *et al.* 2002).

BRCA1 – The Breast Cancer Gene

During the past decade numerous proteins and over 70 genes have been identified as having links to breast, ovarian, and other reproductive system cancers. None is more famous or studied than the *BRCA1* (BReast CAncer 1 / Breast Cancer type 1 Susceptibility Protein) gene discovered in 1994 by Mary-Claire King at UC Berkeley (Friedman, 1994). A year later, a second gene, *BRCA2*, was identified (Wooster, 1995). *BRCA1* is located at chromosomal locus 17q21.31, between *NBR2* and *RPL27* (Figure 1).

The original *BRCA1* gene was discovered after studying families with strong patterns and histories of breast cancer. In each candidate similar mutations along chromosome 17 (Friedman, 1999) were noticed when compared to unaffected patients. After further studies it was concluded that defects or mutations in *BRCA1* lead to an increased risk of developing cancers. Although only 5% of cancer is related to a *BRCA1* mutation (Imaginis, 2006), the inherited risk for

developing the disease is as high as 40% (King, 2003). *BRCA1* is not limited to just breast cancer but also associated with ovarian, cervical, and numerous other tissue cancers.

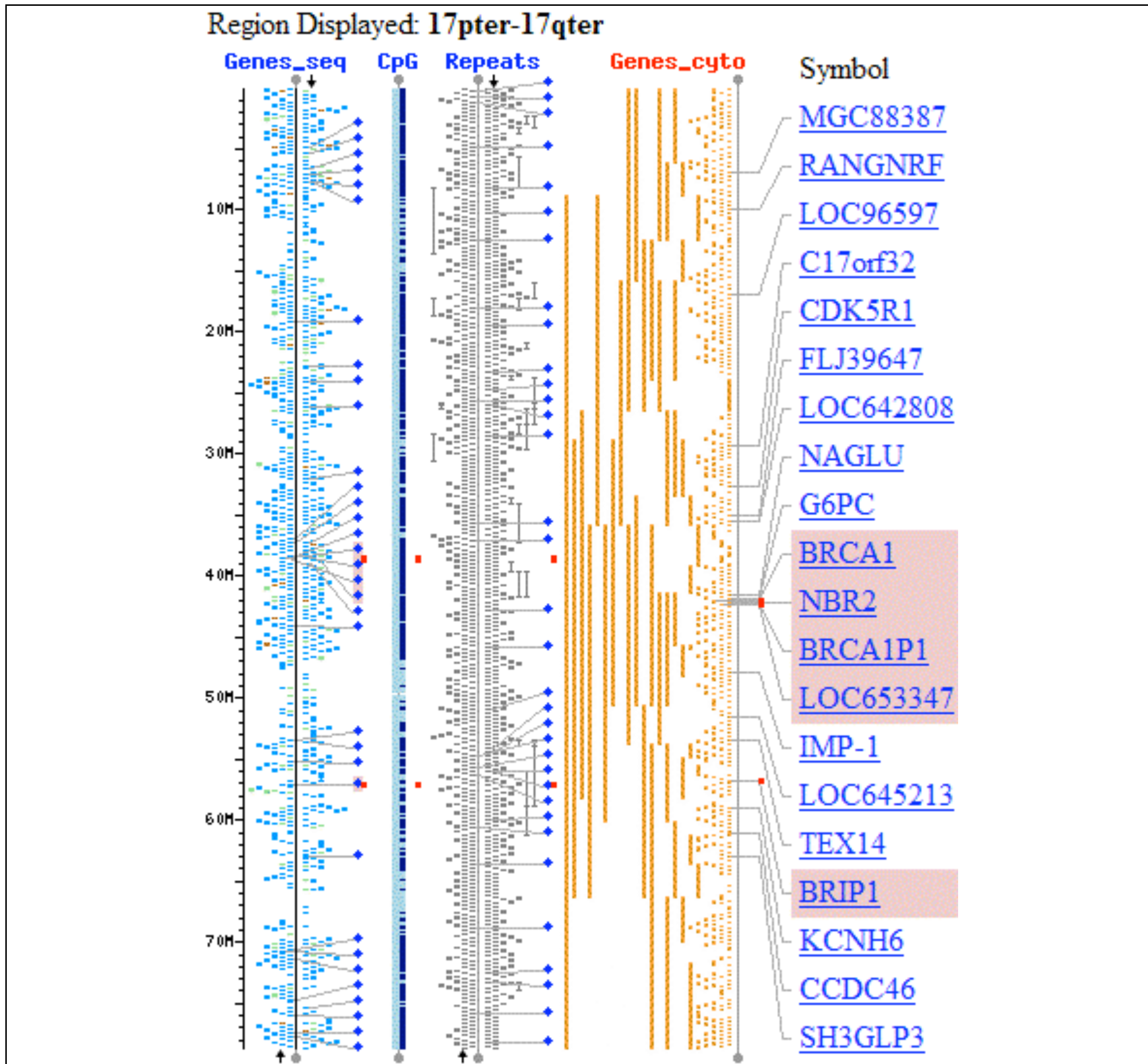

Figure 1 – Map of BRCA1
Location of *BRCA1* on chromosome 17 between *NBR2* and *RPL17* in a strong CpG isle. Note upstream duplication of *BRCA1* exons 1A, 1B and 2 deemed *BRCA1P1*. (Source: NCBI MapViewer)

BRCA1 and *BRCA2* are known as tumor suppressor genes. *BRCA1* serves as a negative regulator of cell growth (Thompson, 1995) by inhibiting cell growth in breast and ovarian cells (Holt, 1996). Functional loss tends to put carriers at a higher risk of developing cancers but does not guarantee cancer growth.

Frequent loss of heterozygosity (LOH) in both somatic and germline cells gave rise to the theory *BRCA1*'s tumor suppressor role (Geisler, 2002). LOH is the loss of one parent's contribution and typically caused by large deletions in one allele. *BRCA1* inherited tumorigenesis consist of a LOH combined with germline mutations in the other allele. Germline mutations typically are small insertions, deletions or single base pair changes.

Since *BRCA1* resides on chromosome 17 it is inherited equally by both men and women. *BRCA1* operates in a heterozygous manner, that is, a mutation in only one allele puts a person at risk of the disease. It should be noted only the risk is inherited, not the disease.

Breast cancer is often accompanied by several other syndromes expressed in upstream regulators of *BRCA1*. These include Li-Fraumeni (*TP53* gene), Cowden Disease (*PTEN* gene), Peutz-Jegher syndrome (*STK11* gene), Ataxia-Telangiectasia (*ATM* gene). These syndromes are caused by mutations in the individual genes and are not influenced by interactions with BRCA1.

Materials and Results

Research revealed a wealth of sequence, functional and mutation information available for the *BRCA1* gene. This project was divided into three parts to study the conservation of *BRCA1*. The first section investigated the biochemical functional aspects of the gene. Basic physiological understanding provided a foundation for analyzing structural and functional consequences resulting from mutations. The second section investigated sequence conservation within the human species between different population groups. Extensive analysis on the functional implications of mutations was studied and correlated between geographic populations. Finally, I explored homology to other species looking at both protein sequence conservation as well as functional repercussions, where available.

The PubMed archives in conjunction with the Oxford Human Molecular Genetics engine were the two most valuable tools used during this project. The information gathered from these sites was utilized as the basis of my research. To begin my research, I started with a NCBI OMIM literature search to obtain an overview of current studies on the *BRCA1* gene. OMIM (Online

Mendelian Inheritance in Man) is a hand edited catalog of human genes and associated disorders. Searching on a particular gene or disease will result in an Entrez ID number along with a textual brief. *BRCA1* for example is listed as +113705 and #114480 for the gene and disease, respectively. Each brief is a combined overview of individual research articles. Many authors have summarized article abstracts into a complete history, genetic information, animal models and a few allelic variations. Full references including direct links to article abstracts are provided. The publishing journal provides access to entire paper texts.

The NCBI site resources are very extensive causing disorientation first use. However, after spending a few hours using these tools I discovered how powerful they were in mining the vast amounts of data available for this project. The Entrez Database model gives a great overview of how each NCBI database is cross-linked to other databases. A single NCBI database entry is referenced to multiple other databases. For example, from an OMIN entry contains direct links to the corresponding Entrez Gene entries.

Entrez Gene contains curated gene specific data from a variety of NCBI sources. Each entry contains a summary of gene function, the reference DNA sequence, and interacting proteins among other related sequence information. Entries also have link-outs to nucleotide, translated protein, and transcripts (Unigene) sequences. Depending upon the type of gene entry selected, one or more sequences will be available. I used this tool mainly for beginning my pathway study of *BRCA1* interacting proteins.

To begin pathway research I used the BIND (Biomolecular Interaction Network Database) provided by Unleashed Informatics. BIND is a collection of molecular interactions between proteins gathered from NCBI literature sources. These basic interactions can be linked together to form a molecular complex or pathway. Over 200k individual interactions have been recorded through either hand curation or high-throughput screening methods. *BRCA1* has 170 molecular interactions and 15 complexes listed in the database.

Interactions are listed with number of domains to each molecule, exact binding sites and the cellular binding location. This information was very useful in building a binding map used during the mutation effect portion of my paper. Entries also contain link-outs to Pfam, GenBank and several gene information sites like Swiss-Prot or UniProt.

Using a few simple text queries, I was able to identify about 200 possible protein interactions on *BRCA1*. Of these, only a handful had site specific binding information available. The remainder of protein binding predictions were based on conformation characteristics but not backed by scientific data nor referenced in literature.

One problem I found with BIND is the number of duplicate entries. Although much of the data has been hand-curated a fair amount is repetitious due to the automated submission engine. I found myself consistently running across multiple entries for the same data and needed to purge them from my results.

After using BIND, I turned my attention to the Human Protein Reference Database or HPRD. HPRD is a platform for depicting protein domains, interaction networks and post-translational modification substrates. Unlike BIND this site uses 100% hand extracted data so information is deemed reliable. Searching for a particular protein yields a list of pictures annotating binding domain position on the protein sequence. I preferred working with this site over BIND as the binding sites and enzymes are presented in a more logical manor. The PTMs, post-translational sites, are listed with upstream enzyme, reaction type and reaction site number. This was very helpful for determining sites to monitor for mutation affects later in my research.

Gene Function

Perhaps the second most useful tool for studying the *BRCA1* pathway was Pathway Studio and Medscan. Both are commercial packages from Ariadne Genomics created for the extraction and modeling of biological pathways.

Pathway Studio is a software tool used for analyzing relationships between proteins, cell processes, and small molecules to develop biological pathways visualizations. The tool has three

data input sources: 1) microarray data, 2) NCBI literature search, 3) and a proprietary database named ResNet. ResNet is a precompiled database generated from scientific literature, public & private data sources of signaling, and biochemical pathways using automated text mining tools. For my study I used the second and third inputs.

For my first search I wanted to graph the diseases linked to *BRCA1*. A search on "*BRCA1*" in the disease database returns a list of eleven objects. Objects or 'nodes' are physical objects such as small molecules, proteins and complexes. Ten of these were proteins and one is a membrane component.

Figure 2 displays the program's ability to generate a map of diseases related to the *BRCA1* and *BRCA2* proteins. To build the pathway some or all of the objects can be selected and graphed. Based on my research, some of the objects listed are outliers not relating directly to the *BRCA1* functional pathway. These include the proteins *BCCIP* and *MRPL36* and the membrane component *M17S2*. Selecting the remaining eight I built a shortest path graph which constructs the shortest disease relationship between the entities. Between each node appears a 'control.' Controls are actions such as expression, regulation, binding or other chemical reaction. As can be observed in Figure 2, both *BRCA1* and *BARD1* have direct relationships to breast cancers.

Often graphing the shortest path will not build a useful pathway because the software is aware that two proteins may require a small molecule to react. For example, *ATR* requires several small molecules to react with *BRCA1* and therefore does not appear on many shortest path type graphs.

I limited my search to return results with less than 100 nodes. This allowed plenty of controls to understand the mechanics of the pathway but kept them to a manageable number. Numbers larger than this tend to produce graphs that are extremely 'busy' making them difficult to read. As the number of components increases, the number of directed relationships grows exponentially. In turn, this result in longer compute times for a complete pathway graph. These would be useful for intricate studies but they are over complicated for my project. Figure 3 is a

very broad search with 5 levels of depth. The resulting graph contains over 2000 nodes with 6500 controls leaving it essentially illegible without manual rearrangement of nodes.

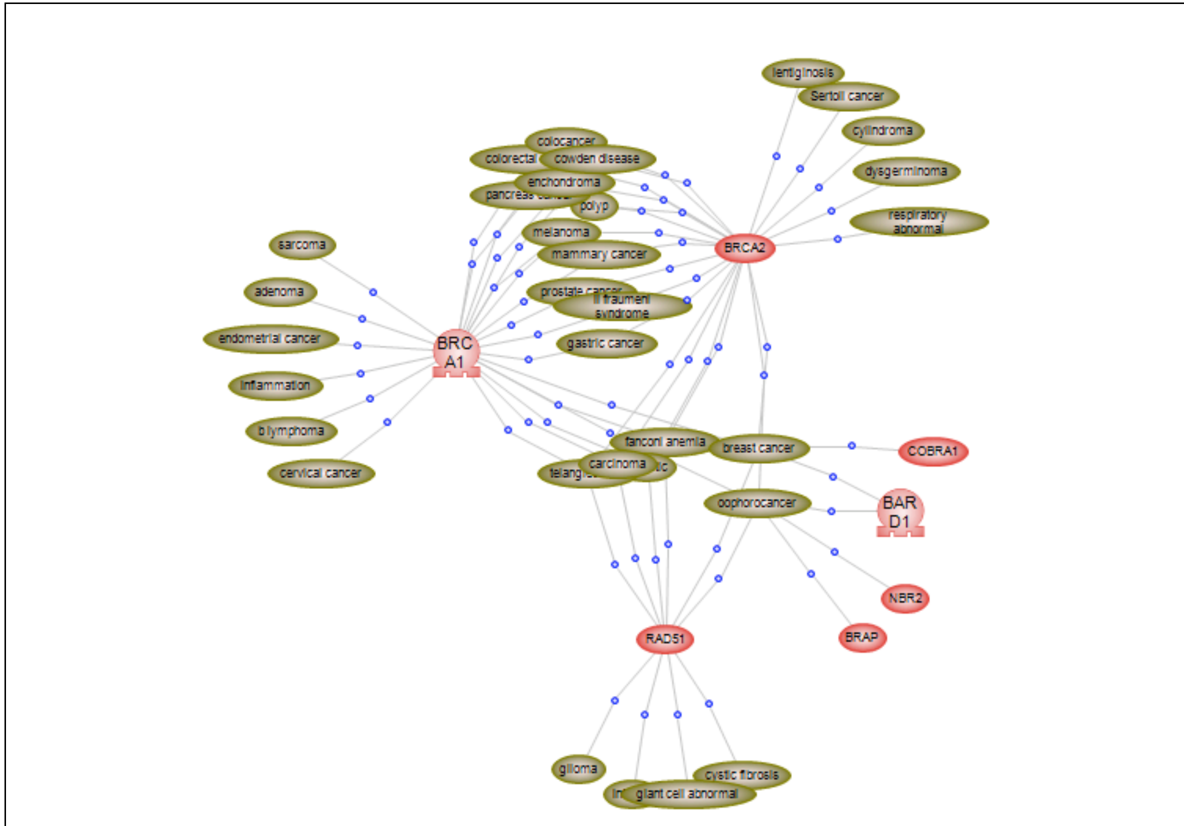

Figure 2 – BRCA1/2 Related Diseases
Diseases with linkages to *BRCA1* type breast cancer are pictured above. Many diseases link to both *BRCA1* and *BRCA2* while others are associated to only one gene. Graph created using ReNet database in Pathway Studio Plus.

Medscan (Pathway Assist) is another Ariadne software package utilizing the same data suite as Pathway Studio. This feature allows pathways to be built using a real-time scan of PubMed abstracts. Medscan is able to create pathways by extracting information from article abstracts using a proprietary dictionary to build relations between proteins.

Since the *BRCA1* pathway is still under investigation neither tool is able to build a complete pathway. However, they are ideal for building detailed partial pathways. For example, the relationship between *BRCA1* and the *RAD50-MRE11-NBS1* complex can be quickly graphed as shown in Figure 4. This gives a basic understanding of the regulators and responsibilities. The

original article abstracts can be obtained in Medscan by clicking on the various picture components such as proteins or actions.

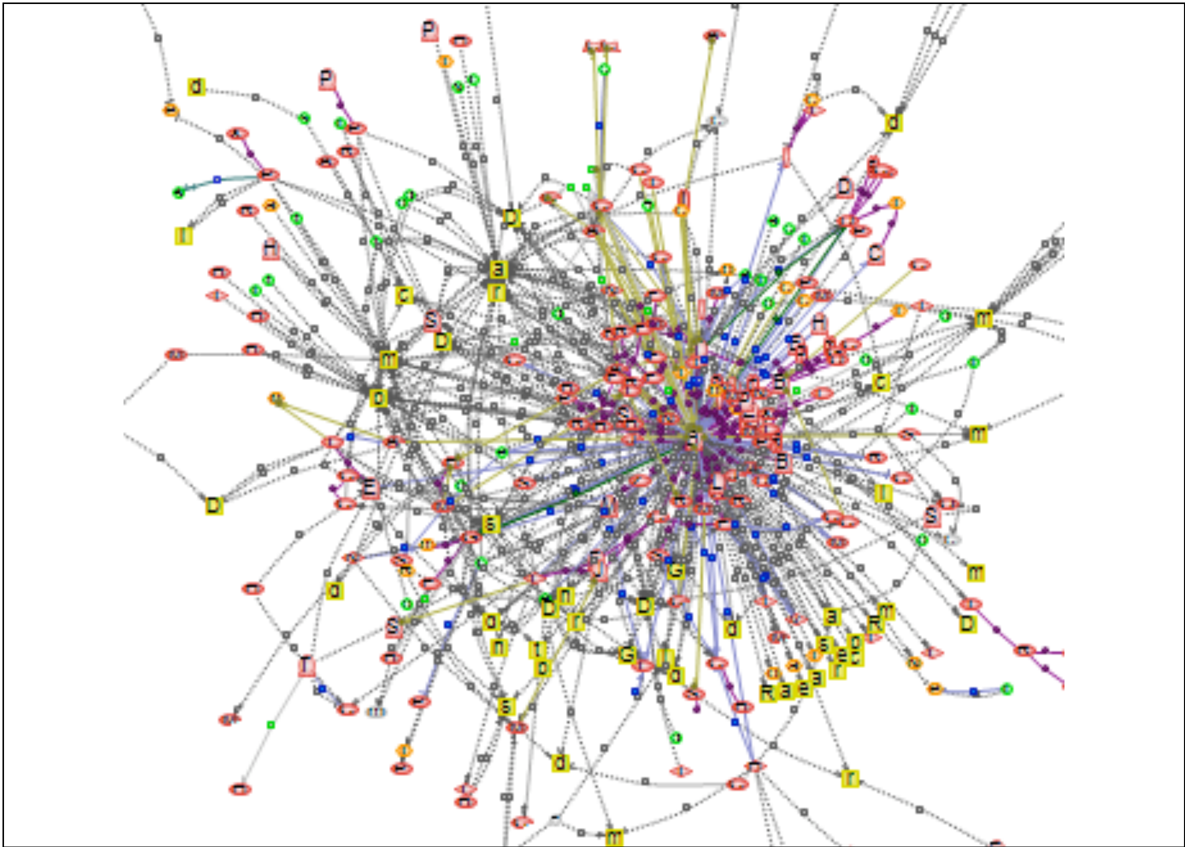

Figure 3 – BRCA1Pathway
Directed graph of 2000 proteins (nodes) with published paper relationships to *BRCA1*. *BRCA1* is centered with surrounding proteins interacting directly with gene. Created using Pathway Studio Plus.

The last tool included in Pathway Studio is Pathway Reference Summary. This tool builds a bibliography page based on reference sources in the current pathway. This file is generated from either the ResNet database or live Medscan results depending on the original search type.

BioCarta and Protein Lounge proved to be useful tools for building the *BRCA1* pathway. Protein Lounge contains one of the largest graphical pathway and protein interaction databases available on the internet. Although the site is a subscription based service they do offer limited trial accounts. BioCarta Pathways is a free service provided by a partnership between the National Cancer Institute and BioCarta Inc.

Both sites contain very similar information and pathway mapping diagrams. Pathways are displayed as an image map accompanied by a large discussion of major interactions. Each protein in the pathway is linked to detailed information including an NCBI Online Mendelian Inheritance in Man (OMIN) style review, sequences, other family proteins, and full publication lists. The Protein Lounge pathways usually contained more proteins per diagram making them appear more complete. These 'extra' proteins are often associated with reactions not directly related to the main pathway.

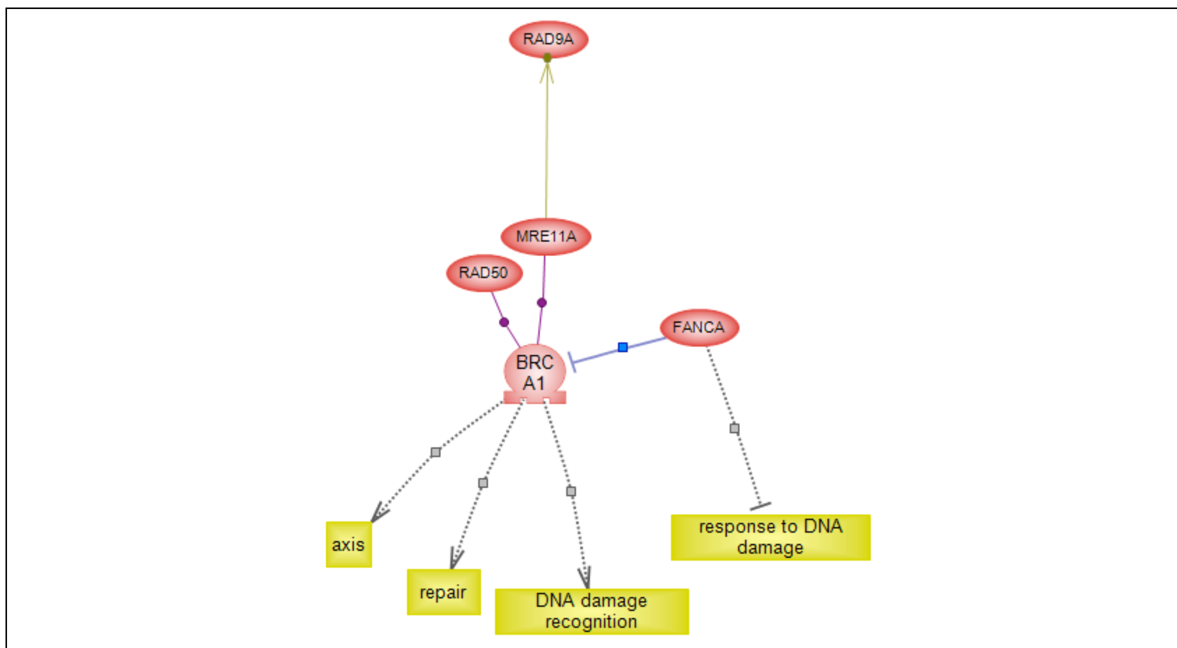

Figure 4 – BRCA1-MRE11/Rad50
A portion of *BRCA1* regulation and function of *MRE11-Rad50* graphed in Pathway Studio. Notice BRCA1 upregulates *MRE11* and blocks the *FANCA* response pathway.

Protein Structure

Sequence and structure data was gathered from sources such as GenBank, Entrez Protein/Structure and the PDB. Similar to before, PubMed articles were used as a reference to isolate bonding sites and special structural features. Swiss-Prot and Entrez Protein contains translated sequences along with publication references. Both sites summarize gene function, disease relations, and domain subunits. Swiss-Prot is more of a one-stop shop for information on genes. In addition to sequence information they also provide 3D structure links, interaction

databases, 2D SDS (gel) samples, and a detailed feature table. The feature table charts the position of two-dimensional conformations such as helices, strands and turns. I utilized this resource to test my predicted protein models.

Entrez Structure is a molecular model database of 3D protein domains. Entrez structure contains two types of models. First each entry is linked to the RCSB (PDB) structure model. RCSB models are formed from predictive computational methods, X-ray crystallography and NMR. RCSB formatted files are an open standard that is widely accepted by a variety of modeling software.

Second, Entrez contains ASN.1 formatted structures stored in MMDB (Molecular Model Database). These are derived from RCSB entries but eliminate hypothetical models, sequence fragments and isoforms. Entrez files are processed specifically to enhance the chemical integrity of models in preparation for computational manipulation. MMDB files are readable in a NCBI utility called Cn3D. Cn3D provides basic rendering functions and the ability to highlight domains. It does not allow modifying the model or the ability to predict structural transforms.

To determine structure & bonding functionality, I observed the C-terminal models of 1JNX and *XRCC1* (1CDZ). 1JNX is the crystal structure of the BRCT repeat region. *XRCC1* is a DNA repair gene containing a domain similar to the *BRCA1* C-terminus. Slight differences between these models were useful for observing structural changes induced by sequences mutations. At the RING domain, I used the 1JM7 model which is a heterodimer formed by *BRCA1* bound to *BARD1*. All three models were viewed in Cn3D and QuickPDB but final rendering was completed in MolScript.

MolScript is a powerful tool used for rendering and manipulating 3D images from PDB files. Although the program is command line based, an OpenGL viewer is included for conducting real-time visual manipulations. The power behind this program results from the ability to highlight important molecules, forces and bond angles. All tweaking and manipulation

must be done via the input file prior to rendering. Table 1 includes a portion of the script used to render *1JM7* PDB files in MolScript.

To render the *BRCA1* RING domain I used the Entrez Structure model 1JM7 - Solution Structure of *BRCA1-BARD1* RING-domain Heterodimer. As no standalone *BRCA1* RING models are available, this was the best choice as it contains mutations-free sequences.

The C-terminal repeats were modeled using 1JNX – Crystal Structure of the BRCT Repeat Region. This model contains both repeats allowing access to the dimer interface where *CtIP* and *BACH1* bind.

While using this program I encountered several problems. The first being it only is available in a precompiled format for IRIX. Machines capable of running this software are outdated and often difficult to locate.

My first model was the RING domain using 1JM7. Molscript automatically renders all models described in the input PDB files. Since 1JM7 is a solution prediction and not an exact crystalline structure, several theoretical models are presented in the PDB database. When rendered, all of the models are displayed concurrently (Figure 5). Variability exists mostly in the loops with well defined helix and sheet positions. To correct this problem I edited the PDB file to only include a single model.

My second issue was some beta sheets were not rendering properly. Since the strands in 1JM7 are usually only 2-4 residues in length, the PDB file had mistakenly labeled them with a sheet identifier. To correct this I edited the PDB file specifically defining beta sheet start and stop positions. Other models such as 1JNX were visually accurate and only required minor tweaking along the beta sheets.

1JNX was an easier model to render since the PDB information is more concrete. The only major problem I encountered was a 'fused' beta sheet. Two beta sheets had fused as the PDB file did not correctly identify turns connecting the sheets. This resulted in three beta sheets instead of four. In this case the PDB file contained correct information but a bug in Molscript

caused a strand skipping. To correct this issue the Molscript input file was changed to realign four parallel beta sheets.

Figure 5 – NMR Predictions of 1JM7
Entrez NMR models of *BRCA1's* RING domain contain moderate variability in the cross-over loops and placement of Zn+ atoms. The above Molscript visualization uses stock input data to illustrate the variability in the input file.

My final step using Molscript highlighted the four major bonding sites in the *RING-BARD* molecule. To do this, I programmed Molscript to select the sidechains of specific backbone residues in the upper helixes. Bond lines were then added between the sidechains of opposing molecules.

Table 1 – MolScript sample input for 1JM7

MolScript renderings are computed from two input files. The first is a script containing the location of a PDB formatted file and the location of areas of interest. Areas of interest may be helices, stands, or individual residues. A portion of the script used to create the rendering in Figure 5 follows.

```
    shadows on;
    window 61.38;
    slab 32.30;

    read mol "1JM7-1.pdb";
    transform atom *
      by centre position atom *;
    set segments 2;

    # Set color and coordinates of first Alpha Helix
    set planecolour hsb 0.6667 1 1;
    coil from A1 to A7;
    set planecolour hsb 0.6275 1 1;
    helix from A7 to A22;

    # Set color and location of sidechain molecules.
    set colourparts off;
    set planecolour white;
    ball-and-stick require in residue A7 and backbone;
    ball-and-stick in residue A10;
    ball-and-stick in residue A85;
    ball-and-stick in residue A96;
```

Structure/Functional Mutations

To begin mutation research another thorough search of literature was undertaken to determine well-know mutations and exons of interest. Individual mutation sites and effects were obtained from the Universal Mutation Database, Cancer Genetics Web, Human Gene Mutation Database and the Breast Cancer Gene Database, in addition to individual journal articles. Using this information modified sequences were submitted through prediction programs then modeled by DeepView, Molscript or Compare3D.

The NIH Breast Cancer Information Core database is based off of GenBank *BRCA1* standard U14680. This database was crucial in locating polymorphisms throughout the gene and also specific protein binding sites. The total number of mutation entries is approximately ten thousand. However, only 1500 unique entries reside in the database with 658 mutations reported more than once. Shockingly, only 82 of all mutations in this database are cited as published in a

periodic journal. This lack of published studies has not diminished the database popularity among researchers.

Although the NIH database is easy to navigate, it only permits simple data searches and does not allow more complex statistical queries. However, a raw data dump of the entire database is available. After downloading the complete dataset, I attempted to load it into a local SQL repository. Unfortunately due to frequent data type violations and syntax mistakes the data did not easily load. Commonly, the wrong data was placed in a column or combined with adjacent data cells. Additionally, mutation designations and position naming conventions differ widely. For example, a single base-pair deletion may be written as del A, delA, or del-A. Similarly intron mutations are named 'intronic', 'Non-coding', I-20, or 'intron/spl'. These variations make data mining difficult without first normalizing the dataset. The variations may in part be due to the broad naming standards available at Human Genome Variation Society (Dunnen, 2000). After some modest hand curation and automated massaging the data was bulk loaded using a custom shell script. My local dataset was taken from June '05.

After working with the dataset for several months, I have some general concerns about the completeness of the database. As with any of the publicly available databases it is the responsibility of researches to submit information. I question the accuracy of statistical numbers derived for high frequency mutations. For example, the frequently occurring 185delAG mutation was reported over fifteen-hundred times.

When compared to other mutations is this an accurate occurrence rate of the mutation in the general population? I hypothesize the submission rate for a non-unique mutation is inversely proportional to the number of existing entries for that mutation. People are more likely to report new scientific finds but not information already discovered by another researcher. A prime example of this is the second most frequent mutations 5382insC. As a mutation with such a high incidence rate, I find it odd it was last reported in 2003.

Once the data was loaded into my local database I was able to launch queries against it. Table 2 outlines some SQL commands used to query the dataset and the composite result.

Table 2 – SQL Queries for BIC Database
Data from the BIC database was queried using standard SQL (Structured Query Language) after loading the dataset into a relational database. (a) Query for finding the number of mutations per exon. (b) Query for determining the number of mutations between codon 1528 and 1862 that are located in exons. (c) Query to determine quantity of each mutation type with results displayed in (d). Mutation type abbreviations (e).

(a)

```
SELECT exon, count(exon) FROM brca1_db GROUP BY exon ORDER
BY exon;
```

(b)

```
SELECT count(*) FROM brca1_db WHERE '1528' <= codon AND codon
<= '1862' AND exon NOT LIKE ‘%I%’;
```

(c)

```
SELECT mutation_type, count(*) FROM brca1_db GROUP BY
mutation_type ORDER BY count DESC;
```

(d)

```
 mutation_type | count
---------------+-------
 F             |  4798
 M             |  2729
 S             |  1138
 N             |  1037
 IVS           |   326
 Syn           |   139
 unknown       |    68
 IFD           |    48
 SYN           |     9
 UTR           |     7
 P             |     4
```

(e)

Abbreviation	Mutation Type
F	Frameshift
IFI	In Frame Insertion
IFD	In Frame Deletion
M	Missense
N	Nonsense
P	Polymorphism
S	Splice
Syn	Synonymous
UV	Unclassified Variant

My first task was to profile the number of mutations in each exon of the gene. This was done simply by counting the number of mutations per exon (Table 2a). Next I queried for mutations in each protein domain listed in Table 6. This query was limited to coding regions by removing introns with the ‘I’ flag (Table 2b). Finally, I queried for the most frequent mutation types and the most frequent mutations. Introns were included in both of these searches (Table 2c).

To search for the most frequent mutation locations in BRCT I trimmed the mutated amino off each mutation. For example, R1699Q and R1699Q were truncated to simply R1699 (Table 4).

Table 3 – Most Frequent Mutations in BIC Database
SQL to determine the top 15 frequently occurring mutations including mutations in introns.

```
SELECT designation, count(*) FROM brca1_db GROUP BY designation
ORDER BY count DESC LIMIT 15;
```

```
  designation   | count
----------------+-------
 185delAG       |  1596
 5382insC       |   836
 C61G           |   155
 IVS5-11T>G     |   152
 R1347G         |   149
 R841W          |   114
 M1008I         |   105
 4184del4       |   105
 R1443X         |   103
 3875del4       |    97
 IVS21-36del510 |    80
 IVS2-14C>T     |    77
 E1250X         |    76
 Q563X          |    70
 M1628T         |    70
```

Table 4 - Frequent Mutations in C-terminus
SQL to determine the top 20 mutations occurring in the *BRCA1* C-terminal region.

```
SELECT exon, designation, count(*) FROM brca1_db WHERE
exon = '16'
OR exon = '17' OR exon = '18'
OR exon = '19' OR exon = '20'
OR exon = '21' OR exon = '22'
OR exon = '23' OR exon = '24'
AND ‘1646’ < codon AND codon < ‘1856’
GROUP BY exon, designation ORDER BY count DESC LIMIT 20;
```

```
 exon | designation | count
------+-------------+-------
 20   | 5382i       |   837
 24   | R1835       |    51
 16   | M1652       |    42
 18   | A1708       |    39
 16   | 5083d       |    38
 19   | 5296        |    30
 20   | R1751       |    29
 21   | M1775       |    23
 18   | R1699       |    20
 22   | G1788       |    19
 20   | G1738       |    15
 17   | 5149d       |    15
 18   | Y1703       |    14
 18   | 5256d       |    14
 24   | W1837       |    13
```

Since the BIC site does not provide tallying tools, it was relatively easy using SQL to find the population with the most mutations (Table 5). Some of the problems with non-standard data submission are visible in the table on the left. The table on the right shows the data after a cosmetic review. Note the first line is blank yet has a 3616 total. These are entries without an ethnicity defined. Also notice the multiple occurrences of Western-European, Western Europe, and Western European.

Table 5 – Populations with High BRCA1 Mutations
Mutation count varies among population groups. Since the BIC database does not standardized naming conventions, multiple ethnic groups often appear in results. Notice the duplication of Western European and Not Specified in the left table. The table on right was hand modified to contain only one row per population.

```
SELECT ethnicity, count(*) FROM brca1_db GROUP BY ethnicity
ORDER BY count DESC LIMIT 11;
```

ethnicity	count
	3616
Ashkenazi	1171
Western-European	1098
Caucasian	671
None-Specified	586
Western Europe	494
Not Specified	267
Western European	265
None Specified	243
African	157
Central/Eastern-European	110

(11 rows)

ethnicity	count
None Specified	4540
Ashkenazi	1334
Western-European	2299
Caucasian	213
African	159
Central/Eastern European	246
Latin-American/Caribbean	173
Greek	86
Asian	104
African American	77

(10 rows)

Mutation effects on structure were studied using a combination of predictive modeling tools. Mutations at bonding sites and those mentioned in the literature were submitted into these forecasting programs. The first program I used was called Protein Predict.

Protein Predict (PHD) is web-based tool for predicting secondary structure from nucleotide or amino acid sequences. The power of this program comes from combining many separate prediction programs and algorithms. Each program analyzes a different portion of the protein structure such as amino-acid chain interactions, solvent accessibility or sequence alignment. Each of the results is scored with a confidence level identifying the accuracy of the prediction. Users can then assess the most reliable prediction method for the protein sequence.

Output from Protein Predict is presented in a graphically formatted HTML display of alignments, tabular charges, and probable structures.

Protein Prediction appears to make better predictions for sequences with multiple mutations rather than single nucleotide polymorphisms. I suspect this is due to SNP sequences sharing significant homology to a single known sequence. When multiple mutations are introduced, the prediction algorithm is able to calculate results from many partial models rather than a single input.

For my research PROsec and PROacc from PHD were the most useful and accurate for analyzing novel mutations in the RING and BRCT domains. PROsec focuses on secondary structure prediction using multiple sequence alignment. PROacc predicts solvent accessibility of residues using a neural network of 238 standardized proteins. Selective output from Protein Predict allowed submission to third party sites such as Expasy Swiss-Model.

Swiss-Model is a protein modeling service provided free by Swiss Institute of Bioinformatics. Swiss-Model works by first using BlastP2 to search for similar target sequences. Sequences with identity above 25% are superposed on homologous areas of known models. Unknown or mutated areas are then modeled by using known Phi and Psi angles from a database of known fragments. The modeled structures are then checked against another database of known sidechain rotamers. The final model is verified using the Luthy method (Luthy, 1992) for assessing protein model quality.

Swiss-Model generates a PDB file based on either raw sequences or Swiss-Prot accession numbers. I found the models generated were generally accurate predictions of both actual PDB entries and documented conformation changes.

Conservation

To find conservation between both human protein samples and other species PSI-BLAST, RPS-BLAST, ExPASy and CDART were used to find sequence variants. All of these

tools accept standard FASTA formatted files. Selective sequences from these programs were combined for comparison in ClustalW or ScanPS (protein).

ClustalW is a multiple sequence alignment tool for protein or DNA. Window, gap, and matrix (BLOSUM vs PAM) alignment are adjustable among other parameters. Output from ClustalW is given in a plain text format to be read by JalView. JalView (Java Alignment editor) as the name implies is a program for viewing the alignment output from ClustalW (Figure 6).

As each of these programs is hosted at a different site, a moderate amount of work is required to search, align and edit the results. However, the effort is worth the ability to visualize and manipulate the alignment. Sequences can be quickly colored along BLOSUM score, percent identity, strand/turn/helix or other scoring methods.

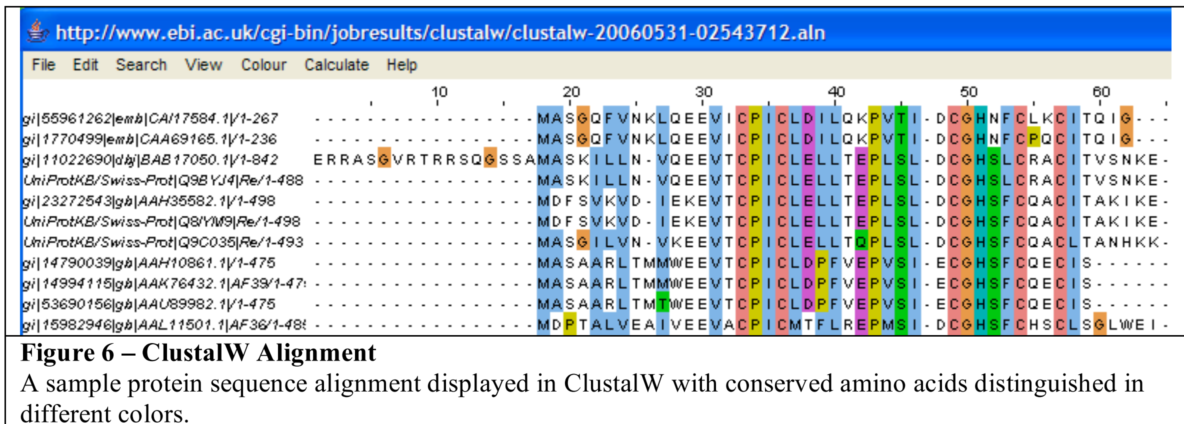

Figure 6 – ClustalW Alignment
A sample protein sequence alignment displayed in ClustalW with conserved amino acids distinguished in different colors.

CDART (Conserved Domain Architecture tool) returns sequence and structure/function homology matches. The database backing CDART is based on alignments of partial and full-length protein domains. CDART uses RPS-BLAST engine to find matches but then compares pre-computed scores on the domain to determine a final match.

Conservation in Other Species

Finally, I explored sequence and function homology in other species such as mouse chromosome 11 and Drosophila using similar BLAST methods and tools as outlined above. In addition to the typical homology search tools I found MGI (Mouse Genome Informatics) and BLink very useful.

The MGI website was an invaluable tool for understanding mouse genomics and the relationship to the human gene counterpart. Searching for an individual gene reveals a sequence summary report. For mouse, six potential sequence matches were listed, four correlating to *BRCA1*. Each gene detail report is a one-stop shop for all information associated with the mouse gene. Of particular use were the genetic map, sequence map, protein domain and mammalian homology sections.

In addition to the NCBI MapViewer, the MGCI Sequence Map highlights the Mouse *BRCA1* gene location in four different viewers - Vega Mouse, Ensembl, and the UCSC genome browser. All of these tools were very good at detailing the position of the gene in relation to other genes on the chromosome.

The mammalian orthology section of MGI contains direct links to the Entrez Gene reference sequences used to map homology. Mouse *BRCA1* had three verified homologous domains with human, rat and domesticated dog DNA.

The other tool to hunt for related sequence data was BLink (BLAST Link) at NCBI. Each Entrez Protein stored in NCBI has an associated BLink entry. These entries contain graphical lists of previously recorded BLAST searches. Similar to blast results the list contains a score, p-value along with accession and organism name. They are somewhat easier to read than BLAST results since the FASTA output is omitted.

Gene Function

DNA mutations frequently leading to cancerous cells are caused by ionizing or ultraviolet radiation, environmental chemicals, or cellular respiration products. Cells, but particularly genomes have adapted mechanisms to specifically deal with these types of alterations. As the cell cycles through its normal metabolic pathway, checkpoints and DNA repair mechanisms are constantly at work. Some of these mechanisms are so important they are literally integrated into the cycle, not just 'complementary' side actions. Repair pathways must not only sense damage signals, but the type of damage, timing of cell cycle, and location of damage.

Proteins involved in the damage response pathway are grouped into sensors, transducers and effectors. *BRCA1* participates in a global role in DNA-damage response, though the process in its entirety is still unknown. It contains key effectors regulating damage-induced arrest in all cell phases. Its role in meiosis is associated with double-strand break repair, homologous recombination (Chen, 1999), and nucleotide excision repair. More specifically, it regulates critical phase transitions by controlling expression, phosphorylation and cellular localization of cyclin-dependent kinases among other proteins (Yarden, 2002). Through these mechanisms it reduces the sensitivity of cells to double-stranded DNA breaks generated by exposure to infrared and gamma radiation.

BRCA1 is found within a "large multisubunit protein complex of tumor suppressors, DNA damage sensors and signal transducers" deemed BASC (*BRCA1* associated genome surveillance complex) (Wang, 2000). Proteins of this complex are crucial for recognizing DNA damage, abnormal DNA and triggering repair. *BRCA1* interactions occur with over 25 different proteins with major reactions occurring with *NBS1* (Nibrin), *RAD50-MRE11* (Meiotic Recombination-11), and *ATR*/*ATM* (Ataxia-Telangiectasia) and *BARD1* (BRCA1-Associated Ring Domain-1). [See Table 6 for a complete list of these proteins]. *ATM* and *ATR* are considered both sensors and transducers. *Chk1* and *Chk2* are transducers while *Cdc25* and other CDKs (cyclins) are effectors (Lee, 178).

BRCA1 is a nuclear phosphoprotein associated with the ubiquination of RNA polymerase II, *FANCD1* and other kinases. It is implicated in the transcriptional regulation of DNA damage-inducible genes that function in cell cycle arrest. *BRCA1* is located predominantly in the nucleus and perinuclear compartment of endoplasmic reticulum complex (Paull, 2001) with the greatest levels found during S and M phases (Chen, 1999).

Although the exact protein function has not been established it plays an active role in transcription regulation, chromatin remodeling, mRNA processing, cell cycle checkpointing and apoptosis. Most importantly *BRCA1* controls checkpointing between G2 to M phase of the cell

cycle (Figure 7). During this period its two responsibilities are participating in detecting damaged DNA and responding and repairing it. (ProteinLounge, 2006).

Table 6 – BRCA1 Related Proteins
Over 25 proteins interact directly with *BRCA1* at one or more binding sites. Specific binding sites are not known for all proteins but general regions have been identified. *BRCA1* polymorphisms are found in uniform frequency across the entire genomic region. Binding data compiled from the Biomolecular Interaction Network Database (BIND) and mutation information from NCBI Breast Cancer Information Core (BIC).

Protein	BRCA1 Codon	Binding Site	BIC mutations	Citation
CDC2	1-76		1981	(Wang, 2000)
CDK2	1-76			(Wang, 2000)
BARD1	1-103	Arg7, Glu10, Glu85, Asp96	1981	(Brzovic, 2001)
Zn1+	1-103	Cys24, ILE26, Cys27, Cys44, Cys47		(Brzovic, 2001)
Zn1+	1-103	Cys39, His41, Cys61, Leu63, Cys64		(Brzovic, 2001)
ER-alpha	1-306	Ala1708, Met1775, Tyr1853	2249	(Kawai, 2002)
p53	224-500	Exon 11	776	(Zhang, 1998)
p53	1760-1863	Unidentified sites in BRCT		(Chai, 1999)
Rad50	341-758		1162	(Zhong, 1999)
ATR	521-757	Ser1423, Ser1457, Ser1524	582	(Gatei, 2000)
ATM	751-1070			(Gatei, 2000)
ATM	1241-1530	Ser1189, Ser1387, Ser1423, Ser1524		(Gatei, 2000)
FANCA	740-1083		1269	(Folias, 2002)
Rad51	758-1064	Exon 11	1221	(Scully, 1997)
CDK2-A/E	1497	Ser1497		(Ruffner, 1999)
BACH1	1429-1863	Ser1655, Gly1656, Lys1702	226	(Cantor, 2001)
CK2beta	1460-1653		542	(O'Brien, 1999)
TRAP220	1528-1863	Arg1708, Pro1749, Met1775	1945	(Wada, 2004)
CtIP	1603-1863	Ser1655, Gly1656, Lys1702	1772	(Wong, 1998)

The proteins produced by *BRCA1* bind strongly to DNA allowing it to promote the activities of the *MRE11*/*RAD50*/*NBS1* complex that is used for double-strand break repair. Loss of this ability contributes to “profound chromosomal” instability (Mallery, 2002), spontaneous DNA breaks and ultimately tumorigenesis(Fan, 1999). *BRCA1* interacts with several other genes to carry out these vital functions including *ATR*, *Chk2*, and *RAD51*.

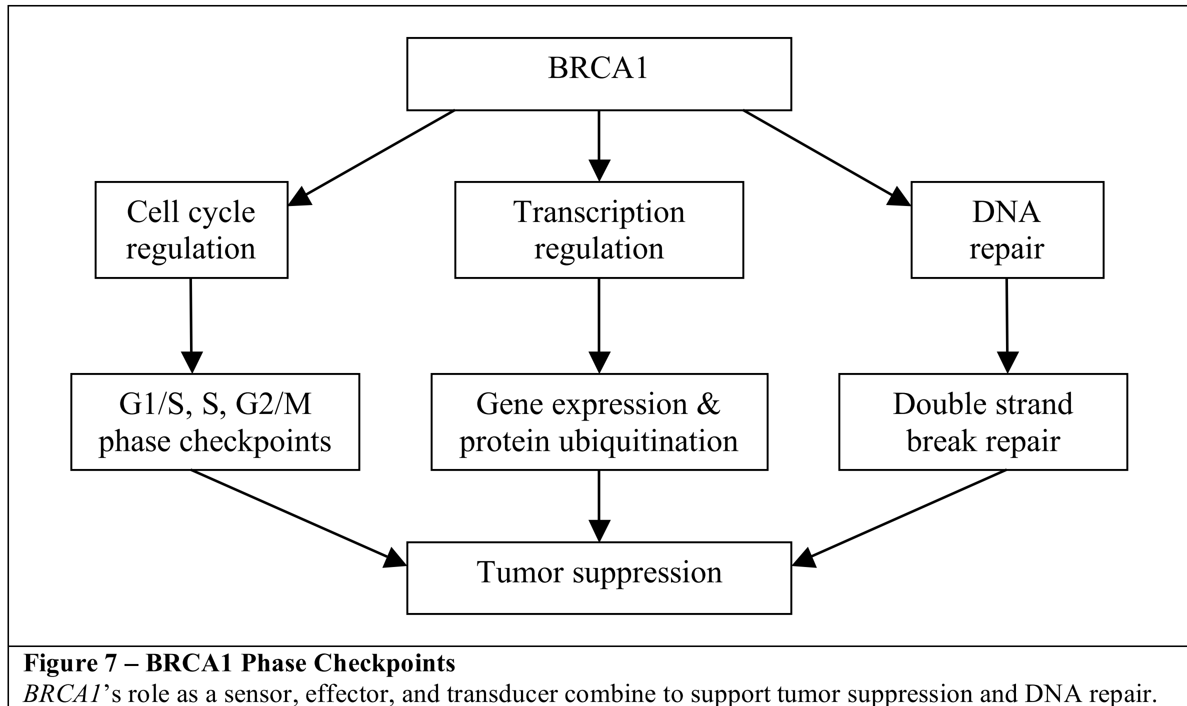

Figure 7 – BRCA1 Phase Checkpoints
BRCA1's role as a sensor, effector, and transducer combine to support tumor suppression and DNA repair.

G1 Phase

G1 phase is marked by cell growth and preparation for chromosomal replication. Damaged DNA evading this period may severely affect future cell processes and possibly be inherited by other cells leading to cancers. During this phase, DNA is checked repeatedly for mutations in an effort to maintain genetic integrity. This paper will focus explicitly on detection and repair mechanisms of double strand breaks caused by IR and UV radiation.

Double stranded breaks (DSB) are the most cytotoxic lesion (Protein Lounge, 2006) leaving both backbone strands disconnected. DSBs are repaired in both pre and post-replication stages using either direct joining or homologous recombination (HR) methods. Direct joining ligates exposed ends with little matching logic so massive translocation errors often result. HR on the other hand uses information intact on sister chromatids, homologous chromosomes or the same chromosome to correctly pair disjoined segments.

Genotoxic stress from double-strand breaks trigger a cascade of protein reactions controlled by the *BRCA1* gene. *BRCA1* damage and repair pathways are invoked through initial stimulation of *ATM* and *ATR* proteins. In response to UV (ultraviolet) and HU (hydroxyurea),

two agents interfering with replication, replication proteins recruit *ATR* to damage sites (Zou, 2003). Emergence from G1 phase is dependent upon the G1/S checkpoint invoked by *ATM*.

Double strand breaks are detected as the cell emerges from G1 phase through phosphatidylinositol kinase proteins (*ATM* / *ATR*) and cycle arrest occurs by blocking cyclin-dependent kinases. *ATM* and *ATR* present at the replication fork are activated upon DNA damage resulting in phosphorylation of numerous substrates (Liu, 2000). *ATM* and *ATR* are involved in multiple parallel but redundant pathways responding to different DNA stresses (Andreassen, 2004) such as double strand breaks or replication forks (Figure 8). One pathway involves the detection of DNA damage and cell cycle arrest while others are responsible for repair.

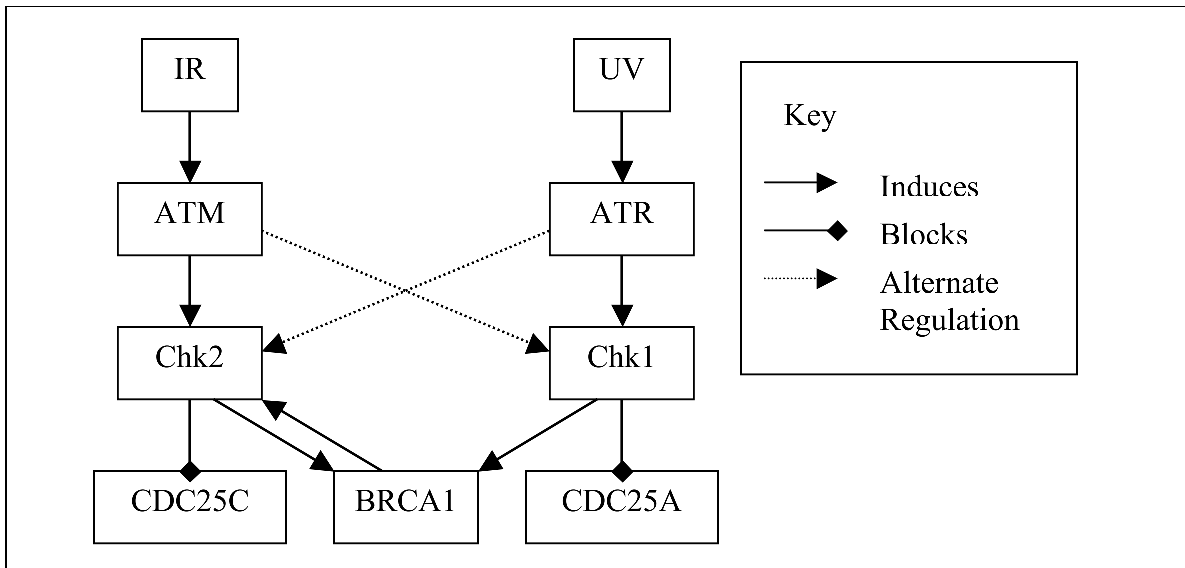

Figure 8 – ATM-ATR Damage Response Pathway
In response to Ultra-Violet or Ionizing Radiation, *ATM* and *ATR* function in parallel pathways to up-regulate a damage-response pathway. Cyclin-dependent kinases, *Chk1* & *Chk2,* are phosphorylated by *ATM/ATR* after exposure to DNA damaging elements. These genes operate in a redundant manner to ensure both cell cycle checkpoints and repair mechanisms are synchronized.

In response to ultraviolet or ionizing radiation damage, *Chk1* phosphorylates *p53* (Banin, 1998) in an *ATM* dependent manor. *p53* is a protein transcriptor mediating other damage-response proteins the damage response (Coene, 1997). *p53* phosphorylates activator *p21* causing a G1 arrest by inhibiting transcriptionally regulated cyline dependent kinases *CDK2* / cyclin-E and *CDK4* / cyclin-D (Waldman, 1995). The stalled replication fork is stabilized by additional

p53 phosphorylation through *Chk2* (Chehab, 2000). Higher levels of damage trigger immediate apoptosis through *ATM* blocking *p53* activity and directly phosphorylating *BRCA1*-*cAbl*. If necessary, c-Abl induces cell death by activating or blocking cell survival proteins *p53*, *JNK*/*SAPK*, or *p73* (Sordet, 2003). *ATM*-*p53* induced apoptosis may also occur at any point in the cell cycle.

Under normal conditions, *BRCA1* C-terminal region binding regulates tyrosine kinase activity of *c-Abl*. IR conditions invoke *ATM* phosphorylation of the BRCT region causing disassociation of this complex resulting in increased tyrosine activity (Foray, 2002). The cleaved *BRCA1* surface makes room for binding of other proteins such as *BACH1* and *CtIP*.

S Phase

In preparation for repair, dramatic increases in *BRCA1* expression at the onset of S-phase are accompanied by the accumulation of *BARD1* polypeptides into nuclear dots in the endoplasmic reticulum (Jin, 1997). *BARD1* (*BRCA1*-Associated RING domain-1) is a protein sharing homology with both C and N-terminal regions of *BRCA1*. It's effectiveness at maintaining genomic integrity through *Rad51* mediated single-strand exchange and homologous paring (Jin, 1997) classify this protein as a tumor suppressor (Wu, 1996). A *BARD1* deficiency results in uncontrolled cell growth whereas overexpression results in cell death (Irminger-Finer, 2001). Typically found during S phase but not during G1 (Jin, 1997) its product peaks with the onset of DNA synthesis. Low levels of *BRCA1*-*BARD1* complex during G1 phase act a *p53* adaptor for *ATM* phosphorylation (Fabbro, 2004). Hyperphosphorylated *BARD1* predominate during M-phase / mitosis (Choudhury, 2005) for nuclear retention of *BRCA1*, controlling chromosomal euploidy and homology-directed repair. Additionally, it is another *p53* mediator of apoptosis (Wu, 1996).

After the cell is arrested, the DNA repair pathway is put into action. At the same time repair is occurring another parallel pathway will continue regulating the cell cycle. These

pathways are kicked off by *ATM* or *ATR* phosphorylation of at least six proteins including *BRCA1*, *BRCA2*, *Rad51*, *NBS1*, *Chk1*, and *Chk2*.

ATR induced phosphorylation of *BRCA1* is considered the beginning of the repair pathway. Phosphorylation at this point forks separate pathways for phase control and repair. DSB activation of *ATM* triggering *p53* arrest begins the two additional parallel pathways for S-phase checkpoint delay eventually intersecting at *FANCD2* (Falck, 2002).

ATM (or *ATR* (Pichierri, 1995)) phosphorylizes *NBS1* to produce an S-phase checkpoint delay (Xu, 2001). Failure to invoke the S-phase checkpoint will result in uncontrolled DNA synthesis (Xu, 2001). *NBS1* later forms with *MRE11*-*Rad50* heterotetramer to control the repair process and regulates phosphorylation of *FANCD2*.

G2 phase

BRCA1 is required for both S-phase arrest and G2/M phase checkpoints (Xu, 2001). G1 activated *Chk1* and *Chk2* target different *BRCA1* molecules for control of G2 phase checkpoints through *ATM* and *ATR* respectively (Yu, 2004). Although *Chk1* and *Chk2* are responsible for G2/M arrest through different regulators, both mediate the same cyclin-dependent kinases (Cyclin-B and *Cdc25C*). *BACH1* (*BRIP1*) and *CtIP* (*CtBP* interacting protein) are the two proteins controlling these pauses which bind at the same *BRCA1* C-terminal region but are managed differently.

The *BRCA1* BRCT domain directly interacts with phosphorylated BRCA1-associated carboxyl-terminal helicase (*BACH1*) (Yu, 2003). *BACH1* forms a heterodimer functioning as a transcription activator or repressor along with the transcriptional corepressor *CtIP* (Varma, 2005). *BRCA1*-*BACH1* interaction is cell cycle regulated with the association persisting from S through M phase and required for DNA damage-induced checkpoint control (Yu, 2003). *BACH1* plays a catalytic role in double strand break repair through it's mediation of *ATR* (Cantor, 2001). In response to IR and other stimuli, *ATR* rapidly phosphorylates *Chk2* stimulating *Chk2* kinase activity (downstream signal transmission) (Huberman, 2005). *Chk2* modulates *BRCA1* function

in response to DNA damage through phosphorylation serine-988 (Lee, 2002). Release from this serine site allows cell cycle restoration. Subsequent hyper-phosphorylation of *BACH1* by *ATR* and *Chk2* promotes homologous & non-homologous recombination (Choudhury, 2005) leading to a G2 "accumulation checkpoint". Prolonging G2 phase postpones mitosis while the cell DNA is still damaged (Figure 9).

While the *BRCA1*-*BACH1* association persists from S to M phase, *BRCA1*-*CtIP* is far less stable only existing during G2 phase (Yu, 2004). *CtIP* is another C-terminal interacting protein which binds in place of *c-Abl* at the BRCT repeats. *CtIP* is phosphorylation-dependent for binding to BRCT domains during early G2 (Yu, 2004).

CtIP is phosphorylated again in late G2 phase while bound to the *BRCA1* repeats. Direct phosphorylation by *ATM* causes a hyper-phosphorylated *CtIP* to disassociate from *BRCA1*. *ATM* phosphorylates *CtIP* at Serines 664 and 745 (Li, 2000). Release from *CtIP* allows *BRCA1* to activate *Chk1* to induce the G2/M transitional checkpoint (Yu, 2004). Mutations in the *CtIP* binding region have significant implications such as repression of *BRCA1* induction of *GADD45*.

BRCA1 activation of *GADD45* in conjunction with *p53*^{p} from G1 phase support transcriptional regulation behavior (Jin, 2000). *GADD45* is targeted by both *BRCA1* and *p53*^{p} through a complex method not fully understood. However, it is believed *BRCA1* represses *GADD45* through interaction with *ZBRK1* (Fan, 2002) and various other transcription factors. Although *p53* is not required for *GADD45* activation, it factors into the strength of the response.

CtIP is an upstream repressor of *GADD45* expression through a *BRCA1* dependent manor (Figure 9). Release of *CtIP* from *BRCA1* allows *BRCA1* to also regulate *GADD45* (Fan, 2002). *GADD45* (Growth-Arrest and DNA-damage) gene is considered a DNA-damage response gene because of its damage induced up-regulation and ability to arrest cell cycle (Li, 2000). *GADD45's* main function is to cause G2/M arrest thereby suppressing cell division. *GADD45* may also play an indirect role in *BRCA1* induced apoptosis via *JNK*/*SAPK* activation (Scheikh, 2000). The G2/M checkpoint delays transitioning into mitosis allowing time for DNA repair.

Failure to stop at this point allows progression of chromosomal rearrangement with damaged DNA resulting in lost genomic integrity.

Approaching the G2/M checkpoint *Cdk1* (*Cdc2*) levels increase steadily during S phase (Jin, 2000). CDK (cyclin dependent kinase) promoters control phase progression. Blockage or inhibition of these proteins stalls phase transitions otherwise known as checkpoints (Jin, 2000). *GADD45* inhibits entry into M-phase through *Cdc2* (*Cdk1*)/ cyclin-B inactivation until repair and transcription are complete. *GADD45* modulates this checkpoint by blocking the Cdc2-cyclin-B kinase complex (Yang, 2000). *Cdc2* is a protein phosphatase initiating DNA synthesis and mitosis. Expression, phosphorylation and cell localization of *Cdc2* & *Cdc25C* are controlled explicitly by *BRCA1* in G2 phase. *GADD45* does not phosphorylate *Cdc2* but instead binds directly to it.

This checkpoint works in conjunction with the parallel *CtIP*-*Chk1* induced control of the same *Cdc25C* pathway. Prior to G2, the *Wee1* kinase adds an inhibitory phosphate to Cdk1 rendering the kinase inactive (Harvey, 2005). Upon release from *GADD45*, Cdc2 removes this phosphate. Subsequent Chk1 phosphorylation of *Cdc25C* promotes the binding of 14-3-3 regulatory proteins (Brown, 1999).

Chk2 has been implicated as a backup phosphorylation agent of *Cdc25C* in the absence of *Chk1* (Liu, 2000). Both *Chk1* and *Chk2* block *Cdc25C* in similar manors and functions controlling cell cycle checkpointing, transduction of DNA damage, and replication stress signals (Brown, 2002). Significant overlap, coordination, and redundancy are exhibited in this pathway to ensure DNA integrity.

GADD45's other behavior is to stimulate DNA excision repair through *Rad51* (Smith, 1994). Upregulation of *RAD51* promotes DNA strand exchange during homology-directed repair (HDR) ultimately leading to chromosomal integrity (Foray, 2002). *BRCA1* and *BRCA2* also participate in direct activation of *Rad51*. Initial *ATM* phosphorylation of *BRCA1* causes its

nuclear foci to disperse to *Rad51* foci. In this way, *c-Abl* association with *BRCA1* may indirectly regulate DNA repair proteins.

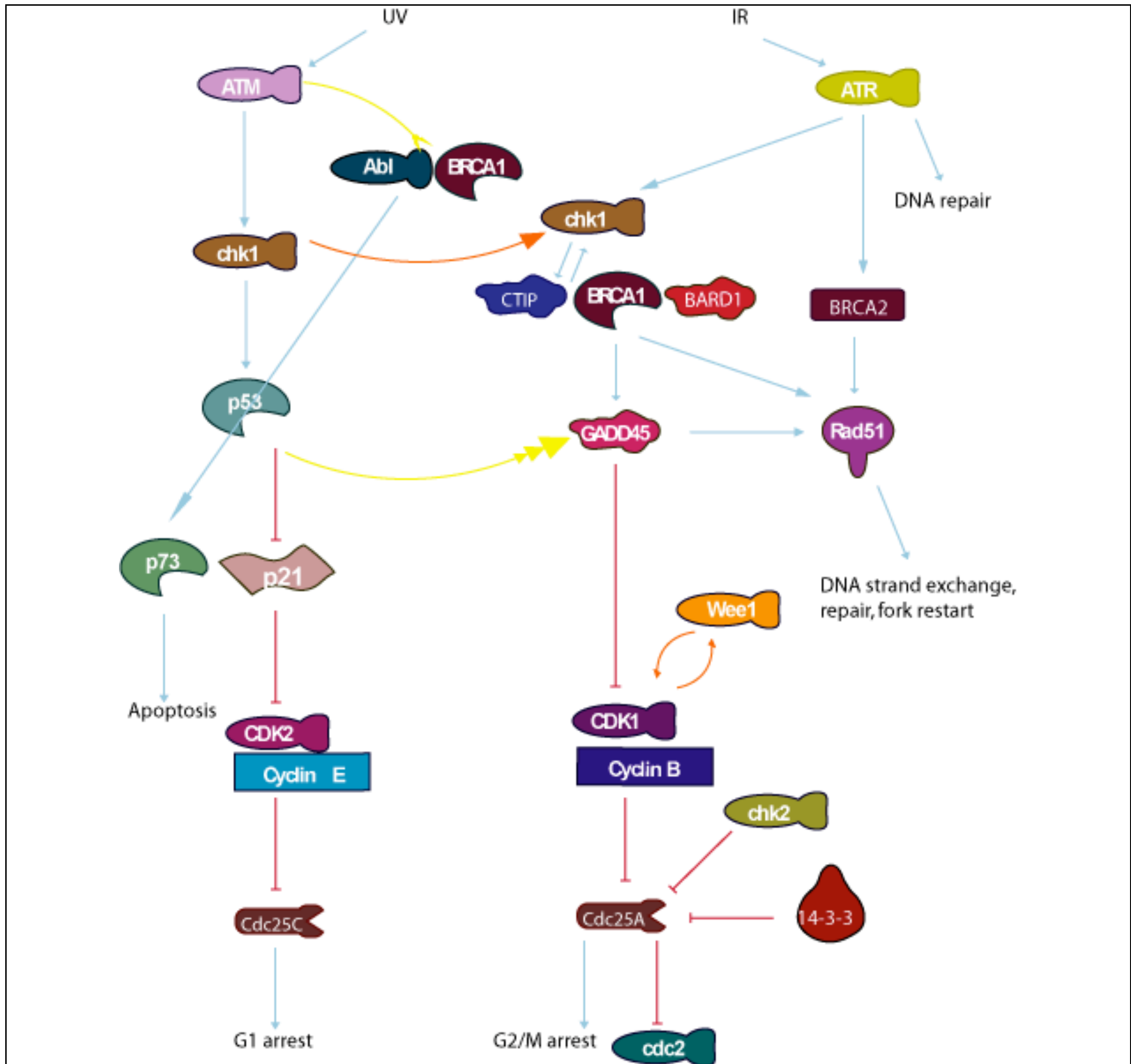

Figure 9 – BRCA1 Pathway
Ultraviolet and Ionizing Radiation begin the *BRCA1* checkpoint and repair pathways. G1 and G2/M arrest are dependent upon *BRCA1* disassociation from *c-Abl* and *CtIP*. Cyclin-dependent kinases are regulated by upstream proteins p53, *GADD4*, and *Chk1*. Figure created using BioCarta Pathway templates and Pathway Assist.

Compared to *BRCA1* little is known about the second gene, *BRCA2*, which lies on chromosome 13 (Wooster, 1995). The gene's exact function has not been uncovered but its interaction with *RAD51* reveals it too participates in homologous recombination (Howlett, 2002).

Following phosphorylation by *ATR*, *BRCA2* may inhibit *Rad51* to control double-stranded repair (Chen, 1999). In no way does *BRCA2* interact directly with *BRCA1* (Kato, 2000).

G2 repair

Homologous repair takes place in late S and early G2 phase when an undamaged sister chromatid is available for use as a template. The *NBS1* S-phase checkpoint response functions at the intersection of two repair signaling pathways. The first pathway induces S-phase arrest while the second promotes double-stranded break repairs (end joining) through *FANCD2* (Wilda, 2000). Phosphorylation of *NBS* by *ATR* in response to ionizing radiation begins these repair pathways (Figure 11).

Upon ionizing radiation the *MRE11* complex assembles in nuclear foci with *BRCA1* (D'Andrea, 2003). *ATR* phosphorylates *NBS1* in response to this ionizing radiation (Falck, 2002) causing association with co-located *MRE11*-*Rad50* (Figure 10). The complete *NBS1*-*Rad50*-*MRE11* (MRN) complex is a sensor of abnormal DNA structures and a regulator of post replication repair process (Wang, 2002). Specifically, *MRE11*-*RAD50*-*NBS1* acts as a double-strand break sensor for *ATM* and recruits *ATM* to broken DNA molecules (Lee, 2005). *BRCA1* mediates this reaction by acting as a *Rad50* adaptor along exon 11 (Zhong, 1999). *MRE11* then regulates exonuclease resection of 5-prime double-strand ends and is involved in additional checkpoint signaling (Huberman, 2005).

NBS1's second repair role is required for activation of the *FANCD2* repair associated protein (Pichierri, 1995). *FANCD2* belongs to a group of nine proteins FA (Fanconi Anemia) coupled with the *BRCA1* repair pathway. FA encoded proteins are essential for protection against chromosome breakage detection (Garcia-Higuera, 2005). Though FANC encoded proteins share little homology they cooperate in a serial pathway essential for regulating the G2-checkpoint and DNA repair. Six of these proteins (A, C, E, F, G, and L) assemble in a nuclear complex that is required for activation of *FANCD2* (Garcia-Higuera, 2005).

During S-phase *FANCD2* is monoubiquitinated by *ATR* in response to IR and is targeted to nuclear foci (dots). It is believed *NBS1* acts an adaptor for *ATR*-mediated phosphorylation of *FANCD2* (Pichierri, 1995). Once activated it assembles with *BRCA1*, *BRCA2*, *NBS1* and *Rad51* in the perinuclear compartment of the endoplasmic reticulum (Coene, 1997). Monoubiquitination by *BRCA1-BARD1*^{ub} is a G2 accumulation checkpoint response leading to genomic stability and promotion of homologous recombination. Although *ATM* is not necessary for the FA pathway it does increase the efficiency in IR conditions (Taniguchi, 2002).

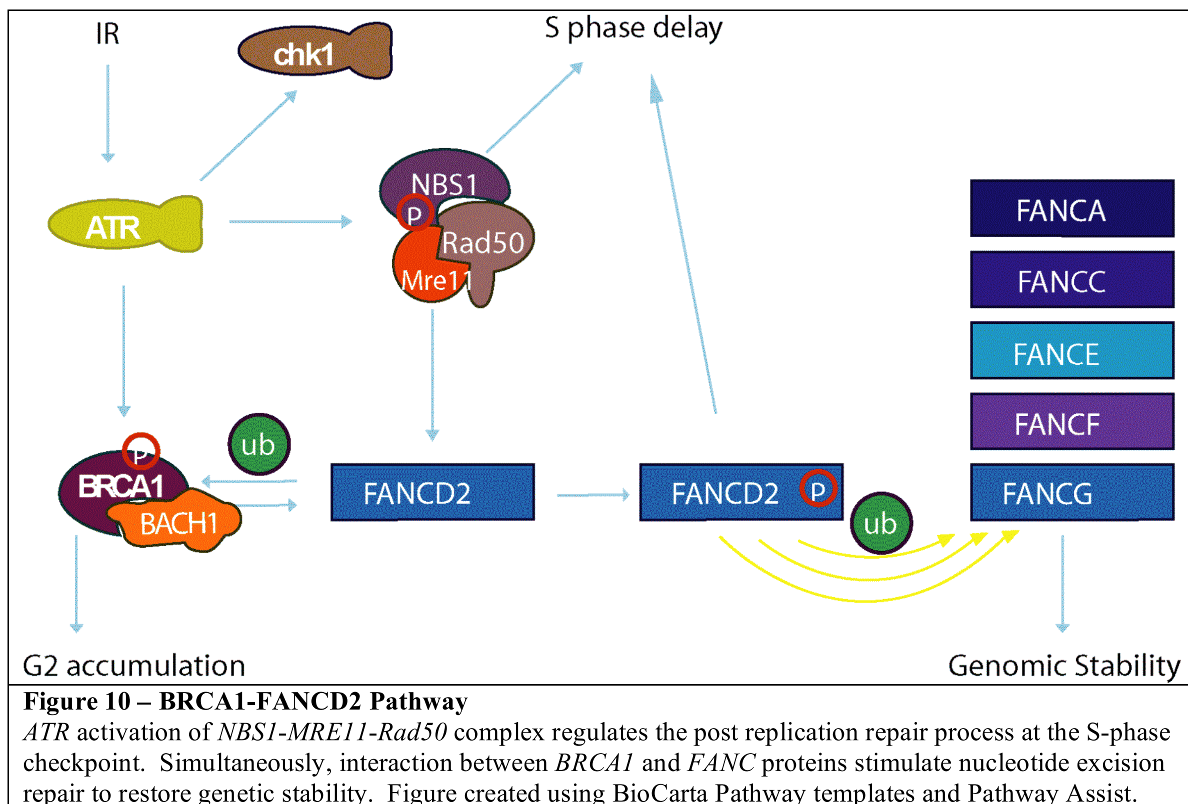

Figure 10 – BRCA1-FANCD2 Pathway
ATR activation of *NBS1-MRE11-Rad50* complex regulates the post replication repair process at the S-phase checkpoint. Simultaneously, interaction between *BRCA1* and *FANC* proteins stimulate nucleotide excision repair to restore genetic stability. Figure created using BioCarta Pathway templates and Pathway Assist.

FANCA interaction occurs in the central region of *BRCA1* at aminos 740-1083. This leads to *FANCD2* phosphorylation and eventually the DNA replication fork restart (Folias, 2002). The outcome of this response is either genomic stability or cellular genotoxicity (apoptosis) in non-repairable conditions. In this region Lys820Glu and Gly960Asp are two frequent causing cancer mutations though neither affects binding (Folias, 2002).

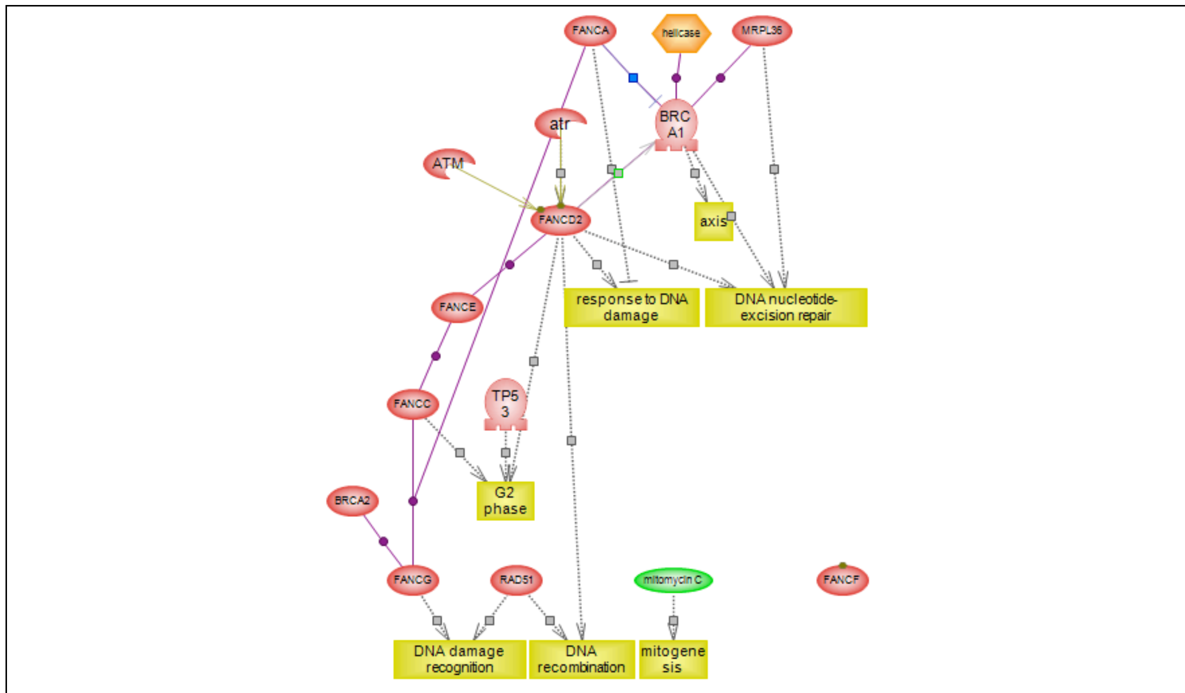

Figure 11 – BRCA1-FANCD2 Pathway
This figure is a representation of the FANCD2 pathway from data correlated Pathway Assist. Pathway Assist identifies protein and cell process relationships from PubMed literature searches.

M phase

BRCA1 acts in multiple pathways, checkpoints and effectors. The onset of mitosis is a strictly controlled process dependent upon *ATM*, *CHK* and *BRCA1* among other proteins. The replication fork restart is managed mostly by proteins associated with the BRCT repeats but also controlled by *FANCD2* and *MRE11-Rad50*.

Structure of BRCA1 Proteins

The *BRCA1* gene contains 24 exons (22 coding) spanning 81k base pairs encoding an 1863-residue multifunction protein. Exon 1 contains two alternative non-coding exons which are duplicates of exon 1 and 3 of the adjacent *NBR1* gene (Puget, 2002). Exon 4 is usually non-coding unless an Alu mutation causes a premature stop codon. The other non-coding introns range between 400bp-9kb in size. *BRCA1* contains three main binding regions, one at the (acidic) carboxyl-terminal region (BRCT), another at the N-terminal (RING) and a vast central region. The C-terminal domain, deemed BRCT, binds with several proteins involved in maintaining

genomic integrity (Yarden, 2002). Protein-protein interactions are regulated by a zinc finger located at the N-terminal side. Figure 12 shows the different binding regions of *BRCA1*.

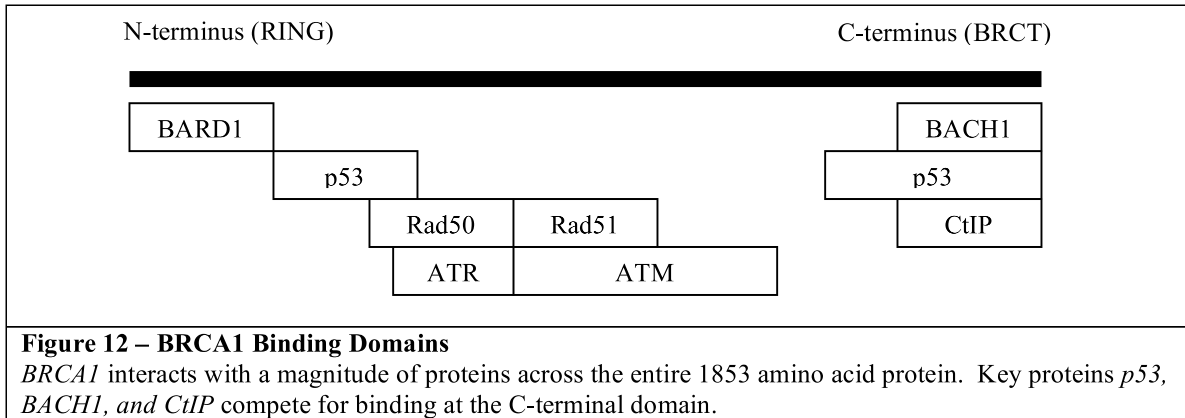

Figure 12 – BRCA1 Binding Domains
BRCA1 interacts with a magnitude of proteins across the entire 1853 amino acid protein. Key proteins *p53, BACH1, and CtIP* compete for binding at the C-terminal domain.

BRCT – BRCA1 C-terminal

BRCA1 houses two tandem repeats of BRCT motifs at the C-terminal regions. *BRCA1* belongs to a family of more than 23 BRCT motif-containing proteins (Callebaut, 1997). These motifs are known to be involved in tumor suppression through DNA damage detection and repair activities (Yu, Xin.). BRCT is a phosphoprotein binding domain meaning interactions occur in a phosphorylated-dependent manner (Yu, 2003). The BRCT region interacts in a phosphorylation-dependent manner with DNA damaged-induced proteins *BACH1* & the transcriptional corepressor *CtIP* peptide to control G2/M checkpointing.

BRCT is approximately 200 amino acids in length spanning codons 1646 to 1859 which cover exons 16-24. Within this region are two tandem repeats each 95 amino acids long named BRCT1 and BRCT2 (BRCT-N and BRCT-C, respectively). Each BRCT domain folds independently before assembling into a quaternary structure (Gaiser, 2004). The tandem repeats are necessary as each contains a single receptor site for a binding complex (Shiozaki, 2004).

Residues 1646-1740 fold into a 4-stranded parallel beta-sheet surrounded by 3 alpha-helices forming a single domain. Each domain contains four beta-sheets flanked by two alpha-helices on one face and a third on the opposite surface (Figure 13a) (Varma, 2005). An asymmetric tertiary structure (dimer) is formed by the grouping of the two BRCT (BRCT1,

BRCT2) domains in a head-to tail manor (Figure 13b). The genomic layout of each secondary structure is n-$\beta 1 \alpha 1 \beta 2 \beta 3 \alpha 2 \beta 4 \alpha 3$-c.

The hydrophobic surfaces of the domains create a channel/groove that is important to the binding of *CtIP* and *BACH1*. The hydrophobic pocket is made of Arg1699, Leu1701, Phe1704, Asn1774, Met1775, Arg1835, and Leu1839 (Varma, 2005). The first surface is responsible for attracting a phosphoserine residue while the second determines the protein recognition specificity (Shiozaki, 2004).

BACH1 (*BRCA1*-Associated C-terminal Helicase 1) is a DEAH helicase molecule providing upstream regulation of double-strand break repair. *BACH1* (*BRIP1*) neighbors *BRCA1* mapping at 17q22. It contains 20 exons encoding 1249 amino acids. Each dimer interface contains a single receptor site for *BACH1* (Shiozaki, 2004). BACH1 introduces itself at the junction of the two BRCT domains binding with its own Ser990 and Phe993. These amino acids are heavily responsible for *BRCA1* binding and require phosphorylation at Ser990 to be recognized (Shiozaki, 2004). BACH1 pSer990 binds via hydrogen bonding to backbone Gly1656 and side chains of Ser1655 and Lys1702 on the first BRCT repeat. An additional single hydrogen bond between Ser998 and Gly1656 forms for stability.

BACH1 Phe993 at the C-terminus binds in the hydrophobic cleft formed at BRCT backbone atoms Thr1700 and Leu1701 primarily through van der Waals forces. Phe993 is anchored to sidechain molecules Phe1704, Asn1744, Met1775, and Leu1839. The vein of the pocket is formed by the latter three amino acids found on the second BRCT repeat. Additional salt bridges and hydrogen bonds stabilize the molecule (Shiozaki, 2004).

Similar to *BRCA1*-*BACH1*, *CtIP* binds using phospho-Ser327 and Phe330 at the same conserved cleft. The same major binding points Ser1655, Lys1702, Gly1656 and Leu1701 are utilized, though, several stabilizing bonds are missing from *BRCA1*-*CtIP* complex (Varma, 2005). Phosphorylation of *CtIP* at Ser327 during G2 phase controls G2/M transition during brief interactions with BRCT repeats.

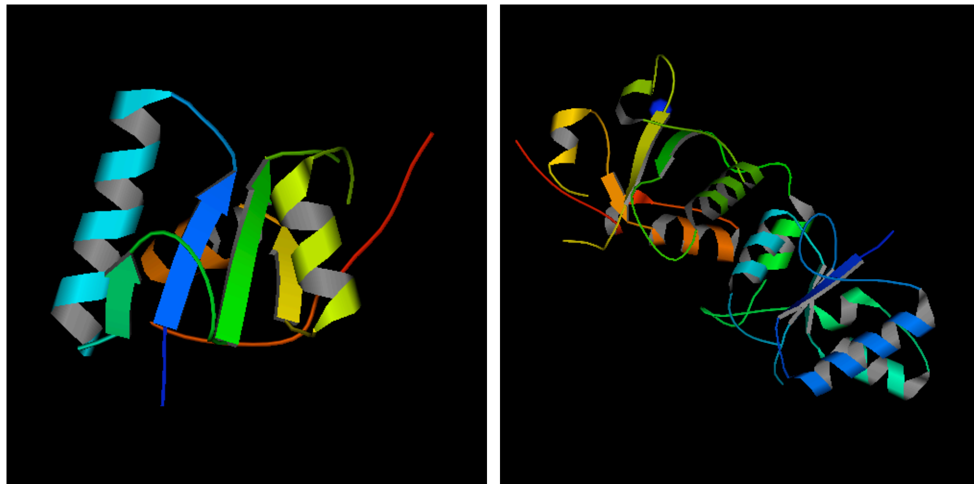

Figure 13 – Single BRCT Domain and BRCT Dimer
(a) Structure of a single BRCT repeat (1JNX_A). Four parallel beta sheets are bordered by two alpha helices and a third on distal side.
(b) BRCT tandem repeat bound in head-tail fashion. Dark green and light blue helices form backbone of hydrophobic pocket for *CtIP* and *BACH1* binding.
Visualization created using Molscript from PDB entry 1JNX.

RING – BRCA1 N-terminal

At the N-terminus *BRCA1* region lays a RING finger domain mediating protein-protein interactions. The RING domain confers E3 ubiquitin (Ub) ligase activity leading to genomic stability (Chen, 2002). In the *BRCA1* pathway *BARD1* (BRCA-associated RING domain 1) enhances ubiquitin enzyme activity and the ability of DNA binding (Brzovic, 2001).

The RING protein is approximately 100 amino acids in length extending from amino 1 through 103 covering exons 2-7. The tertiary structure consists of three alpha helices, three beta sheets and two Zinc binding loops (Brzovic, 2001). Residues 8-22 and 81-96 form anti-parallel alpha helices flanking the central backbone at 23-76. A third helix from residues 46-53 is separated from the backbone by three near parallel beta sheets. The very short beta sheets originate from residues 35-37, 42-44, 74-76.

The RING tertiary structure is very unique consisting of three cysteine residues, one histidine, and another four cysteines in a doubled-over ring pattern often referred to as a 'cross-brace' motif. In the middle of each ring binds a Zinc atom (Figure 14). The RING motif consists

of seven Cysteine residues and Histdine residues arranged in pairs spaced two amino acids apart (Brzovic, 2001). This is also known as a C3HC4-type ring variant (Figure 14). Two binding domains are formed by these residues which interact with two Zn^{2+} atoms. Site I is formed by residues Cys 24, Cys 27, Cys 44 and Cys 47. Site II ligands include Cys 39, His 41, Cys 61 and Cys 64. These sites differ greatly from BRCT in that there is significant overlap between contributing amino sequences.

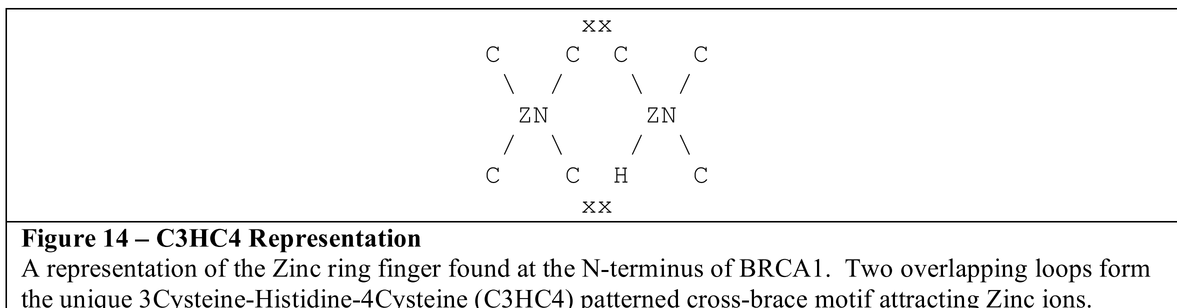

Figure 14 – C3HC4 Representation
A representation of the Zinc ring finger found at the N-terminus of BRCA1. Two overlapping loops form the unique 3Cysteine-Histidine-4Cysteine (C3HC4) patterned cross-brace motif attracting Zinc ions.

The portion of *BARD1* with binding affinity to *BRCA1*'s RING contains a similar RING domain. *BARD1* exhibits homology to *BRCA1*'s RINGs with a slight alignment shift and five fewer amino acids. *BARD1* residues 36-48, 101-116 form the same alpha helices surrounding the central motif at 49-100. Residues 61-63 and 69-71 form two beta sheets. Notice *BARD1* is missing both a beta sheet and the diminished lower alpha helix. Zn^{2+} site I is shifted toward the 3' end to Cys 50, Cys 53, Cys 71 and Cys 74 and respectively site II to Cys 66, His 68, Cys 83 and Cys 86 in *BARD1*.

In the *BRCA1*-*BARD1* heterodimer complex the *BARD1* subunit's two alpha helices bundle with the *BRCA1*'s. In Ubiquitin ligase (E3) complex primary interaction occurs through the first Zn2+ loop, central helix, and second Zn2+ loop (Brzovic, 2001), essentially at the RING-RING interface. The resulting dimer interface contains hydrophobic interaction points for binding to other proteins. Helix side chain residues Gln19, Leu22, Val83, Glu84, and Leu87 contact Leu30, Ile31 and Lys32 in the subunit (Brzovic, 2001). The helix angle between fingers is determined by these residues forming a cleft for E3 ligases UbcH7 and UbcH5 binding. The

BRCA1-BARD interface occurs though helical side chain interactions at Arg7, Glu10, Glu85, and Asp96 (Figure 15).

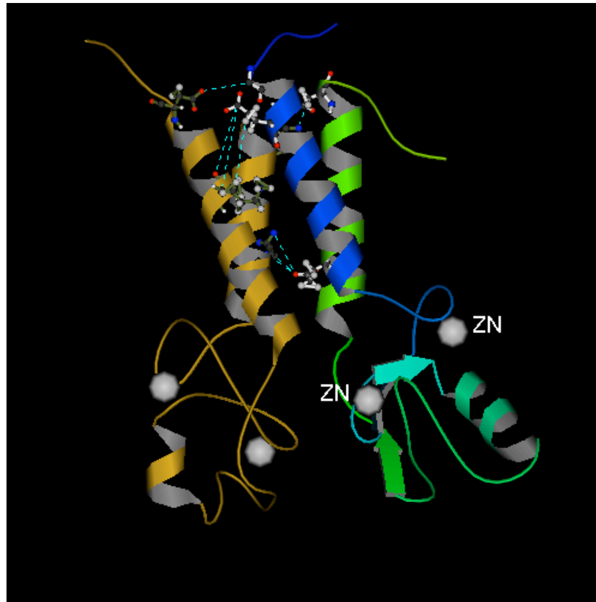

Figure 15 – BRCA1 RING Domain bound to BARD1
Two alpha helices compose the backbone binding region of the BRCA1 RING domain pictured in blue and green. The dark green lower portion denotes the Zn-finger responsible for E3 ubiquitin-ligase activity. On the left in brown is the homologous *BARD1* protein forming a complete heterodimer. Visualization created using Molscript from PDB entry 1JM7.

Other BRCA1 Binding Sites

Although the BRCT and RING domains account for only 16% of the *BRCA1* gene, the remaining structure remains largely unstudied. Several other important binding sites occur in the central region though it is generally less structured than the terminal ends.

FANC complex protein FANCA interacts in the *BRCA1* central region at residues 740-1083 (Folias, 2002). FANCC may also interact in this region to up-regulate FANCD2 but the binding sites are unidentified.

Amino acids 340-534 exhibit beta sheeting folding in a flexible manor possibly allowing binding to several different conformations. Experiments suggest this region is involved in homologous repair through binding to four-junction DNA (Naseem, 2006). However, no protein binding sites were identified in this region.

In response to DNA damage, Cds1 (*Chk2*) regulates *BRCA1* through phosphorylation of Ser998 (Lee, 2002). Phosphorylation at this site is required for release from hCds2 molecule to restore S-phase and M-phase.

In response to IR, *ATM* phosphorylates at Ser1387 for intra-S checkpoint (Gatei, 2001). During G2/M checkpoint *ATM* phosphorylates at Ser1423 and Ser1524 in response to both IR and UV (Choudhury, 2005). It appears *ATM* phosphorylation of *BRCA1* Ser1423 at these Serines act as a p53 adaptor allowing *ATM* phosphorylation (Fabbro, 2004). Mutations at any of these sites will likely affect *BRCA1* binding affinity and function.

BRCA1 Mutations

Various mutations are introduced into all DNA during replication, transcription, and translation. These may be as simple as a single nucleotide substitution, insertion or deletion (indel), or as complex as inversions, breaks, splices, and rearrangements. In 1994 Swensen found an 11-bp deletion and a 1-bp insertion creating a stop codon in exon 2 leading to initial isolation of the *BRCA1* gene (Miki, 1994). Since then, the NIH Breast Cancer Mutation Database has documented more than ten thousand other mutations.

Each exon encodes between 37-300 base pairs with the exception of exon 11 which covers over 60% of the gene. Mutations are evenly distributed through coding regions though due to its size exon 11 contains the majority. About 70% are insertion frameshifts, creating truncated reading frames through a down sequence stop codon resulting in severely truncated protein products approximately 85% of the time. The top three reported mutations represent 30% of the total and fall into one of the terminal regions. About one-third of mutations in either terminal domain are of missense type.

Most Frequent Mutations Reported by BIC

Using the Breast Cancer Information Core Database (BIC) database the most frequently occurring exon mutations are 185delAG, 5382insC, C61G, R1347G, R841W. The first two produce premature stop codons in the BRNT and BRCT regions. The C61G mutation falls into

third exon of RING domain severely affecting confomation, resulting in the inability to confer ubiquitin ligase (E3) activity (Mallery, 2002). The other two mutations both are missense mutations in exon 11 with no known effects.

BRCT Mutations

The BRCT region spans exons 16-24 covering codons 1646-1859 and is a frequent location of cancer causing mutations (Table 7). 5382insC is the most frequent mutation occurring within this range. As the designation implies, it is an insertion mutation at codon 1756 (cDNA position 5382) resulting in a frameshift stop codon at 1829. Every frameshift mutation in this domain results in a premature stop codon within the coding region. Additionally, some single nucleotide polymorphisms (nonsense mutations) at Arg1835 and Arg1751 cause stop mutations. All of these stops result in truncation of one or both BRCT repeats and ultimately inactivation of the protein. Truncation of only one repeat does not allow protein binding since no hydrophobic pocket is formed. Lab experiments show BRCT can withstand small truncations but deletions greater than 8 residues hinder protein folding (Williams, 2003).

A mutation introduced at intron 21 (VIS21-36del510) results in a down-sequence stop codon at the 1805 amino acid (Williams, 2003). Other mutations in this region such as R1751Q disrupt salt bridges in the BRCT linker helix (Williams, 2003). Larger deletions such as the removal of exon 21 or 22 remove enough of the functional domain to inhibit any protein folding or bonding (Zikan, 2005).

The frequently occurring missense mutations Met1652, Arg1699, Ala1708 and Met1775 occur on the BRCT repeats (Cantor, 2001). Met1775 and Arg1699 reside on opposite domains in the conserved surface cleft. *BACH1* Phe993 stacks closely against the second repeat at sidechains of Met1775. Mutation M1775R is unlikely to alter folding but does result in a charge-charge repulsion at the hydrophobic core (Williams, 2001). This significantly disrupts recognition of *BACH1* and *CtIP* (Vallon-Christersson, 2001). In *CtIP*, the substitution change causes Phe330 to collide with the sidechain of Arg1775 (Varma, 2005). Lab experimentation has shown double-

strand break repair is abrogated by mutation at this site (Figge, 2004). Substitutions at A1708, M1775 and M1783 are considered interface mutations whereas protease sensitivity falls and hydrophobic core disruption results from these changes.

Table 7 – Frequent BRCA1 Mutations
The most frequently occurring mutations according to the Breast Cancer Information Core Database (BIC) are listed in this table. Although mutations are evenly distributed throughout all exons, the top three mutations 185delAG, 5283insC and C61G account for approximately 30%.

Exon	Designation	Type	NT	Codon	Effect	Count
2	185delAG	F	185	23	Stop 39	1596
20	5382insC	F	5382	1756	Stop 1829	836
5	C61G	M	300	61	M	155
11	R1347G	M	4158	1347	UV	149
11	R841W	M	2640	841	UV	114
11	M1008I	M	3143	1008	UV	105
11	4184del4	F	4184	1355	F	105
13	R1443X	N	4446	1443	N	103
11	3875del4	F	3875	1252	F	97
11	E1250X	N	3867	1250	N	76
11	Q563X	N	1806	563	N	70
16	M1628T	M	5002	1628	UV	70
18	A1708E	M	5242	1708	M	39

Arg1699 participates in salt bridges between the BRCT repeats and shared hydrogen side chains of Phe993. R1699W affects tertiary structure stability which significantly disrupts *BACH1* and *CtIP* bonding (Vallon-Christersson, 2001). R1699Q has no effect and R1699S disrupts the salt bridge interface (Vallon-Christersson, 2001). Modeling of the transition from Arginine to Tryptophan predicts Trp1699 sterically collides with Phe330 obstructing entrance to the hydrophobic pocket (Varma, 2005). Inhibition at this site undoubtedly disrupts double-strand break repair (Scully, 1999).

The two other common missense mutations occur at Met1652 and Ala1708 and are buried deep in the hydrophobic crevice. A1708E is on the first domain and Met1652 is on the second. The A1708E mutation is the most frequently occurring BRCT repeat mutation according to BIC. Alteration at these points disrupts folding, and ultimately the final protein structure (Shiozaki, 2004). A positively charged Alanine to a negatively charged Glutamine destabilizes

structure producing an insoluble domain. This is a clear example of tertiary structure relating to function.

M1652I is thought to be a benign polymorphism due to its peripheral location on the C-terminus of the first BRCT repeat. Although found frequently in the presence of cancer-causing mutations it has not been shown to cause structural deformities (Williams, 2003). D1692N is a common mutation in cancer patients but a direct link to cancer has not been established. This mutation causes a weakening of the salt bridge to S1715 but does not significantly affect the conformation.

BACH1 binds to the backbone molecules Gly1656, Ser1655 and Lys1702. Ser1655 and Lys1702 sidechains donate hydrogens to *BACH1* Ser990 while Gly1656 has an additional binding to backbone *BACH1* pSer998. Ser1655 has several frequently occurring 19 base frame shifts resulting in an upstream stop codon at 1670. This renders a shortened non-functioning protein. Missense S1655F mutation completely blocks interaction with *BACH1* pSer990 (Shiozaki, 2004) and *CtIP* Ser327 (Varma, 2005). Though very disruptive the BIC database only reports a single occurrence at Gly1656 and Lys1702.

Atoms in the backbone of the pocket such as Thr1700 and Leu1701 show only a few frameshift mutations. Changes (other than deletions) at these sites would weaken *BACH1* binding by diminishing van der Waals contacts at Phe993.

Remaining backbone binding sites Phe1704 in the first repeat and Asn1774 and Leu1839 in the second have no mutations listed in BIC. To analyze mutations at Phe1704, Asn1774, Leu1839 residues, I mutated amino acids at each site to the most common amino change. For example, Phenylalanine most frequently mutates to Leucine in *BRCA1*. Therefore, in my predicted structure, Phenylalanine was changed to Leucine. Similarly, Asn1774 was mutated to complementary Lysine and Valine and Leu1839 was mutated to Valine and Phenylalanine.

F1704L had no significant changes in between the base and predicted models other than a 7 degree Phi change, although it is possible this change would obstruct the BRCT cleft for *CtIP*

or *BACH1*. The mutated A1774K/V angles lie in forbidden regions of the Ramachandran for both the stock and predicted models. ProteinPredict shows a three-fold increase in relative solvent accessibility for the Lysine mutations. No significant change is seen for modification to Histidine. Based on these results mutations are unlikely to interfere with binding activity though crystallography analysis would be necessary to confirm these results.

Table 8 – BRCT Secondary Structure Effects
Mutations have a wide variety of effects on secondary structures ranging from little or no effect to completely abrogating binding ability. Note mutations at Arginine 1699 (R1699) have different effects depending upon amino acid modification.

Mutation	Structure Location	Predicted Structure Changes	Citation
M1651I	N-b3 surface	No effect	(Williams, 2003)
M1652I	N-b4 surface	No effect	(Deffenbaugh, 2002)
D1692N	N-ß3/a2	Destabilizes salt bridge with S1715	(Joo, 2002)
D1692Y	N-surface	Slightly alter protease recognition	(Joo, 2002)
Q1694X		Skip exon 18	(Liu, 2001)
F1695L	N-surface	No effect	(Joo, 2002)
V1696L	N-surface	Slightly alter protease recognition	(Joo, 2002)
C1697R	N-surface	Alter folding & strong destabilizing	(Williams, 2003)
R1699W	N-a2 surface	Affects stability	(Worley, 2002)
R1699Q	N-a2 surface	None	(Hayes, 2000)
R1699S	N-a2 surface	Disrupts interface salt bridge	(Williams, 2003)
F1704L	N-a2 core	No significant change	(Williams, 2003)
A1708E	N-a2 surface	Dirupts interface / hydrophobicity	(Williams, 2003)
S1715R	N-b4/a3 fold	Alter folding & strong destabilizing	(Williams, 2003)
W1718C	N-a3 fold	Alter folding & strong destabilizing	(Williams, 2003)
T1720A	N-a3 surface	No effect	(Williams, 2003)
G1738R	N/C interval	Alter folding & strong destabilizing	(Williams, 2003)
G1738E	N/C interval	Alter folding & strong destabilizing	(Williams, 2003)
P1749R	C-b3 surface	Alter folding & strong destabilizing	(Williams, 2003)
R1751Q	C-a1 surface	Salt Bridge disruption	(Williams, 2003)
A1752P	C-b3 surface	Disrupts helix linker	(Williams, 2003)
I1766S	C-b3 fold	Alter folding & strong destabilizing	(Williams, 2003)
N1774V	C-terminus surface	No significant change	(Williams, 2003)
M1775R	C-a2 surface	Dirupts interface / hydrophobicity	(Williams, 2003)
M1783T	C-a2 surface	Dirupts interface / hydrophobicity	(Williams, 2003)
G1788V	C-surface	Alter folding & strong destabilizing	(Williams, 2003)
G1788V	C-surface	Disrupts a1-b2 turn	(Williams, 2003)
V1804D	C-b3 surface	No effect	(Williams, 2003)
V1809F	C-b3 surface	Alter folding & strong destabilizing	(Williams, 2003)
V1833M	C-b4 pocket	Alter folding, destabilize hydrophobic pocket	(Williams, 2003)
W1837G	C-a3 core	Alter folding & strong destabilizing	(Williams, 2003)
W1837R	C-a3 core	Alter folding & strong destabilizing	(Williams, 2003)
L1839V	C-a3 surface	Alter folding & mild destabilizing	(Williams, 2003)
Y1853C	C-terminus surface	Alter folding & strong destabilizing	(Williams, 2003)

If present in vivo, the L1839V mutation would fall within a conserved surface cleft region. This is an area of highly conserved residues. Helix prediction decreases significantly for the L1839V mutation. Coupled with Psi/Phi angle changes the $\alpha 3$ helical conformation is significantly changed.

Table 8 lists secondary structure effects caused by SNPs or amino acid mutations. Most of these amino acids occur on the surface of alpha helixes or beta sheets. Although some mutations present little or no effect, the majority alter folding and are highly destabilizing to the protein structure.

RING Mutations

The N-terminal region of *BRCA1* contains a zinc ringer finger domain covering 103 residues over a five exon span. Approximately 15% of *BRCA1* mutations are found within this short range including the most common, 185delAG, occurring 1600 times. The adjacent 188del11 mutation lead to Miki's discovery of *BRCA1* in conjunction with C61G and C64G.

Similar to BRCT, some mutations at critical binding sites affect conformation while others have little structural effect. For example, the frameshift of two amino acids at Glu185 generate an early Stop codon rendering the protein unviable and truncation of the all but the first alpha helix. This is unlike the more common R71G substitution falling adjacent to a beta sheet but having little effect on final conformation.

According to BIC data, site II mutations occur ten times more frequently than in site I. Lab experiments have shown mutations at residues Cys39, Cys61 and Cys64 in the second loop still allow for binding with *BARD1* (Brzovic, 2001). Though occurring more frequently site II mutations are far less disruptive to *BRCA1* function. This is supported by the fact that site I ligands are targeted as predisposing patients with cancer. Site I missense mutations at Cys24, Cys44, or Cys47 cause improper protein folding and ultimately abrogate Ub-ligase ability to confer (Brzovic, 2003). Mutations at Site II Cys27, His41 have not been seen in patients.

Only a handful of mutations were reported in the helix and Ub-cleft regions. Ub-cleft mutations significantly alter the width of the binding interface and hinder Ub activity. At the dimer interface mutations alter the protein's affinity toward BARD. It is unknown at this time the exact impact if the charge field is tolerated.

In an attempt to predict mutation effect at Cys27 and His41, I passed two mutated 1JM7 sequences through ProteinPredict and SwissModel. I modified these residues to Tyrosine as it's a commonly mutation in the Zn2+ rings. Secondary structure prediction was done using ProteinPredict, SwissModel and DeepView. C27Y and H41Y modifications were submitted to all prediction schemes at Protein Predict. C27Y had mixed predictions for helix and sheet content but almost 90% confidence for non-secondary structure regions. PROFsec (secondary structure) search results show an intact helix at residues 27-30 but with ultra low prediction reliability. This is confirmed by PROFacc (accession) prediction for no secondary structure with a medium confidence between 33-55%. Relative solvent accessibility falls in the 9% range which compares favorably to the stock germ sequence.

H41Y differed slightly with PROFsec predicting helix, strand, or neither equally plausible with 40% confidence. PROFacc predicted moderate solvent accessibility within the H41 region which is consistent with the residue location at edge of beta sheet 2. ProteinPredict predicted a high confidence both mutations would remain as strands with 0-9% solvency.

The modified sequences were then fed into SwissModel to generate predicted PDB files. The predicted files were then rendered in MolScript and also viewed in DeepView for accessing the quality of the structure.

Using DeepView to analyze the secondary structure, both Psi and Phi angles increased about 5 degrees between the original model and a single mutation. Neither the H41 nor C27 mutations have values resulting in sterical clashes. Limited structure variability exists compared to the unmanipulated protein sequences so binding would probably continue with Zn^{2+}.

Table 9 lists other mutation sites within *BRCA1*-N. Less structural research on N-terminal mutations has been conducted than on the C-terminus. *BARD1* structural mutation research far exceeds that done on *BRCA1* N-terminus. As can be seen in the table C39, C44, and C47 modifications alter folding but with unknown consequences. Perhaps the most significant mutation is the Leu53 and Leu53 in alpha helix-2 that abolishes bonding with *BARD1*.

Table 9 – BRCN Secondary Structure Effects
There are two areas of importance for mutations in the N-terminal region. First, mutations found along the backbone structure affect the ability of *BARD1* to bind efficiently. Second, mutations in the RING will influence Ub-ligase activity.

Mutation	Structure Location	Predicted Structure Changes	Citation
R7C	Adjacent to a1	Destroy salt bridge w/BARD1 Trp34	(Brzovic, 2001)
C24R	Adjacent to a1	Alter folding	(Brzovic, 2001)
C27	Central motif	No effect predicted	
H41	Central motif	No effect predicted	
C44F	b2 surface	Alter folding	(Brzovic, 2001)
C47F	Adjacent to a2	Alter folding	(Brzovic, 2001)
C39S	Between b1-b2	Slight structure alteration	(Brzovic, 2001)
C39R	Between b1-b2	Slight structure alteration	(Brzovic, 2001)
C39Y	Between b1-b2	Slight structure alteration	(Brzovic, 2001)
L52F	a2 surface	Protein-protein interaction	(Brzovic, 2001)
L53F	a2 surface	Protein-protein interaction	(Brzovic, 2001)
C61G	Central motif	Unknown	
C64G	Central motif	Slight structure alteration	(Brzovic, 2001)
C64Y	Central motif	Slight structure alteration	(Brzovic, 2001)
R71G	Adjacent to b3	No effect	

Central Region / Other Domain mutations

Other regions in the BRCA1 gene have important functions but neither well defined domains nor binding sites. Unfortunately this inhibits research on structural changes caused by mutations in the central domain.

Exon 11 encodes approximately 61% of the gene by spanning codons 263 to 1370. Deletion of this exon renders a truncated nonfunctioning p53 binding site. p53 mediates transcription in *BRCA1* ranges 224-500 and the second BRCT domain but exact binding sites are unknown (Zhang, 1998).

Only one mutation at Ser1423 is listed in BIC for all the binding sites. Mutation at Ser1423 abrogates *ATM* ability to phosphorylate this site (Xu, 2001). However throughout the 1241-1530 range over one-thousand mutations are listed mostly consisting of Ser to Ser polymorphisms or Arg to Stop changes.

Cds1 (*hCds1*/*Chk2*) binds at Ser988 to regulate *BRCA1* through phosphorylation. In-vitro mutations at this residue have been shown to abrogate the ability of *Chk2* to regulate *BRCA1* (Lee, 2002). Mutation at Ser988 abrogates ability of *Chk2* to phosphorylate *BRCA1* and restore cell cycle after DNA damage (Liu, 2002). Qin reported *Chk2* may also bind to Thr1852 in the BRCT region (Qin, 2003). No mutations have been reported to BIC at either site or additional data to support his findings.

CDK2 (Cyclin-dependent kinase E) phosphorylates *BRCA1* at Ser1497 and Thr967. Substitution of Serine to Alanine or Threonine to Asparagine lowers but does not prevent *CDK2* phosphorylation along this region (Ruffner, 1999).

Table 10 – BRCA1 Central Region Mutations
Few mutations are documented in the central region of *BRCA1*. Laboratory testing has demonstrated mutations at key binding sites severely inhibit binding.

Mutation	Structure Location	Predicted Structure Changes	Citation
Thr967	Exon 11	May effect CDK2 phosphorylation.	(Ruffner, 1999)
Ser988x	Exon 11	May effect Chk2 binding.	(Liu, 2001)
del Exon11	Exon 11	p53 unable to bind to BRCA1.	(Zhang, 1998)
Ser1423	Exon 13	Abrogates ATM ability to phosphorylate.	(Xu, 1999)
Ser1497x	Exon 14	May effect CDK2 phosphorylation.	(Ruffner, 1999)
Thr1852x	Exon 24	None	(Qin, 2003)

BRCA1 Conservation

To study gene conservation I first observed conservation at the known binding points located at each terminus. I then searched for homologous sequences around central region phosphorylation sites. Finally, I compared conservation at the terminals to other species.

Starting with the RING section of *BRCA1* a FASTA formatted sequence of 1JM7 was submitted to NCBI's PSI-BLAST (Table 11).

The majority of BLAST results were laboratory isoform splice variants. These results were discarded from the homology search because they do not naturally occur *in vivo*.

Table 11 – 1JM7 1JM7 is the solution structure representation of *BRCA1-BARD1* RING-domain heterodimer. The amino acid sequence for *BRCA1* (Chain_A) is listed below from NCBI GI:15988069.
MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKK GPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFAKKGK

Additional searches were conducted using NCBI-Blast2, RPS-BLAST, and CDART to gather sequences from alternative Uniprot and SwissProt databases. Sequences from all search results were combined into a single text file for multiple alignment in ClustalW. Several other alignments using PHD and MaxHom (Multiple Sequence Alignment) from Protein Predict did not yield significant complements.

Table 12 – Conservation of RING Domain

1JM7 RING domain aligned against other Human sequences. Little homology exists throughout N-terminal region, however, note exceptionally strong conservation of C3HC4 RING (blue) amino acids. Alpha helices are denoted in red. Beta sheets in yellow. Sequences are labeled with protein and NCBI GI number.

```
                         |10         |20         |30          |40         |50          |60
1JM7|15988069:   8 VEEVQNVINAMQKILECPICLELIKEPVST-KCDHIFCKFCMLK-LLNQKKGPSQCPLCKNDIT
TRIP31|55961262: 1 -MASGQFVNKLQEEVICPICLDILQKPVTI-DCGHNFCLKCITQ-IGETSCGFFKCPLCKTSVR
Tri31|1770499:  16 -MASGQFVNKLQEEVICPICLDILQKPVTI-DCGHNFCPQCITQ-IGETSCGFFKCPLCKTSVR
RNF1|11022690:   1 -MASKILLN-VQEEVTCPICLELLTEPLSL-DCGHSLCRACITVSNKE-MGGKSSCPVCGISYS
INF1|Q9BYJ4:     1 -MASKILLN-VQEEVTCPICLELLTEPLSL-DCGHSLCRACITVSNKE-MGGKSSCPVCGISYS
TRIM22|23272543: 1 -MDFSVKVD-IEKEVTCPICLELLTEPLSL-DCGHSFCQACITAKIKE-SRGESSCPVCQTRFQ
TRIP22|Q8IYM9:   1 -MDFSVKVD-IEKEVTCPICLELLTEPLSL-DCGHSFCQACITAKIKE-SRGESSCPVCQTRFQ
TRIP5|Q9C035:    1 -MASGILVN-VKEEVTCPICLELLTQPLSL-DCGHSFCQACLTANHKK-DKGESSCPVCRISYQ
Ro/SSA|14790039: 1 -MASAARLTMMWEEVTCPICLDPFVEPVSI-ECGHSFCQECISQVGKG---GGSVCPVCRQRFL
SSA1|14994115:   1 -MASAARLTMMWEEVTCPICLDPFVEPVSI-ECGHSFCQECISQVGKG---GGSVCPVCRQRFL
SjA1|53690156:   1 -MASAARLTMTWEEVTCPICLDPFVEPVSI-ECGHSFCQECISQVGKG---GGSVCPVCRQRFL
SSA56|15982946:  1 -MDPTALVEAIVEEVACPICMTFLREPMSI-DCGHSFCHSCLSGLWEI-QNWGYTCPLCRAPVQ
RING137|80474981:1 ------LVEAIVEEVACPICMTFLREPMSI-DCGHSFCHSCLSGLWEI-QNWGYTCPLCRAPVQ
IRIS|37537746:   7 VEEVQNVINAMQKILECPICLELIKEPVST-KCDHIFCKFCMLKLLN-QKKGPSQCPLCKNDIT
Breast|94315232: 7 VEEVQNVINAMQKILECPICLELIKEPVST-KCDHIFCKFCMLKLLN-QKKGPSQCPLCKNDIT
LON1|73747840:  15 GAAAAAAATEVWDGFKCRKCHGFLSDPVSL-SCGHTFCKLCLERGRAADRRCALCGVKLSALMV
Neuro|68032982: 15 EGGPAPEPRAPRDLLGCPRCRRLLHKPVTL-PCGLTVCKRCVE------PGPAR-----PQVRR
Tri73|Q86UV7:    1 -MAWQVSLLELEDRLQCPICLEVFKESLML-QCGHSYCKGCLVSLSYH-LDTKVRCPMCWQVVD
RAD18|Q9NS91:    9 WPPGLAVMKTIDDLLRCGICFEYFNIAMIIPQCSHNYCSLCIRKFLSY----KTQCPTCCVTVT
HSD4|6468773:    2 AEDLSAATSYTEDDFYCPVCQEVLKTPVRTTACQHVFCRKCFLTAMRE---SGAHCPLCRGNVT
TRAC1|Q96EQ8:   15 ALERRRDPELPVTSFDCAVCLEVLHQPVRT-RCGHVFCRSCIATSLKN---NKWTCPYCRAYLP
TNF6|Q9Y4K3:    15 QGYDVEFDPPLESKYECPICLMALREAVQT-PCGHRFCKACIIKSIRD---AGHKCPVDNEILL
```

Reviewing the multiple alignments yielded some interesting results. The 1JM7 alignment is nearly identical within the two ring regions in most other sequences. Sequences in Table 12 are listed in descending order of homology score (E-value).

While all the sequences exhibit strong sequence similarity to the first RING Cysteines and Histidine, the second RING is less similar. *TRIP31* and *TRI31* RINGs exhibit strong homology at both binding sites and the surrounding RING region.

RNF21 (Long Interferon-responsive finger protein 11) clone and *INF1* (Interferon-responsive finger protein 1) are interferon-responsive finger proteins. Both are highly conserved at binding points including the Histidine. This high level of conservation may be attributed to the double RING box-coiled coils (RBCC) shape. This unique structure does not tolerate many mutations. *TRIP5* (Tripartite motif-containing protein 5) and *TRIM22 / TRIP22* (Tripartite motif-containing 22) also express similar homology. The remaining sequences vary greatly and express little homology outside the key binding sites.

The BRCT repeats interact with phosphorylated proteins containing the signature sequence pSer-X-X-Phe where X indicates any residue (Varma, 2005). Threonine and Tryptophan terminated sequences are also possible but research suggests this combination is (Varma, 2005). This unique configuration allows numerous mutations between the Phe and Ser binding points to have minimal or little effect on BRCT recognizing *BACH1*.

Table 13 – 1JNX 1JNX is the crystal structure of the *BRCA1* C-terminal repeat from NCBI GI:15988247.
VNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVMKTDAEFVCER TLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGPKRAR ESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPI VVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIP

Entrez Protein and PSI-BLAST returned a combined total of 300 reference sequences for *Homo sapiens*. Of these, only 17 are unique sequences and not isoforms. Far fewer alignments were found at the ring domain. Using ClustalW the list was narrowed to 13 sequences containing at least one conserved amino acid. I also attempted to BLAST a single BRCT repeat which shortened the search sequence to just 96 amino acids. The results from this search were identical to the full length search. My hypothesis had been a single BRCT domain would occur more frequently than the dimer. This means no sequences are homologous to a single BRCT repeat.

Several *BRCA1* related proteins show strong homology at the key binding sites for *BACH1* and *CtIP* Ser998 and pSer990 (1655-56, 1702) and Phe993(1699-1704, 1774-75, 1839). It is interesting that the RING binding proteins *BARD1* and *MDC1* (DNA Damage Mediator) exhibit such strong identity to the C-terminus sequence. *TP53* (*p53* binding protein) has somewhat less resemblance to the backbone chain yet maintains major binding sites.

G2/M phase regulators are the highly conserved because they control mitosis (Jin, 2000). Residues interacting with the first motif (pSer990) are highly conserved while the second motif is least conserved. Specifically, Phe993 binds at the hydrophobic pocket of 1774-1775 in exon 22. In this region neither of these residues are preserved in *MDC1*, *BARD1*, or *p53* being replaced by different amino acids (Shiozaki, 2004). Exon 22 is the least conserved exon in BRCT.

Table 14 – Conservation of BRCT Domain
The C-terminal domain contains little homology at key binding sites and secondary structure. Blue denotes homologous sites; Alpha helices in red; Beta sheets in yellow. Sequences are labeled with protein and NCBI GI number.

<table>
<tr><th></th><th>|1655</th><th>|1699</th><th>|1774</th><th>|1788</th><th>|1839</th></tr>
<tr><td>1JNX|15988247:1653</td><td>VVSGLTPEEFML</td><td>ERTLKYFLGIAG</td><td>FTNMPTDQLEWM</td><td>QLCGASVVKELS</td><td>PVVTREWVLDS-</td></tr>
<tr><td>MDC1|84040269: 517</td><td>MDSPPHQKQPQR</td><td>RRTVKFLCALGR</td><td>VQPPPPQMGEII</td><td>SCCGGTYLPSMP</td><td>PEFLLTGVLKQE</td></tr>
<tr><td>BARD1|4557349: 571</td><td>IGSGLSSEQQKM</td><td>QSTLKCMLGILN</td><td>FKHHPKDNLIKL</td><td>TAGGGQILSRKP</td><td>CNYHPERVRQGK</td></tr>
<tr><td>BARDLOC|2828068:</td><td>RPVDYTDDESMK</td><td>QSTLKCMLGILN</td><td>FKHHPKDNLIKL</td><td>TAGGGQILSRKP</td><td>CNYHPERVRQGK</td></tr>
<tr><td>TP53B|8928568:1691</td><td>KPGAVGAGEFVS</td><td>CRTRKYFLCLAS</td><td>QQQNFLELWSEI</td><td>MTGGAASVKQHH</td><td>PVVSQEWVIQCL</td></tr>
<tr><td>TP53B|488592: 745</td><td>KPGAVGAGEFVS</td><td>CRTRKYFLCLAS</td><td>QQQNFLELWSEI</td><td>MTGGAASVKQHH</td><td>PVVSQEWVIQCL</td></tr>
<tr><td>MCPH1|47605903 597</td><td>EEKENLPGGYSG</td><td>LRTLNVLLGIAR</td><td>ASSPPVAKLCEL</td><td>HLCGG-------</td><td>KYLSEKWVLDS-</td></tr>
<tr><td>KIAA0259|1665785:</td><td>HDSEIAKQAVCD</td><td>LRNEKYLASVAA</td><td>ILHVDQSREAGF</td><td>LQSGGAKVLPGH</td><td>YCLRTEYIADYL</td></tr>
<tr><td>CAGF28|2565046: 8</td><td>KQNEVANVQPSS</td><td>TRTLKFLAAISV</td><td>GICPSLSTMKAI</td><td>ECAGGKVLSKQP</td><td>LHLCREYFARGI</td></tr>
<tr><td>XRCC1|139820: 131</td><td>VCSQPYSKDSPF</td><td>EDELRRVAEQKE</td><td>FANTPKYSQVLG</td><td>AGPGSSSEEDEA</td><td>PVLQEDIDIEGV</td></tr>
<tr><td>SerFCP1a|3769521:</td><td>QAQECGHLHVVN</td><td>IRTGARGPPAPS</td><td>LGEGSDDSDSEK</td><td>RPEEQEEEPQPR</td><td>EDAASESSRESS</td></tr>
<tr><td>VPARP|5702306:</td><td>HPGGFTTRPSAG</td><td>SRCPVFAFQSSD</td><td>LNLNTNGLHSFL</td><td>QKGIQSLGVKGR</td><td>EDAASESSRESS</td></tr>
<tr><td>t-DONT|181649:</td><td>MTGGFRRGKKMG</td><td>ESTFEKLRLPSR</td><td>WTGSPRQFERDL</td><td>R-----------</td><td>KAESEEEIFAHL</td></tr>
<tr><td>PESC|21542165: 378</td><td>VDRPGQQTSVIG</td><td>EKEGDYVPPEKL</td><td>--EQRMEGKKPR</td><td>MAGTLKLEDKQR</td><td>IREANKLAEKRK</td></tr>
</table>

I was surprised the *XRCC1* BLAST alignment has little sequence similarity to BRCT other than a few key binding sites. Since BRCT style repeats in *XRCC1* were used as a crystallography model for *BRCA1*, the aforementioned originally led me to anticipate a stronger protein sequence homology.

To model BRCT, I used the *XRCC1* DNA repair protein, 1CDZ. *Structure of an XRCC1 BRCT domain: protein–protein interaction module* (Zhang, 1996) by Xi Zhang discusses the sequence and structural similarities between these two repair proteins. They confer at high levels of similarity, invariance and conservation. However, my own analysis shows few amino acids

are actually conserved let alone ones forming important secondary structures or binding sites. Only eight residues in all beta sheets and alpha helices are conserved (Table 15).

Table 15 – BRCT vs. XRCC1
The crystal structure of XRCC1 is often used to model BRCA1's tandem repeat. Little homology exists between the two sequences though tertiary structure is remarkable similar. The 1JNX sequence is repeated to clarify homologous sites (gray). Alpha helices are shown in red; Beta sheets in yellow. Sequences are labeled with protein and NCBI GI number.

```
                      |1730               |1750     |1760     |1770     |1780
1JNX|15988247     VRGDVVNGRNHQGP----------KRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCG-
1JNX|15988247     VRGDVVNGRNHQGP----------KRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCG-
XRCC1|9954649     IEGVQSEGQDNGAEDSGDTEDELRRVAEQKEHRLPPGQEENGEDPYAGSTDENTDSEEHQEPP

                   |1790     |1800               |1810     |1820     |1830
1JNX|15988247     ASVVKELSSFTLG----------TGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREW-VLDSVAL
1JNX|15988247     ASVVKELSSFTLG----------TGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREW-VLDSVAL
XRCC1|9954649     DLPVPELPDFFQGKHFFLYGEFPGDERRKLIRYVTAFNGELEDNMSDRVQFVITAQEWDPSFEEAL

                       |1850
1JNX|15988247     YQCQELDTYLIPQIP------
1JNX|15988247     YQCQELDTYLIPQIP------
XRCC1|9954649     YSCNEKQKLLPHQLYGVVPQA
```

Risk of Breast Cancer

Although most of the identified *BRCA1* mutations have been designated as benign polymorphisms, many clinically validated ones lead to cancer predisposition. The majority of mutations lead to protein products missing one or more BRCT domains (Yu, 2004). Mutations in this region cause truncation of the domain(s) resulting in partial or complete loss of protein function (Varma, 2005). Commonly removal of these domains generally leads to an increased incidence of tumors. Less severe substitutions and missense mutations disrupt the secondary structure. *BRCA1* & *BRCA2* mutations only account for a small percentage of the total population affected by the disease. Other genetic components such as *p53*, *NBS1*, *ATM* may contribute to extended cell growth.

Risk is significantly altered depending upon the number of mutations in a single gene or across multiple genes (*BRCA2*, *CHEK2*, *RAD51*). Mutations in *BRCA1* are inherited in a pattern of autosomal dominance (Hall, 1990) / (Information about Cancer, 2006). *BRCA1* related cancers tend to occur along familial [verse sporadic] lines. There is also a portion of the population without any *BRCA1* mutations having extremely high incidences of cancer.

BRCA1 type cancer affects 1 in 800 people in the US population (Le, 2003). Approximately 1 in 500 to 1000 people are carriers of at least one *BRCA1* mutation (Genetics, 2002). Genetically linked (germline) breast cancer only accounts for 5% of breast cancer in the general population and upwards of 15-20% for families with a strong history. Women with a family history of both breast and ovarian cancers have much higher incidence of 60-80% breast cancer (Nathanson, 2001) and 40% ovarian cancer.

The relative risk of developing breast cancer peaks at age 35 then decreases slowly for people without a BRCA1 mutation. The cumulative risk is estimated at 65% by age 70 (Antoniou, 2003). This means younger women are more likely to get the disease (Table 16). During a woman's lifetime there is about a 1 in 7 chance of developing a form of breast cancer. A 2002 study looking at 933 ovarian cancer patients found hereditary patients survived on average longer than non-hereditary patients. Chemotherapy and radiation were more effective in these patients and they tended to have a lower recurrence rate (Boyd, 2000).

Men with germline mutations have an estimated 6% risk of breast cancer. This is about 80 times greater than the general population (Nathanson, 2001).

Table 16 – Cumulative Breast Cancer Risk
The cumulative risk of contracting breast cancer decreases with age regardless of population. (Easton, 1995)

Age	**Cumulative Risk**
30 years	3 %
40 years	20%
50 years	51%
60 years	54%
70 years	65%

In 70-80% of *BRCA1*-mutated breast cancers, a mutated *TP53* (tumor suppressor gene) is also found. This compares to a 30% mutation rate in *TP53* alone where a wildtype *BRCA1* is present (Hartman, 2002). *p53* has the ability to provide nucleotide excision repair in the absence

of *GADD45*. This finding suggests the high incidence of mutated *p53* in the presence of mutated *BRCA1* inhibits the alternate repair pathway.

BRCA1 Cancer Causing Mutations

The majority of *BRCA1* mutations are unique to a person or family in that they are not common throughout the general population. Mutations predisposing cancer are inherited in a pattern of autosomal dominance.

"Founder mutations" occur when a population experiences an elevated level of inbreeding due to geographic isolation. Over time, rare mutations recur becoming more common in the general population (Table 17). Population examples of this phenomenon are Jewish, Swedish, Icelanders, Netherlands (Holland). In the Pakistani population, first cousin marriages have a higher incidence of breast cancer. This supports the founder mutation theory. They also have the highest rate of breast cancer among any Asian population (Liede, 2002). The incidence of detection of founder mutations increases over 18x when a family has a history of breast or ovarian cancer (Tonin, 1998).

Mutations occur throughout the *BRCA1* sequence but modifications at the terminal ends are most damaging to function (Table 18). Isolating a specific cancer causing variant is a difficult task. For example, the R1699W mutation found exclusively in Western European populations is associated with cancer while the R1699Q effect is unclear (Vallon-Christersson, 2001). On the contrary, M1652I is frequently observed in cancer patients, however, laboratory testing concluded it does not cause structural deformities (Humphrey, 1997).

The Ashkenazi Jewish population is the most closely studied population and perhaps one of the most impacted by *BRCA1* mutations. Approximately 2% (1 in 100) of Ashkenazi Jews carry a mutation in either *BRCA1* or *BRCA2* (Struewing, 1997). Few mutations outside the 185delAG and 5382insC founder mutations are present within the population (Phelan, 2002). People with either the *BRCA1* 185delAG or *BRCA2* 6174delT mutation have a 36% chance of developing the disease. The occurrence rate is three times higher than in the normal population

(Fodor, 1998). These mutations significantly compromise structural integrity leading to protein product inactivation and cancer.

185delAG is the most frequently occurring mutation and spans all nationalities. Some studies have shown male patients with the 185delAG mutation are twice as likely to develop prostate cancer (Giusti, 2003).

Second to 185delAG is the BRCT reading frame-shift mutation 5382insC. This mutation is seen across equally as many population groups but incidence ranges much higher in the Polish population. In this group, 5382insC accounts for over 50% of mutations followed by C61G at 20% (Gorski, 1003). In Italian families, the 5382insC mutation is coupled almost exclusively with a combination of ovarian and breast cancers (Santarosa, 1999).

The R71G, G1706E, A1708E and 589delCT mutations are found almost exclusively in the Spanish population. Interestingly, heavy Spanish influence in Latin America and the Caribbean has transported 1708E across the Atlantic. The other three mutations are reported far less frequently in this portion of the world. These mutations along with 185delAG account for 50% of Spanish decent linked mutations (Diez, 2003).

Some nonsense mutations cause skipping of one or more exons producing a truncated but functional polypeptide. This event is called Nonsense-mediated Altered Splicing (NAS). The exon skipping occurs when a splicing enhancer (in the coding sequence) is altered through a missense or nonsense mutation (cSNP) (Liu, 2001). As an example, the Q1694X mutation causes skipping of exon 18. This mutation has been reported in 10 patients (BIC). NAS only occurs in cases where the exon skipped contains a multiple of three nucleotides (Liu, 2001).

Native North American women exhibit extremely high incidence of gallbladder, cervix and kidney cancers but far lower breast cancer rates. To date, only two population specific mutations have been identified in Native Americans. The two mutations 1506delA and 1510insG occur in exon 11 at codon 464 (Liede, 2002). Both lead to a premature stop codon found exclusively in Native North Americans.

Table 17 – Founder / Population Specific Mutations

Some mutations have become isolated to a specific population or ethnic group such as R1699W commonly found in Western Europeans. Other mutations such as 185delAG or 5382insC are the only mutations found within the Jewish population.

Mutation	Exon	Population	Citation
M1V/A120G	2	French	(Rostagno, 2003)
185delAG	2	Jewish	(Phelan, 2002)
185delAG	2	Indian	(Kumar, 2002)
185delAG	2	Italian	(Caligo, 1996)
185delAG	2	Spanish	(Diez, 2003)
185delAG	2	Canadian	(Simard, 1994)
185delAG	2	Chilean	(Mew, 2002)
C39R	3	Czech	(Machackova, 2001)
T243C	3	French	(Rostagno, 2003)
252del4	3	Italian	(Caligo, 1996)
IVS5+3A>G	I-5	Belgian	(Claes, 2004)
IVS5+3A>G	I-5	Belgian	(Claes, 2004)
C61G	5	Czech	(Machackova, 2001)
C61G	5	Polish	(Byrski, 2003)
C64R	5	Italian	(Caligo, 1996)
R71G	5	Spanish	(Diez, 2003)
589delCT	8	Spanish	(Diez, 2003)
589delCT	8	Chinese	(Tang 1999)
Q356R	11	Canadian	(Tang 1999)
Q356X	11	Chinese	(Tang 1999)
1499insA	11	Italian	(Caligo, 1996)
2080insA	11	Pakistani	(Liede, 2002)
N723D	11	African Amer	(Olopade, 2003)
2418delA	11	African Amer	(Fetzer, 1999)
2478insG	11	Belgian	(Claes, 2004)
2594delC	11	Danish	(Bergthorsson, 2001)
2800delAA	11	Scottish	(Scottish, 2003)
Q957X	11	Swiss	(Garvin, 1997)
3761delGA	11	Czech	(Machackova, 2001)
P1150S	11	Chinese	(Tang 1999)
K1183R	11	Swiss	
E1221X	11	Belgian	(Claes, 2004)
3819del5	11	Czech	(Foretova, 2004)
3819del5	11	Polish	(Byrski, 2003)
P1238L	11	Jewish	(Phelan, 2002)
3871del4	11	Swiss	(Garvin, 1997)
3889delAG	11	Pakistani	(Liede, 2002)
S1218C	11	Scottish	(Scottish, 2003)
4153delA	11	Polish	(Gorski, 2000)

Mutation	Exon	Population	Citation
IVS12-1643d	I-12	Dutch	(Verhoog, 2001)
4148del4	12	Pakistani	(Liede, 2002)
4248delAG	12	Pakistani	(Liede, 2002)
R1443T	13	Fr. Canadia	(Tonin, 1998)
IVS14-1G>A	I-14	Pakistani	(Liede, 2002)
C4491T	14	French	(Rostagno, 2003)
4572del22	14	Jewish	(Phelan, 2002)
Q1458X	14	Chinese	(Tang, 1999)
S1512I	15	Jewish	(Phelan, 2002)
D1546N	15	Jewish	(Phelan, 2002)
4873delCA	16	Italian	(Caligo, 1996)
4875delCA	16	Italian	(Caligo, 1996)
D1692N	17		
X1694X	18	Western Eu	(Liu, 2000)
C1697R	18	Danish	(Bergthorsson, 2001)
G1706E	18	Spanish	(Diez, 2003)
A1708E	18	N.W. Spain	(Diez, 2003)
5283insC	20	Turkish	(Balci, 1999)
5382insC	20	Swiss	(Garvin, 1997)
5382insC	20	Jewish	(Phelan, 2002)
5382insC	20	German	(Dong, 1998)
5382insC	20	Italian	(Santarosa, 1999)
5382insC	20	Canadian	
5382insC	20	Polish	(Byrski, 2003)
5385insC	20	Czech	(Foretova, 2004)
5385insC	20	Russian	(Gayther, 1997)
5454delC	22	Philippino	(Matsuda, 2002)
R1835X	24	European	(Dong, 1998)
C5622T	24	Turkish	(Balci, 1999)

Table 18 – Cancer Causing Mutations
Mutations with strong links to cancer are listed in this table. To qualify as a 'cancer' mutation, a strong family history of both breast cancer and the specific mutation must exist.

Mutation	Exon	Result	Citation
C24R	2	Alter folding	(Vallon, 2001)
C39S/Y	3	Slight structure alteration, weaken binding	(Vallon, 2001)
C44F	3	Alter folding	(Vallon, 2001)
C47G/F	5	Alter folding	(Vallon, 2001)
C61G	5	Slight structure alteration, weaken binding	(Vallon, 2001)
C64G/R/Y	5	Slight structure alteration, weaken binding	(Vallon, 2001)
Q356R	11	Unclear since not located in domain.	(Dunning, 2000)
P871G	11	Unclear.	(Dunning, 2000)
S1613G	16	Unclear.	(Dunning, 2000)
D1692Y	17	Destabilize core	(Williams, 2003)
Q1694X	18	Skip exon 18	(Liu, 2001)
E1694K	18	Reduce BACH1 binding.	(Liu, 2001)
F1695L	18	Reduce BACH1 binding.	(Liu, 2001)
V1696L	18	Reduce BACH1 binding.	(Liu, 2001)
C1697R	18	Destabilize core	(Williams, 2003)
C1697R	18	Unknown.	(Vallon, 2001)
R1699W	18	Destabilize core	(Williams, 2003)
A1708E	18	Destabilize core	(Williams, 2003)
S1715N	18	Loss of activity	(Vallon, 2001)
S1715R	18	Charge change, destabilize core.	(Williams, 2003)
G1738E	20	Loss of activity	(Vallon, 2001)
H1746D	20	Unknown.	(Liu, 2001)
P1749R	20	Charge change, destabilize core.	(Williams, 2003)
M1775R	21	Causes sterical clash. Disrupts BRCA1-CtIP	(Varma, 2005)
W1837X	24	Truncation	(Vallon, 2001)

BRCA1 Conservation in Other Species

Studying *BRCA1* conservation in other species was far easier than humans alone as many complete sequences are available. All comparisons were conducted using the Human *BRCA1* protein sequence (AAA73985) generated from the *BRCA1* nucleotide sequence U14680. To begin, I BLASTed the primary sequence against the full organism database. This produced 200 results with E-values approaching 0.0 from the NCBI non-redundant protein database using BLOSUM62 and 45. Hand curating these yielded 53 unique species with significant homology. I then re-blasted against the 30 available individual organism databases. The 1853 residue *BRCA1* sequence yielded no homology in viruses, Achaea, bacteria or arthropoda including

Poteome, Aquifex *aeolicus*, Bacillus *subtilis*, C.*elegans*, Drosophila, E.*coli*, HIV, Methanococcus *jannaschii*, Ovis *aries*, Plasmodium *falciparum*, and Saccharomyces *cerevisiae*.

Table 19 – BRCA1 N-terminal Homology

Strict conservation exists in the N-terminal region of *BRCA1* throughout all Animal Chordata. The C3HC4 RING (blue) is highly conserved with no variation in position. Sequences are labeled with species and NCBI GI number.

```
                                    |10       |20       |30       |40       |50
Human|555932           ----MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
Chimpanzee|38503185    ----MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
Gorilla|48479022       ----MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
Orangutan|48479020     ----MDLSAVRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
Chinch Monkey|48479018 ----MDLSAVRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCRFCMLKLLNQK
Mouse|4097808          ----MDLSAVQIQEVQNVLHAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
NorwayRat|6978573      ----MDLSAVRIQEVQNVLHAMQKILECPICLELIKEPVSTQCDHIFCKFCMLKLLNQK
Dog|61740517           ----MDLSADRVEEVQNVLNAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQR
Cow|30466260           ----MDLSADHVEEVQNVLNAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQK
Opossum|77020291       ----MDLPTVTIEEVKNVLIGMQKILECPICLELIKEPVSTTCDHIFCRFCMLKLLSKK
Opossum|62638180       ESKEMDLPTVTIEEVKNVLIGMQKILECPICLELIKEPVSTTCDHIFCRFCMLKLLSKK
Chicken|15081211       ----MDLSVIAIGDVQNVLSAMQKNLECPVCLDVIKEPVSTKCDHVFCRFCMFKLLSRK
Xenopus|15991720       ----MTCSRMDIEGICSVISVMQKNLECPICLELMKEPVATKCDHIFCKFCMLQLLSKK

                           |60       |70       |80       |90       |100      |110
Human|555932           KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFAKKENNSPE
Chimpanzee|38503185    KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFAKKENNSPE
Gorilla|48479022       KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFAKKENNSPE
Orangutan|48479020     KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLQYANSYNFAKKENNSPE
Chinch Monkey|48479018 KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIIHAFQLDTGLQFANSYNFAKKENHSPE
Mouse|4097808          KGPSQCPLCKNEITKRSLQGSTRFSQLAEELLRIMAAFELDTGMQLTNGFSFSKKRNNSCE
NorwayRat|6978573      KGPSQCPLCKNEITKRSLQGSARFSQLVEELLKIIDAFELDTGMQCANGFSFSKKKNSSSE
Dog|61740517           KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIIHAFELDTGLQFADSYNFSKKENNSPE
Cow|30466260           KGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIIHAFELDTGLQFANSYNFSRKEDNSPE
Opossum|77020291       KGPSQCPLCKNNITKRSLRESTRFNQLVEGLLKTIRAFELDTGFQFSNTQDFSKWERRTPE
Opossum|62638180       KGPSQCPLCKNNITKRSLRESTRFNQLVEGLLKTIRAFELDTGFQFSNTQDFSKWERRTPE
Chicken|15081211       KKG-QCPLCKTEVTKRSLKENSRFKQLIEGLLEAISAFELDTGVKFLSSRYFPKTSTEVAT
Xenopus|15991720       KKG-PCPLCKTEVTRRSLQESHRFKLLVEGQLKIIKAFEFDSGYKFFPSQEHTKGLDSTIE
```

As would be expected, significant homologous pairs were found in Chimpanzee, Gorilla, Orangutan and the Chinch Monkey (Table 19, Table 20). However, mouse, rat, dog, opossum, chicken and frog exhibited homology only slightly less than primates. Unfortunately, many other species' sequences did not include amino acids in the N-terminal region. This may be due to the entire gene not being sequenced or a lack of conservation in the N-terminus.

I then blasted each terminal sequence against the entire database to find conservation at the endpoints and also in other genes. The 116 amino acids of the RING domain exhibited identical homology to Chimpanzee and Gorilla. The RING domain is essentially 100% conserved in this region with only a few minor mutations occurring in the Orangutan and Chinch Monkey sequences.

Table 20 – BRCA1 C-terminal Homology
The *BRCA1* C-terminus exhibits less homology than the N-terminus. Primates display near 100% homologous sequences while other Chordata . Highly conserved backbone amino acids exit at 1699-1704, 1774, 1775, 1835 and 1838 as pictured in blue. Sequences are labeled with species and NCBI GI number.

```
                                         |1640     |1650     |1660       |1670       |1680
Human|555932              ----------ASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVM
Chimpanzee|38503185       ----------ASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVM
Gorilla|48479022          ----------ASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVM
Orangutan|48479020        ----------ASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVM
Chinch Monkey|48479018    ----------ASTERVNKRMSLVVSGLTPEEFMLVYKFARRYHIALTNLISEETTHVVM
Mouse|4097808             ----------SSEERADRDISMVVSGLTPKEVMTVQKFAEKYRLTLTDAITEETTHVII
NorwayRat|6978573         ----------SPKERAERDISMVVSGLTPKEVMIVQKFAEKYRLALTDVITEETTHVII
Dog|61740517              ----------SSTRGVNKRISMVASGLTPKEFMLVHKFARKHHISLTNLISEETTHVIM
Cow|30466260              ----------SSTERSKKRLSMVASGLTPKELMLVQKFARKHHVTLTNLITEETTHVIM
Opossum|77020291          ----------KKGGTGNRKISLVSSGLTPKENMLVQKFARKTHSTVSHQITEGTTHVIM
Opossum|62638180          ----------KKGGTGNRKISLVSSGLTPKENMLVQKFARKTHSTVSHQITEGTTHVIM
Chicken|15081211          HNVAGKENAASSGTTCRTEMSIVASGLNQSEHLMVQKFARKTQSTFSNHITDGTTHVIM
Xenopus|15991720          --------FKSPVVSSRRNLSFVASGLNQCEMALVQRFSKTTQSILSSRITDSTTHVIM

                          |1690     |1700      |1710      |1720      |1730      |1740
Human|555932              KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGP
Chimpanzee|38503185       KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGP
Gorilla|48479022          KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKEGKMLNEHDFEVRGDVVNGRNHQGP
Orangutan|48479020        KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGP
Chinch Monkey|48479018    KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGP
Mouse|4097808             KTDAEFVCERTLKYFLGIAGGKWIVSYSWVVRSIQERRLLNVHEFEVKGDVVTGRNHQGP
NorwayRat|6978573         KTDAEFVCERTLKYFLGIAGGKWIVSYSWVIKSIQERKLLSVHEFEVKGDVVTGSNHQGP
Dog|61740517              KTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKILDEHDFEVRGDVVNGRNHQGP
Cow|30466260              KTDPEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKEGKMLDEHDFEVRGDVVNGRNHQGP
Opossum|77020291          KTDAEFVCERTLKYFLGIAGGKWVVSFLWVVQSFKEGKMLPECDFEVRGDVINGRNHRGP
Opossum|62638180          KTDAEFVCERTLKYFLGIAGGKWVVSFLWVVQSFKEGKMLPECDFEVRGDVINGRNHRGP
Chicken|15081211          KTDEELVCERTLKYFLGIAGRKWVVSYQWIIQSFKEGRILDEEHFEVKGDVINGRNHQGP
Xenopus|15991720          KTDAELVCERTLKYFQGIASRKWVVSYEWVVQSFREGQILDEYDFEVKGDVINGRNHRGP

                          |1750     |1760      |1770      |1780      |1790      |1800
Human|555932              KRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVV
Chimpanzee|38503185       KRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVV
Gorilla|48479022          KRARESQDRKIFRGLDICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVV
Orangutan|48479020        KRARESQDRKIFRGLEICCYGPFTNMPTDQLEWIVQLCGASVVKELSSFTLGTGVHPIVV
Chinch Monkey|48479018    KRARESPDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGFHPIVV
Mouse|4097808             RRSRESRE.KLFKGLQVYCCEPFTNMPKDELERMLQLCGASVVKELPSLTHDTGAHLVVI
NorwayRat|6978573         RRSRESQE.KLFEGLQIYCCEPFTNMPKDELERMLQLCGASVVKELPLLTRDTGAHPIVL
Dog|61740517              KRARESQD-KIFRGLEICCYGPFTNMPTDQLEWMVHLCGASVVKEPSLFTLSKGTHPVVV
Cow|30466260              KRARESRDKKIFKGLEICCYGPFTNMPTDQLEWMVQLCGASVVKEPSSFTPDQGTHPVVV
Opossum|77020291          ERARESQGMKIFRGLEICCYGPFTDMSTDQLEWMVQLCGASVVKKPSSLRFRVGSSPVVV
Opossum|62638180          ERARESQGMKIFRGLEICCYGPFTDMSTDQLEWMVQLCGASVVKKPSSLRFRVGSSPVVV
Chicken|15081211          KRARQSPAEKIFKDFEICCCGPFTDMTTGHLEWIVELCGASVVKQLHLFTHKVNSTAVVV
Xenopus|15991720          RRSRLGSDGLLLIDFEICFFGSFTDMTLDDLEWMVSECGATVVKKLQFFKKKHNVTSLVI

                          |1820     |1820      |1830      |1840      |1850      |1860
Human|555932              VQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY-------
Chimpanzee|38503185       VQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY-------
Gorilla|48479022          VQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY-------
Orangutan|48479020        VQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY-------
Chinch Monkey|48479018    VQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPHSHY-------
Mouse|4097808             VQPSAWTEDSNCPDIGQLCKARLVMWDWVLDSLSSYRCRDLDAYLVQNITCDSSEPQDSND
NorwayRat|6978573         VQPSAWTEDNDCPDIGQLCKGRLVMWDWVLDSISVYRCRDLDAYLVQNITCGRDGSEPQDS
Dog|61740517              VQPDAWTEDSGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPRTAADS-----
Cow|30466260              VQPDAWTEDAGFHVIGQMCEAPVVTREWVLDSVALYQCQELDTYLVP--------------
Opossum|77020291          VQPDAWEDDSSFQEIGLVCEAPVVTREWVLDSVACYQRQELDTYLISQTSPSLC-------
Opossum|62638180          VQPDAWEDDSSFQEIGLVCEAPVVTREWVLDSVACYQRQELDTYLISQTSPSLC-------
Chicken|15081211          VQPDAWMEGTSYEAIQRKNNVAVVTREWVLDSVACFECQELDAYLVSQD------------
Xenopus|15991720          VQPDASTEVRDYTEIRKKHKALVVTREWLMDSVATYRLQKFDAYLA---------------
```

In the remainder of the species the major cysteine and histidine binding sites are completely conserved. The most N-terminal RING is strictly conserved especially at the helix

and beta-strands. Slightly less homology exists throughout the remaining sequence with most variation in exons 5 and 6.

Similarly, the BRCT is also highly conserved with all four primates almost 100% identical while other species lose homology approaching the terminal end. The linker region between the two repeats is the least conserved portion. Amino acids along the surface of the tandem repeat cleft junction are highly conserved. See Table 20 codons 1699-1704 and 1774-1775, 1835 and 1838.

BRCA1 Central Region Conservation

To demonstrate conservation in the central region of *BRCA1* I blasted the exon 11 portion of the sequence. This region ranges from residue 263 to 1405. Conserved areas are found sporadically through the range but are usually no longer than five amino acids in length. These highly conserved areas are typically separated by wide areas of divergence.

In exon 11 I found three conserved areas. The first is the 823-900 regions listed in Table 21. This appears to be the most highly conserved region with 15 closely packed amino acids found in all 12 species. Unfortunately there are no known binding or phosphorylation sites in this region.

Other conserved areas are the 1100-1250 and 1350-1400 ranges. Within these areas are mammalian conserved Serine 988, 1387, 1457, 1524 and Tyrosine 1394. These are well known phosphorylation sites for *ATM* and *ATR* (Okada, 2003).

Table 21 – BRCA1 Exon 11
The central region exhibiting the highest specie homology is located at amino acids 823-900 (Human). The sixteen codons pictured in blue display a compact region of conservation.

```
                       |829        |830      |840      |850      |860      |870      |880
Human|555932           FKYPLGHEVN-HSRETSIEMEESELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLK
Chimp|38503185         FKYPLGHEVN-HSRETSIEMEESELDAQYLQNTFKVSKRQSFALFSNPGNPEEECATFSAHCRSLK
Gorilla|48479022       FKYPLGHEVN-HSRETSIEMEESELDAQYLQNTFKVSKRQSFALFSNPGNPEEECATFSAHSRSLK
Orang.|48479020        FKYPLGHEVN-HSQETSIEMEESELDTQYLQNTFKVSKRQSFALFSNPGNPEEECATFSAHSRSLK
Chinch|48479018        FKYPLGSEVN-HSQETSIEIEESELDTQYLQNTFKVSKRQSFALFSNPGNPEEECATFSAHSRSLK
Mouse|4097808          LKPPLRHALN-LSQEK-VEMEDSELDTQYLQNTFQVSKRQSFALFSKPRSPQKDC----AHSVPSK
NorwayRat|6978573      FKHPLRHELN-HNQET-IEMEDSELDTQYLQNTFQVSKRQSFALFSKLRSPQKDCTLVGARSVPSR
Dog|61740517           FVVPLTCKDN-HTQETSIEMEESELDTQCLRNMFKVSKRQSFALFSYPRDPEEDCVTVCPRSGAFG
Cow|30466260           FQDLLGHDINYVIQETSREMEDSELDTQYLQNTFKASKRQTFALFSNPGNPQKECATVFAHSGSLR
Opossum|77020291       LRGLMRQGVK-NASETTTEMEDSELDTQYLQNTFKRSKRQTFALGS---SPRQECMKPCAISQALH
Opossum|62638180       LRGLMRQGVK-NASETTTEMEDSELDTQYLQNTFKRSKRQTFALGS---SPRQECMKPCAISQALH
Chicken|15081211       FRIGKSPMAK-NASEFTMEAEDSELDMQYLRNIFRSSKRQSFSLYP---TPMKACTTDDVASEKLN
Xenopus|15991720       FRIGKSPMAK-NASEFTMEAEDSELDMQYLRNIFRSSKRQSFSLYP---TPMKACTTDDVASEKLN
```

BRCA1 Mouse Model

Mus musculus BRCA1 maps to chromosome 11 at 60.5cM adjacent to *Corn2* (Corneal disease 2), see Figure 16. Similar to humans, mouse *BRCA1* is in close proximity to the adjacent to *NBR1* gene. A separation of less than 300bp allows *BRCA1* expression to be driven by a shared promoter with NBR1 (Whitehouse, 2004). Mouse *BRCA1* operates in similar pathway to human *BRCA1* however, some significant functional differences occur in the presence of a haploid gene.

Mouse *BRCA1* homolog is syntenic with Human *BRCA1* meaning the gene order is conserved over evolutionary distances. In terms of identity, approximately 87% of exons are homologous to *BRCA1* while at the amino acid level identity drops to 55%.

Beyond sequence homology it is difficult to study the effect of single amino mutations in mice. Mouse models have been tricky to develop since homozygosity is lethal during embryonic development and heterozygous mice are not cancer prone. Heterozygous mice are normal functioning at 11 months and tumor free. However, when a sequence mutation is introduced on one allele while the other allele is deactivated they die shortly after conception. *BRCA1* deficient mice are unable to repair double-strand breaks by homologous recombination. These mice are hypersensitive to IR and are unable to carry out transcription repair of damaged DNA (Whitehouse, 2004).

Mice with exon 5 & 6 deletions exhibit severe developmental defects before reaching mortality at 7 days into gestation. This may be a consequence of lower cyclin-E levels at the G1/S transition resulting in slower cell growth (cell proliferation) and embryo size. Early embryonic arrest suggests exons 5 and 6 are responsible for cell proliferative burst (Haken, 1996).

Mice with exon 11 deletion survive to about 13 days but exhibit spina bifida and other neuroepithelial defects (Gowen, 1996). These cells have an intact G1/S phase checkpoint but a defective G2/M checkpoint. Embryos surviving longer than 13 days have extensive chromosomal abnormalities such as multiple centrosomes. This suggests *BRCA1* not only

regulates cell cycle checkpointing but also genomic stability through centrosome replication (Xu, 1999). Interestingly, a combination of homozygous exon 11 deletion and heterozygous removal of a *p53* allele allowed for full embryo development. However, mice with this chromosomal abnormality began to develop tumors in 6-12 months after birth (Xu, 2001).

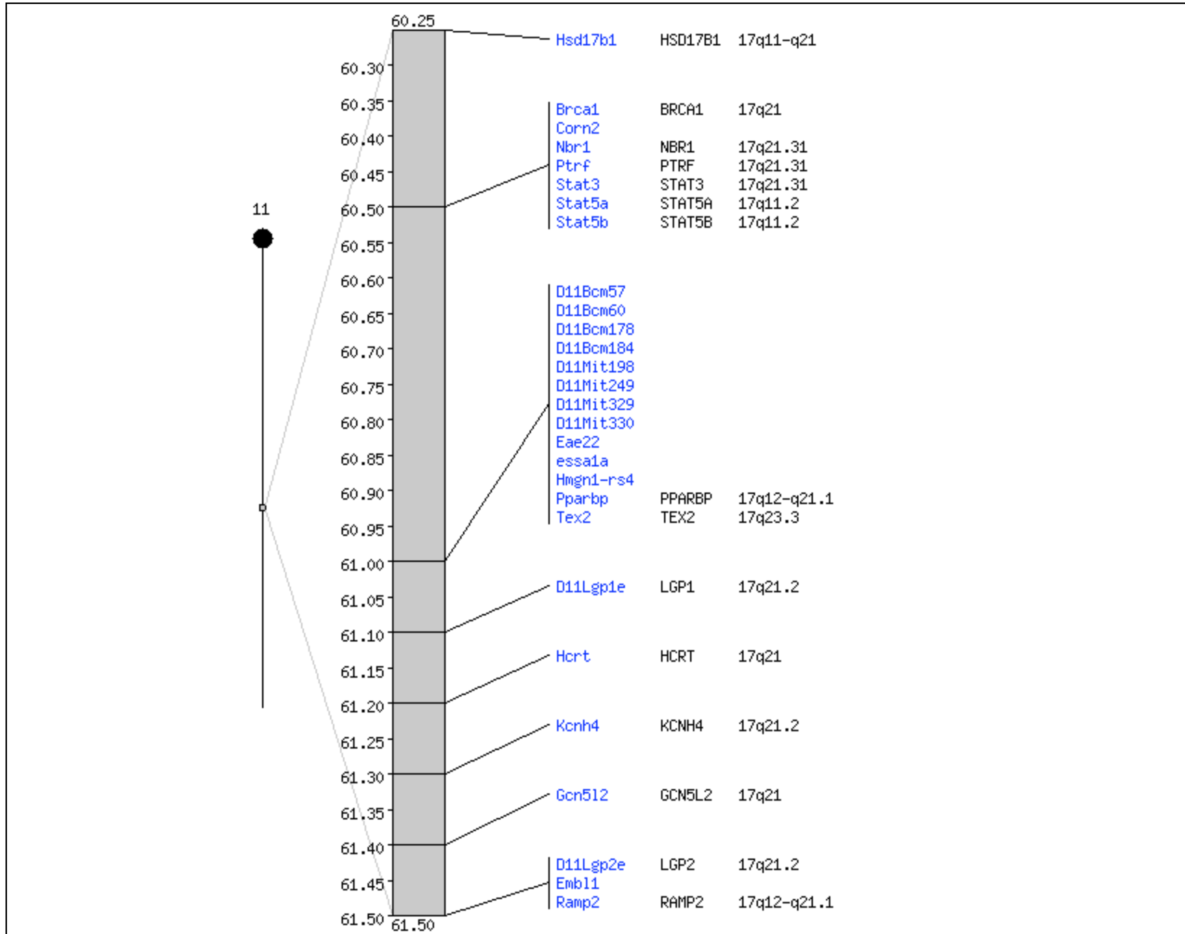

Figure 16 – Mouse Comparative Map
Mouse *BRCA1* maps to chromosome 11 adjacent to the *Corn2* and *NBR1* at 60.50cM. Source: Mouse Genome Informatics

Rattus norvegicus

Similar to mouse, *Rattus norvegicus* shares approximately 83% sequence similarity at the terminal ends. Rat codons 1589-1817 map to the BRCT 1646-1959 region in humans.

To study *BACH1* binding functionality with Rat *BRCA1* and investigate structural different, the 1L0B model was created with *BRCA1* bound to p53. Structural analysis shows rat hydrophobic residues Arg1645, Leu1651, and Ala1654 are conserved between human and rat and

hydrogen bond contacts Arg1645, Asp1785, Arg1727 and Ile1626 occur at same positions (Joo, 2002).

Human cancer causing mutations E1694K, F1695L, and V1696L were replicated in rat at Glu1640, Phe1641 and Val1642. Protein products with a single mutation caused a deformed BRCT β3 sheet. This resulted in 30-50% less *BACH1* binding activity (Joo, 2002). Rat N1719R is a less severe disruption as it falls on a turn and not a helix of sheet.

Canis familiaris

Overall, *Canine BRCA1* shares 71% sequence similarity with the human homolog. At both N and C-terminus the identity rises to 84%. The central region, including exon 11 has about 70% homology but the longest 'domain' is no more than 21 acids in length (Szabo, 1996). Canine *BRCA1* is relatively the same length as the Human strand with only forty-five additional bases due to small insertions.

Little structural comparison research has been done on the consequences of mutations carried into the canine sequence. However an interesting repeat occurs in the central region. At (human) 1758, Canine *BRCA1* contains an 18bp duplication of the previous five amino acids. This rare mutation is well conserved against other Canine database samples.

Table 22 – Human-Canine duplication
A five amino acid sequence duplication occurs in canine at 1758. This duplication is highly conserved throughout all NCBI Canine samples.

```
                             |1740     |1750           |1760     |1770     |1780
Human|555932           DFEVRGDVVNGRNHQGPKRARESQDR-----KIFRGLEICCYGPFTNMPTDQLEWMVQLC
Dog|61740517           DFEVRGDVVNGRNHQGPKRARESQDRESQDRKIFRGLEICCYGPFTNMPTDQLEWMVHLC
```

Drosophila

Table 23 – Human-Drosophila
Similar to Animal Chordata, Drosophila conserves the C3HC4 Zinc-finger at critical site C24, C27, C39, H41, C44, C61, and C64 (blue). Other conserved amino acids are shown in grey.

```
                          |20       |30         |40       |50       |60
1JM7|15988069          AMQKILECPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKKGPSQCPLCKNDIT-KRSLQESTRF
DRING|2388783          SLHSELMCPICLDMLKKTMTTKECLHRFCSDCIVTALRS--GNKECPTCRKKLVSKRSLRADPNF
```

The Drosophila RING sequence contains large gaps of sequence divergence with only 34% amino acid identity to the human homolog. However, all eight amino acids composing the Zinc

finger loops are conserved. At the C-terminal about the same homology exists but significant binding sites and turns are not distinguishable between the two sequences.

Overall Conservation

Primate conservation ranged between 98-100% within the RING and C-terminal regions. Over the entire sequence homology dropped to only 96% within all coding regions. Other species express more variation overall but maintained high scores in the terminal regions (Table 24).

The near 100% conservation of binding sites suggests a possibility of evolutionary conserved orthologs. The high identity expressed in the terminal ends appears to have evolved under negative selection which is why they are almost identical. On the contrary, the high variability in the central region shows diversification or positive selection.

Table 24 – Percent Identity to Human BRCA1
Primate BRCA1 is nearly 100% conserved among all species. Both terminal regions are conserved almost equally in each species though overall conservation drops steadily outside Primates.

	BRCN	**BRCT**	**Overall**
Chimpanzee	100	100	98
Gorilla	100	100	98
Orangutan	99	99	97
Chinch Monkey	99	99	96
Mouse	80	83	55
Rat	86	83	55
Dog	91	84	74
Cow	91	89	71
Opossum	89	84	45
Chicken	62	62	29
Xenopus	60	53	34
Drosophila	34	34	26

Comparing Table 19 and Table 20, the BRCN region conserves more mutations than the BRCT-terminal. Patient derived mutations I15L, I42V and D67E are frequently observed in the less evolved mammals and amphibians. In BRCT, F1695L, H1672Q, G1764D area few mutation conserved only between the Chicken and Xenopus sequences. These mutations are attributed to SNPs.

There are five conserved regions in BRCT, N-terminus/β1, α1/β2, β3, β4/α3 and the C-terminus. Most variation occurs at α2 helix and surface loops between Alpha1 & Beta2 (C1) and Beta2 & Beta3 (C2). Mutations in this region are unlikely to alter folding properties as the loops are far from the backbone structure (Zhang, 1999).

Discussion

Although *BRCA1* is a highly studied gene and a wealth of information exists, it was extremely difficult finding reliable and accurate research. Literature searches revealed a considerably amount of information, however, many papers were outdated. For example, I found numerous papers describing the same pathway but with major discrepancies. Instead of compiling the *BRCA1* pathway during a few days, my research spanned several weeks substantiating all reactions with scientific evidence.

An example of this is the role of *BRCA1* activation of *Chk1*. Protein Lounge and several articles provide considerable evidence for this claim. However the NCBI OMIN page for *CHEK1* (approved gene symbol) does not contain a single mention of *BRCA1*. A 2002 paper by Lee validates the interaction of these two proteins so it's puzzling that the official *Chk1* reference does not note this information.

Furthering the confusion is pathway activation by either *ATM* or *ATR*. Some papers discussed *ATM* activation of *Chk1* while others conclude *ATR* as the stimulus. It was only in 2002 that Falck's identification of both proteins as activators clarified these pathways. He discovered that either kinase can phosphorylate *Chk1* and *NBS1-FANCD2* substrates under differing UV and IR conditions.

Another challenge was finding recent prediction and modeling software. Many of the programs developed at colleges look promising but use outdated techniques or algorithms and are often poorly documented. CBSU's LOOPP program is a good example.

LOOPP is a 3D structure prediction program producing PDB output based on sequence alignment. Since this program uses different scoring metrics than Swiss-Model it would make for

a good comparison. Sequence submission is through a web interface with result returned via email. The email contains a compressed archive containing twenty text files along with DOS executables. LOOPP's documentation does not explain how to use the executables nor clearly interpret the results.

In this research project I studied the structure and function of the proteins produced by the *BRCA1* gene. An attempt was made to draw genetic relationships between cancer patients of diverse ethnicities and geographic locations. My research found specific mutations isolated to single populations such as Western European, Dutch, Native North Americans, and Ashkenazi Jewish populations. Taken together, my findings suggest a strong correlation between geographic isolation and mutation exclusiveness. Furthermore *BRCA1* sequence homology and mutation conservation are exhibited among primates in addition to other species.

Bibliography

American Cancer Society (2005). What are the Key Statistics for Breast Cancer. http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_breast_cancer_5.asp.

Andreassen, PR.; D'Andrea,AD.; Taniquchi,T.; "ATR couples FANCD2 monoubiquitination to the DNA-damage response." *Genes and Development*. 18:1958-1963 (2004).

Antoniou, A.; Pharoah, PD.; Narod, S.; Risch, HA.; Eyfjord, JE.; Hopper, JL.; Loman, N.; Olsson, H.; Johannsson, O.; Borg, A.; Pasin,zi, B.; Radice, P.; Manoukian, S.; Eccles, DM.; Tang, N.; Olah, E.; Anton-Culver, H.; Warne,r E.; Lubinski, J.; Gronwald, J.; Gorski, B.; Tulinius, H.; Thorlacius, S.; Eerola, H.; Nevanlinna, H.; Syrjakoski, K.; Kallioniemi, OP.; Thompson, D.; Evans, C.; Peto, J.; Lalloo, F.; Evans DG.; Easton DF.; "Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies." *American Journal of Human Genetics*. 72:1117-1130 (2003).

Balci, A.; Huusko P.; Paakkonen K.; Launonen V.; Uner A.; Ekmekci A.; Winqvist R.; "Mutation analysis of BRCA1 and BRCA2 in Turkish cancer families: a novel mutation BRCA2 3414del4 found in male breast cancer." *European Journal of Cancer*. 35:707-710 (1999).

Banin, S.; Moyal L.; Shieh S.; Taya Y.; Anderson CW.; Chessa L.; Smorodinsky NI.; Prives C.; Reiss Y.; Shiloh Y.; Ziv Y.; "Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science*. 10:1674-1677, 1998.

Bergthorsson, J. BRCA1 and BRCA2 mutation status and cancer family history of Danish women affected with multifocal or bilateral breast cancer at a young age. *Journal of Medical Genetics*. 38:361-368, 2001.

Biocarta Pathways (2006). *BioCarta*. http://www.biocarta.com/genes/index.asp

Biomolecular Interaction Network Database (BIND). Unleashed Informatics. http://bond.unleashedinformatics.com/

Bork, P.; Hofmann, K.; Bucher, P.; Neuwalkd, AF.; Altschul, SF.; Koonin, EV.; " superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins." *The FASEB Journal*. 11:68-76 (1997).

Boyd, J.; Sonoda,Y.; Federici, MG.; Bogomolniy, F.; Rhei, E.; Maresco, DL.; Saigo, PE.; Almadrones, LA.; Barakat, RR.; Brown, CL.; Chi, DS.; Curtin, JP.; Poynor, EA.; Hoskins, WJ.; "Clinicopathologic features of BRCA-linked and sporadic ovarian cancer." *JAMA*. 283:2260-2265 (2000).

Brown, AL.; Lee, CH.; Schwarz, JK.; Mitiku, N.; Piwnica-Worms, H.; Chung, JH.; "A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage." *Proceedings of the National Academy of Sciences*. 96:3745-3750 (1999).

Brown, A. Lo, LJ.; Catteau, A.; Xu, CF.; Lindeman, GJ.; Hodgson, S.; Solomon, E.; "Germline BRCA1 promoter deletions in UK and Australian familial breast cancer patients." *Human Mutation.* 19:435-442 (2002).

Brzovic, PS.; Keefee, JR.; Nishikawa, H.; Miyamoto, K.; Fox, D.; Fukuda, M.; Ohta, T.; Klevit, R.; "Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex." *Proceedings of the National Academy of Sciences*. 100:5646-5651 (2003).

Brzovic, PS.; Rajagopal, P.; Hoyt, DW.; King, MC.; Klevit, RE.; "Structure of a BRCA1-BARD1 heterodimeric RING-RING complex." *Nature Structural Biology*. 8:833-837 (2001)

Byrski, T. "Germline mutation of BRCA1 gene in Polish families with strong aggregation of breast and/or ovarian cancer based on coding sequence analysis using the SSCP method." *Annales Academiae Medicae Stetinensis*. 49:27-43 (2003).

Caligo, M.; Ghimenti, C.; Cipollini, G.; Ricci, S.; Brunetti, I,; Marchetti, V.; Olsen, R.; Neuhausen, S.; Shattuck-Eidens, D.; Conte, PF.; Skolnick, MH.; Bevilacqua, G.; "BRCA1 germline mutational spectrum in Italian families from Tuscany: a high frequency of novel mutations." *Oncogene*. 13:1483-1488 (1996)

Callebaut, I.; Mornon, JP.; "From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair." *Federation of European Biochemical Societies*. 400:25-30 (1997).

Cantor, SB.; Bell, DW.; Ganesan, S.; Kass, EM.; Drapkin, R.; Grossman, S.; Wahrer, DC.; Sqroi, DC.; Lane, WS.; Haber, DA.; Livingston, DM.; "BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function." *Cell*. 105:149-160 (2001).

Chai, YL.; Cui, J.; Shao, N.; Shyam, E.; Reddy, P.; Rao, VN.; "The second BRCT domain of BRCA1 proteins interacts with p53 and stimulates transcription from the p21WAF1/CIP1 promoter." *Oncogene*. 18:263-268 (1999).

Chehab, NH.; Malikzay, A.; Appel, M.; Halazonetis, TD.; "Chk2/hCds1 functions as a DNA damage checkpoint in G1 by stabilizing p53." *Genes and Development*. 14:278-288 (2000).

Chen, A.; Kleiman, FE.; Manley, JL.; Ouchi, T.; Pan, ZQ.; "Autoubiquitination of the BRCA1*BARD1 RING ubiquitin ligase." *Journal of Biological Chemistry*. 277:22085-22092 (2002).

Chen, CF.; Chen, PL.; Zhong, Q.; Sharp, ZD.; Lee, WH.; "Expression of BRC repeats in breast cancer cells disrupts the BRCA2-Rad51 complex and leads to radiation hypersensitivity and loss of G(2)/M checkpoint control." *Journal of Biological Chemisttry*. 274:32931-32935 (1999).

Chen, J.; Silver, DP.; Walpita, D.; Cantor, SB.; Gazdar, AF.; Tomlinson, G.; Couch, FJ.; Weber, BL.; Ashley, T.; Lingston, DM, Scully, R.; "Stable Interaction between the Products of the BRCA1 and BRCA2 Tumor Suppressor Genes in Mitotic and Meiotic Cells." *Molecular Cell*. 2:317-28 (1998).

Chen, Y.; Chen, CF.; Riley, DJ.; Allred, DC.; Chen, PL.; Von Hoff, D.; Osborne, CK.; Lee, WH.; "Aberrant subcellular localization of BRCA1 in breast cancer." *Science*. 3:789-91 (1995).

Chen, Y.; Farmer, AA.; Chen, CF.; Jones, DC.; Chen, PL.; Lee, WH.; "BRCA1 is a 220-kDa nuclear phosphoprotein that is expressed and phosphorylated in a cell cycle-dependent manner." *Cancer Research Journal*. **56**:2168-72 (1996).

Choudhury, AD. Xu, H.; Modi, AP.; Zhang, W.; Ludwig, T.; Baer, R.; "Hyperphosphorylation of the BARD1 turmor suppressor in mitotic cells." *Journal of Biological Chemistry*. **280**:24669-24679 (2005).

Claes, K.; Poppe, B.; Coene, I.; Paepe, AD.; Messiaen, L.; "BRCA1 and BRCA2 germline mutation spectrum and frequencies in Belgian breast/ovarian cancer families." *British Journal of Cancer*. **90**:1244-1251 (2004).

Clamp, M.; Cuff, J.; Searle, SM.; Barton, GJ.; "The Jalview Java Alignment Editor." *Bioinformatics*. **20**:426-7 (2004).

Coene, E.; Van Oostveldt, P.; Willems, K.; van Emmelo, J.; De Potter, CR.; "BRCA1 is localized in cytoplasmic tube-like invaginations in the nucleus." *Nature*. 116:122-124 (1997).

D'Andrea, AD.; Grompe, M.; "The Fanconi anaemia/BRCA1 pathway." *Nature Reviews Cancer*. 3:23-34 (2003).

Deffenbaugh, AM.; Frank, TS.; Hoffman, M.; Cannon-Albright, L.; Neuhausen, SL.; "Characterization of common BRCA1 and BRCA2 variants." *Genetic Testing*. 6:119-121 (2002).

Diez, O.; Osorio A.; Duran M.; Martinez-Ferrandis JI.; de la Hoya M.; Salazar R.; Vega A.; Campos B.; Rodriguez-Lopez R.; Velasco E.; Chaves J.; Diaz-Rubio E.; Jesus Cruz J.; Torres M.; Esteban E.; Cervantes A.; Alonso C.; San Roman JM.; Gonzalez-Sarmiento R.; Miner C.; Carracedo A.; Eugenia Armengod M.; Caldes T.; Benitez J.; Baiget M.; "Analysis of BRCA1 and BRCA2 genes in Spanish breast/ovarian cancer patients: a high proportion of mutations unique to Spain and evidence of founder effects." *Human Mutation*. 22:301-312 (2003).

Dong, J.; Chang-Claude, J.; Wu, Y.; Schumacher, V.; Debatin, I.; Tonin, P.; Royer-Pokora, B.; "A high proportion of mutations in the BRCA1 gene in German breast/ovarian cancer families with clustering of mutations in the 3' third of the gene." *Human Genetics*. 103:154-161 (1998).

Dunnen, JT.; Antonarakis, SE.; "Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion." *Human Mutation*. 15:7-12 (2000).

Dunning, A.; Chiano, M.; Smith, NR.; Dearden, J.; Gore, M.; Oakes, S.; Wilson, C.; Stratton, M.; Peto, J.; Easton, D.; Clayton, D.; Ponder, BA.; "Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population." *Human Molecular Genetics*. 6:285-289 (1997).

Easton, DF.; Ford, D.; Bishop, DT.; "Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium." *American Journal of Human Genetics*. 56:265-271 (1995).

Eppig, JT.; Bult, CJ.; Kadin, JA.; Richardson, JE.; Blake JA.; and the members of the Mouse Genome Database Group. "The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology." *Nucleic Acids Res*. 33: D471-D475 (2005).

Fabbro, M.; Savage, K.; Hobson, K.; Deans, AJ.; Powell, SN.; McArthur, GA.; Khanna, KK.; "BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage." *Journal of Biological Chemistry*. 279:31251-31258 (2004).

Falck, J.; Petrini, JH.; Williams, BR.; Lukas, J.; Bartek, J.; "The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways." *Nature Genetics*. 30:290-294 (2002).

Fan, S.; Wang, J.; Yuan, R.; Ma, Y.; Meng, Q.; Erdos, MR,; Pestell, RG.; Yuan, F.; Auborn, KJ.; Goldberg, ID.; Rosen, RM.; "BRCA1 inhibition of estrogen receptor signaling in transfected cells." *Science.* 284: 1354-1356 (1999).

Fan, S.; Jin, S.; Tong, T.; Zhao, H.; Fan, F.; Antinore, MJ.; Rajasekaran, B.; Wu, M.; Zhan, Q.; "BRCA1 regulates GADD45 through its interactions with the OCT-1 and CAAT motifs." *Journal of Biological Chemistry*. 277:8061-8067 (2002).

Fetzer, S.; Tworek, HA.; Piver, MS.; DiCioccio, RA.; "Classification of IVS1-10T→C as a polymorphism of BRCA1." *Cancer Genetics and Cytogenetics*. 113:58-64 (1999).

Figge, MA.; Blankenship, L.; "Missense mutations in the BRCT domain of BRCA-1 from high-risk women frequently perturb strongly hydrophobic amino acids conserved among mammals." *Cancer Epidemiol Biomarkers Preview*. 13:1037-1041 (2004).

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

Folias, A.; Matkovic, M.; Bruun, D.; Reid, S.; Hejna, J.; Grompe, Ml.; D'Andrea, A.; Moses, R.; "BRCA1 interacts directly with the Fanconi anemia protein FANCA." *Human Molecular Genetics*. 11:2591-2597 (2002).

Fodor, FH.; Weston, A.; Bleiweiss, IJ.; McCurdy, LD.; Walsh, MM.; Tartter, PI.; Brower, ST.; Eng, CM.; "Frequency and carrier risk associated with common BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer patients." *American Journal of Human Genetics*. 63:45-51 (1998).

Foray, N.; Marot, D.; Randrianarison, V.; Venezia, ND.; Picard, D.; Perricaudet, M.; Favaudon, V.; Jeggo, P.; "Constitutive association of BRCA1 and c-Abl and its ATM-dependent disruption after irradiation." *Molecular Cell Biology*. 22:4020-4032 (2002).

Machackova, E.; Damborsky, J.; Valik, D.; Foretova, L.; "Novel germline BRCA1 and BRCA2 mutations in breast and breast/ovarian cancer families from the Czech Republic." *Human Mutation*. 23:397-398 (2004).

Gaiser, OJ.; Ball, LJ.; Schmieder, P.; Leitner, D.; Strauss, H.; Wahl, M.; Kuhne, R.; Oschkinat, H.; Heinemann, U.; "Solution structure, backbone dynamics, and association behavior of the C-terminal BRCT domain from the breast cancer-associated protein BRCA1." *Biochemistry*. 43:15983-15995 (2004).

Garcia, V.; Garcia, JM.; Pena, C.; Silva, J.; Dominquez, G.; Rodriquez, R.; Maximiano, C.; Espinosa, R.; Espana, P.; Bonilla, F.; "The GADD45, ZBRK1 and BRCA1 pathway: quantitative analysis of mRNA expression in colon carcinomas." *Journal of Pathology*. 206:92-99 (2005).

Garcia-Higuera, I.; Taniquchi, T.; Ganesan, S.; Meyn, MS.; Timmers, C.; Hejna, J.; Grompe, M.; D'Andrea, AD.; "Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway." *Molecular Cell*. 7:249-262 (2001).

Garvin, A.; Attenhofer-Haner, M.; Scott, RJ.; "BRCA1 and BRCA2 mutation analysis in 86 early onset breast/ovarian cancer patients." *Journal of Medical Genetics*. 34:990-995 (1997).

Gatei, M.; Zhou, BB.; Hobson, K.; Scott, S.; Young, D.; Khanna, KK.; "Ataxia telangiectasia mutated (ATM) kinase and ATM and Rad3 related kinase mediate phosphorylation of Brca1 at distinct and overlapping sites. In vivo assessment using phospho-specific antibodies." *Journal of Biological Chemistry*. 276:17276-17280 (2001).

Gatei, M.; Scott, SP.; Filippovitch, I.; Soronika, N.; Lavin, MF.; Weber, B.; Khanna, KK.; "Role for ATM in DNA damage-induced phosphorylation of BRCA1". *Cancer Research*. 60:3299-3304 (2000).

Gayther, S.; Harrington P,; Russell, P,; Kharkevich, G.; Garkavtseva, RG.; Ponder, BA.; "Frequently occurring germ-line mutations of the BRCA1 gene in ovarian cancer families from Russia." *American Journal of Human Genetics*. 60:1239-1249 (1997).

Geisler, JP.; Hatterman-Zogg, MA.; Rathe, JA.; Buller, RE.; "Frequency of BRCA1 dysfunction in ovarian cancer." *Journal of the National Cancer Institute*. 94:61-7 (2002).

Genetics Home Reference, 2005. Breast Cancer [on-line]. Available from http://ghr.nlm.nih.gov/condition=breastcancer.

Giusti, RM.; Rutter, JL.; Duray, PH.; Freedman, LS.; Konichezky, M.; Fisher-Fischbein, J.; Greene, MH.; Maslansky, B.; Fischbein, A.; Gruber, SB.; Rennert, G.; Ronchetti, RD.; Hewitt, SM.; Struewing, JP.; Iscovich, J.; "A twofold increase in BRCA mutation related prostate cancer among Ashkenazi Israelis is not associated with distinctive histopathology." *Journal of Medical Genetics*. 40:787-792 (2003).

Gorski, B.; Jakubowska, A.; Huzarski, T.; Byrski, T.; Gronwald, J.; Grzybowska, E.; Mackiewicz, A.; Stawicka, M.; Bebenek, M.; Sorokin, D.; Fiszer-Maliszeqska,L, Haus,O.; "A high proportion of founder BRCA1 mutations in Polish breast cancer families." *International Journal of Cancer*. 110:683-686 (2004).

Gorski, B.; Byrski, T.; Huzarski, T.; Jakubowski,A.; Menkiszak, J.; Gronwald, J.; Pluzanska,A.; Bebenek, M.; Fischer-Maliszewska.; L.; Grzybowska, E.; Narod, SA.; Lubinski, J.; "Founder mutations in the BRCA1 gene in Polish families with breast-ovarian cancer." *American Journal of Human Genetics*. 66:1963-1968 (2000).

Gowen, LC.; Johnson, BL.; Latour, AM.; Sulik, KK.; Koller, BH.; "BRCA1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities." *Nature Genetics*. 12:191-194 (1996).

Gowen, LC.; Avrutskaya, AV.; Latour, AM.; Koller, BH.; Leadon, SA.; "BRCA1 required for transcription-coupled repair of oxidative DNA damage." *Science*. 281:1009-1012 (1998).

Guex, N.; Peitsch, MC.; "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling." *Electrophoresis* 18:2714-2723 (1997).

Hakem, R.; de la Pompa, JL.; Sirard, C.; Mo, R.; Hakem, A.; Wakeham, A.; Potter, J.; Reitmair, A.; Billia, F.; Firpo, E.; Hui, CC.; Roberts, J.; Rossant, J.; Mak, TW.; "The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse." *Cell*. 85:1009-23 (1996).

Hall, JM.; Lee, MK.; Newman, B.; Morrow, JE.; Anderson, LJ.; Huey, B.; King, MC.; "Linkage of early-onset familial breast cancer to chromosome 17q21." *Science*. 250:1684-1689 (1990).

Hartman, AR.; Ford, JM.; "BRCA1 induces DNA damage recognition factors and enhances nucleotide excision repair." *Nature Genetics*. 32:180-184 (2002).

Harvey, SL.; Charlet, A.; Haas, W.; Gygi.; SP.; Kellogg, DR.; "Cdk1-dependent regulation of the mitotic inhibitor Wee1." *Cell*. 122:407-420 (2005).

Hayes, F.; Cayanan, C.; Barilla, D.; Monteiro, AN.; "Functional Assay for BRCA1: Mutagenesis of the COOH-Terminal Region Reveals Critical Residues for Transcription Activation." *Cancer Research*. 60:2411-2418 (2000).

Higgins D.; Thompson, JD.; Gibson, TJ.; "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Research*. 22:4673-4680 (1994).

Holt, JT.; Thompson, ME.; Szabo, C.; Robinson-Benion, C.; Arteaga, CL.; King, MC.; Jensen, RA.; "Growth retardation and turmour inhibition by BRCA1." *Nature Genetics*. 12:298-302 (1996).

Howlett, NG.; Taniquchi, T.; Olson, S.; Cox, B.; Waisfisz, Q.; De Die-Smulders, C.; Persky, N.; Grompe, M.; Joenje, H.; Pals G.; Ikeda, H.; Fox, EA.; D'Andrea, AD.; "Biallelic inactivation of BRCA2 in Fanconi anemia." *Science.* 297:606-609 (2002).

Huberman, J. Repair of Strand Breaks. *Roswell Park Cancer Institute*. Oct 31, 2005. http://saturn.roswellpark.org/cmb/huberman/DNA_Repair/dsbreak.html.

Humphrey, J.; Salim, A.; Erdos, MR.; Collins, FS.; Brody, LC.; Klausner, RD.; "Human BRCA1 inhibits growth in yeast: potential use in diagnostic testing." *Proceedings of the National Academy of Sciences*. 94:5820-5825 (1997).

Imaginis (2006). Genetic Risk Factors for Breast Cancer. http://www.imaginis.com/breasthealth/genetic_risks.asp

Information about Cancer (2006). Stanford Comprehensive Cancer Center. http://cancer.stanford.edu/information/geneticsAndCancer/types/herbocs/

Irminger-Finger, I.; Leung, WC. *et al*. "Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis." *Molecular Cell*. 8:1255-1266 (2001).

Jin, Y.; Xu, XL.; Yang, MC.; Wei, F.; Ayi, TC.; Bowcock, AM.; Baer, R.; "Cell cycle-dependent colocalization of BARD1 and BRCA1 proteins in discrete nuclear domains." *Proceedings of the National Academy of Sciences*. 94:12075-12080 (1997).

Jin, S.; Antinore, ML.; Lung, FD.; *et al.* “The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression.” *Journal of Biological Chemistry*. 275:16602-16608 (2000).

Joo, WS.; Jeffrey, PD.; Cantor, SB.; Finnin, MS.; Livingston, DM.; Pavletich, NP.; “Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure.” *Genes and Development*. 16:583-593 (2002).

Kato, M.; Yano, K.; Matsuo, F.; *et al.* “Identification of Rad51 alteration in patients with bilateral breast cancer.” *Journal of Human Genetics*. 45:133-137 (2000).

Kawai, H.; Li, H.; Chun, P.; Avraham, A.; Avraham, HK.; “Direct interaction between BRCA1 and the estrogen receptor regulates vascular endothelial growth factor (VEGF) transcription and secretion in breast cancer cells.” *Oncogene*. 21:7730-7739 (2002).

King, MC.; Marks, JH.; Mandell, JB.; *et al.* “Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2.” *Science*. 302:574-575 (2003).

Kraulis, J.; Murphy, ME.; “MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures.” *Journal of Applied Crystallography*. 24:946-950 (1991).

Kumar, B.; Lakhotia, S.; Ankathil, R.; *et al.* “Germline BRCA1 mutation analysis in Indian breast/ovarian cancer families.” *Cancer Biology & Therapy*. 1:18-21 (2002).

Le, Trang. BRCA1 and BRCA2 Muatations. 2003. *Myriad Genetics*. http://bioweb.usc.edu/courses/2003-spring/documents/bisc406-t.le.ppt

Lee, EY. BRCA1 and Chk1 in G2/M checkpoint: a new order of regulation. *Cell Cycle*. 1:178-80 (2002).

Lee, JH.; Paull, TT.; “ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex.” *Science*. 308:551-554 (2005).

Li, S.; Ting, NS.; Zheng, L.; Chen, PL.; Ziv, Y.; Shiloh, Y.; Lee, EY.; Lee, WH.; “Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response.” *Nature*. 406:210-215 (2000).

Liede, A.; Jack, E.; Hegele, RA, Narod, SA.; “A BRCA1 mutation in Native North American families.” *Human Mutation*. 19:460 (2002).

Liede, A.; Malik, IA.; Aziz, Z.; Rios, L.; Kwan, E.; Narod, SA.; “Contribution of BRCA1 and BRCA2 mutations to breast and ovarian cancer in Pakistan.” *American Journal of Human Genetics*. 71:595-606 (2002).

Liu, H.; Cartegni, L.; Zhang, MQ.; Krainer, AR.; “A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes.” *Natural Genetics*. 27:55-58 (2001).

Liu, Q.; Guntuku, S.; *et al*. “Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint.” *Genes and Development*. 14:1448-1459 (2000).

Liu, Y.; Virshup, DM; White, RL.; Hsu, LC.; "Regulation of BRCA1 phosphorylation by interaction with protein phosphatase 1alpha." *Cancer Research*. 62:6357-6361 (2002).

Luthy, R.; Bowie, JL.; Eisenberg, D.; "Assessment of protein models with three-dimensional profiles." *Nature*. 356:83-85 (1992).

Machackova, E.; Damborsky, J.; Valik, D.; Foretova, L.; "Novel germline BRCA1 and BRCA2 mutations in breast and breast/ovarian cancer families from the Czech Republic." *Human Mutation*. 18:545 (2001).

Mallery, DL.; Vandenberg, CJ.; Hiom, K.; "Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiqutin chains." *Journal of European Molecular Biology Organization.* 21:6755-6762 (2002).

Matsuda, M.; Miyaqawa, K.; Takahashi, M.; *et al*. "BRCA1 and BRCA2 mutations among breast cancer patients from the Philippines." *International Journal of Cancer*. 98:596-603 (2002).

Mew, N.; Hamel, N.; Galvez, M.; Al-Saffar, M.; Foulkes, WD.; "Haplotype analysis of a BRCA1: 185delAG mutation in a Chilean family supports its Ashkenazi origins." *Clinical Genetics*. 62:151-156 (2002).

Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; *et al.* "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1." *Science*. 7:66-71 (1994).

Naseem, R.; Sturdy, A.; Finch, D.; Jowitt, T.; Webb, M.; "Mapping and conformational characterization of the DNA-binding region of the breast cancer susceptibility protein BRCA1." *Biochemical Journal*. 395:529-535 (2006).

Nathanson, KL.; Wooster, R.; Weber, BL.; "Breast cancer genetics: what we know and what we need." *Nature Medicine*. 7:552-556 (2001).

Nikitin, A.; Eqorov, S.; Daraselia, N.; Mazo, I.; "Pathway studio - the analysis and navigation of molecular networks." *Bioinformatics*. 19:2155-2157 (2003).

O'Brien, KA.; Lemke, SJ,; *et al.* "Casein kinase 2 binds to and phosphorylates BRCA1." *Biochemical and Biophysical Research Communications*. 260:658-664 (1999).

Okada, S.; Ouchi, T.; "Cell cycle differences in DNA damage-induced BRCA1 phosphorylation affect its subcellular localization." *Journal of Biological Chemistry*. 278:2015-2020, (2003).

Olopade, O.; Fackenthal, JD.; Dunston, G.; *et al.* "Breast cancer genetics in African Americans." *Cancer*. 97:236-245 (2003).

Osborne, K..; Schiff, R.; "Growth factor receptor cross-talk with estrogen as a mechanism for tamoxifen resistance in breast cancer." *The Breast*. 12: 362-367 (2004).

Paull, T.; Cortez, D.; Bowers, B.; Elledge, SJ.; Gellert, M.; "Direct DNA binding by BRCA1." *Proceedings. Of National Academy of Science.* 98: 6086-6091 (2001).

Peitsch, M. C. "Protein modeling by E-mail." *Bio/Technology* **13:** 658-660 (1995).

Phelan, CM.; Kwan, E.; *et al.* "A low frequency of non-founder BRCA1 mutations in Ashkenazi Jewish breast-ovarian cancer families." *Human Mutation*. 20:352-357 (2002).

Pichierri, P.; Rosselli, F.; "The DNA crosslink-induced S-phase checkpoint depends on ATR-CHK1 and ATR-NBS1-FANCD2 pathways." *The EMBO Journal*. 23:1178-1187 (2004).

Protein Lounge. 2006. BRCA1 – Breast Cancer Gene Pathway. http://www.proteinlounge.com/.

Puget, N.; *et al.* "Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot." *American Journal of Human Genetics*. 70:858-865 (2002).

Qin, D. *et al*. "Identification of potential binding sites for the FHA domain of human Chk2 by in vitro binding studies." *Biochemical and Biophysical Research Communications*. 311:803-808 (2003).

Rostagno, P.; Gioanni, J.; *et al.* "A mutation analysis of the BRCA1 gene in 140 families from southeast France with a history of breast and/or ovarian cancer." *Journal of Human Genetics*. 48:362-366 (2003).

Ruffner, H.; Jiang, W.; Craig, AG.; Hunter, T.; Verma, IM.; "BRCA1 is phosphorylated at serine 1497 in vivo at a cyclin-dependent kinase 2 phosphorylation site." *Molecular Cell Biology*. 19:4843-4854 (1999).

Santarosa, M.; *et al*. "BRCA1 and BRCA2 genes: role in hereditary breast and ovarian cancer in Italy." *International Journal of Cancer*. 83:5-9 (1999).

Schafer, J. *et al.* "A mechanism of drug resistance to tamoxifen in breast cancer." *Journal of Steroid and Molecular Biology*. 83: 75-83 (2002).

Scheikh, MS.; Hhollander, MC.; Fornance, AJ.; "Role of GADD45 in apopotosis." *Biochemical Pharmacology*. 59:43-45 (2000).

Scottish/Northern Irish BRCA1/BRCA2 Consortium. "BRCA1 and BRCA2 mutations in Scotland and Northern Ireland." *British Journal of Cancer*. 88:1256-1262 (2003).

Scully, R. *et al.* "Association of BRCA1 with Rad51 in mitotic and meiotic cells." *Cell Press*. 88:265-275 (1997).

Scully, R.; *et al.* "Genetic analysis of BRCA1 function in a defined tumor cell line." *Molecular Cell* 4:1093-1099 (1999).

Schwede, T.; Kopp, J.; Guex, N.; Peitsch, MC.; "SWISS-MODEL: an automated protein homology-modeling server." *Nucleic Acids Research* **31:** 3381-3385 (2003).

Shiozaki, EN.; Gu, L.; Yan, N.; Shi, Y.; "Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling." *Molecular Cell*. 14:405-412 (2004).

Simard, J. *et al.* "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families." *Nature Genetics*. 8:392-398 (1994).

Smith, ML.; *et al.* "Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen." *Science*. 266:1376-1380 (1994).

Sordet, O.; Khan, QA.; Kohn, KW.; Pommier, Y.; "Apoptosis Induced by Topoisomerase Inhibitors." *Current Medicinal Chemistry*. 3:271-290 (2003).

Struewing, JP. *et al.* "The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews." *New England Journal of Medicine*. 336:1401-1408 (1997).

Sum, EY.; Peng B.; Yu, X.; Chen, J.; Byrne.; Linderman, GJ.; Visvader, JE.; "The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity." *Journal of Biological Chemistry*. 7849-7856 (2002).

Szabo, CI.; *et al.* "Human, canine and murine BRCA1 genes: sequence comparison among species." *Human Molecular Genetics*. 5:1289-1298 (1996).

Tang, N. *et al.* "Prevalence of mutations in the BRCA1 gene among Chinese patients with breast cancer." *Journal of the National Cancer Institute*. 91:882-885 (1999).

Taniguchi, T.; *et al.* "Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways." *Cell*. 109:459-472 (2002).

Thompson, ME.; Jensen, RA.; Obermiller, PS.; Page, DL.; Hold, JT.; "Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression." *Nature Genetics*. 9:444-50 (1995).

Tonin, PN.; *et al.* "Founder BRCA1 and BRCA2 mutations in French Canadian breast and ovarian cancer families." *American Journal of Human Genetics*. 63:1341-1351 (1998).

Unleashed Informatics. 2006. Biomolecular Object Network Databank. http://bind.ca/

Vallon-Christersson, J.; *et al.* "Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families." *Human Molecular Genetics*. 10:353-360 (2001).

Varma, AK.; Brown, RS.; Birrane, G.; Ladias, JA.; "Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex." *Biochemistry*. 44:10941-10946 (2005).

Verhoog, L. *et al.* "Large regional differences in the frequency of distinct BRCA1/BRCA2 mutations in 517 Dutch breast and/or ovarian cancer families." *European Journal of Cancer*. 37:2082-2090 (2001).

Wada, O.; *et al.* "BRCA1 function mediates a TRAP/DRIP complex through direct interaction with TRAP220." *Oncogene.* 23:6000-6005 (2004).

Waldman, T.; *et al.* "p21 is necessary for the p53-mediated G1 arrest in human cancer cells." *Cancer Research*. 55:5187-90 (1995).

Wang, H.; *et al.* "BRCA1 proteins are transported to the nucleus in the absence of serum and splice variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases." *Oncogene*. 15:143-57 (1997).

Wang, Y.; *et al.* "BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures." *Genes and Development*. 14:927-939 (2000).

Wheeler, D.; *et al.* "Database resources of the National Center for Biotechnology Information: 2002 update." *National Center for Biotechnology Information*. 30:13-16 (2002).

Whitehouse, C.; Chambers, J.; Catteau, A.; Solomon, E.; "Brca1 expression is regulated by a bidirectional promoter that is shared by the Nbr1 gene in mouse." *Gene*. 326:87-96 (2004).

Wilda, M. *et al.* "Expression pattern of the Nijmegen breakage syndrome gene, Nbs1, during murine development." *Human Molecular Genetics*. 9:1739-1744 (2000).

Williams, RS.; Green, R.; Glover, JN.; "Cystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1." *Nature Structural Biology*. 8:838-842 (2001).

Williams, RS.; *et al.* "Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations." *Journal of Biological Chemistry*. 278:53007-53016 (2003).

Williams, RS.; Glover, JN.; "Structural consequences of a cancer-causing BRCA1-BRCT missense mutation." *Journal of Biological Chemistry*. 278:2630-2635 (2003).

Wong, AK.; *et al.* "Characterization of a carboxy-terminal BRCA1 interacting protein." *Oncogene*. 17:2279-2285 (1998).

Wooster, R.; *et al.* "Identification of the breast cancer susceptibility gene BRCA2." *Nature*. 378: 789-792 (1995).

Worley, T.; Vallon-Christersson, J.; Billack, B.; Borg, A.; Monterio, AN.; "A naturally occurring allele of BRCA1 coding for a temperature-sensitive mutant protein." *Cancer Biology & Therapy*. 1:497-501 (2002).

Wu, LC. *et al.* "Identification of a RING protein that can interact in vivo with the BRCA1 gene product." *Nature Genetics*. 14:430-440 (1996).

Xu, B.; Kim, ST.; Kastan, MB.; "Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation." *Molecular Cell Biology*. 21:3445-3450 (2001).

Xu, X. *et al.* "Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells." *Molecular Cell*. 3:389-395 (1999).

Xu, X.; *et al.* "Genetic interactions between tumor suppressors Brca1 and p53 in apoptosis, cell cycle and tumorigenesis." *Nature Genetics*. 28:266-271 (2001).

Yang, Q.; *et al.* "Identification of a functional domain in a GADD45-mediated G2/M checkpoint." *Journal of Biological Chemistry*. 275:36892-36898 (2000).

Yarden, R.; Brody, LC.; "BRCA1 interacts with components of the histone deacetylase complex." *Proceedings. Of National Academy of Science*. 96:4983-4988 (1999).

Yarden, R.; *et al.* "BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage." *Nature Genetics*. 30:285-289 (2002).

Yu, X.; Chen, J.; "DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains." *Molecular Cell Biology*. 24:9478-9486 (2004).

Yu, X.; Chini, CC.; He, M.; Mer, G.; Chen, J.; "The BRCT domain is a phospho-protein binding domain." *Science*. 302:639-642 (2003).

Zhang, H.; *et al.* "BRCA1 physically associates with p53 and stimulates its transcriptional activity." *Oncogene*. 16:1713-1721 (1998).

Zhang, X.; *et al.* "Structure of an XRCC1 BRCT domain: a new protein-protein interaction module." *Journal of European Molecular Biology Organization*. 6406-6411 (1998).

Zhong, Q.; *et al.* "Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response." *Science*. 285:747:750 (1999).

Zikan, M.; Pohlreich, P.; Stribrna, J.; "Mutational analysis of the BRCA1 gene in 30 Czech ovarian cancer patients." *Journal of Genetics*. 84:63-67 (2005).

Zou, L.; Elledge, SJ.; "Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes." *Science*. 300:1542-1548 (2003).

Appendix A

Molscript input file for 1JNX terminus:

```
! MolScript v2.1 input file

title "BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN"

plot
  window 38.53;
  slab 27.35;

  read mol "1JNX-1.pdb";

  transform atom * by centre position atom *
   by rotation
  -0.620329 0.151884 -0.769495
  -0.635512 -0.672322 0.379615
  -0.459691 0.72451 0.513585;

  set segments 2;

  set planecolour hsb 0.6667 1 1;
  coil from X1649 to X1651;
!B1
  set planecolour hsb 0.6154 1 1;
  strand from X1651 to X1655;
  set planecolour hsb 0.5641 1 1;
  coil from X1655 to X1658;
!A1
  set planecolour hsb 0.5128 1 1;
  helix from X1658 to X1673;
  set planecolour hsb 0.4615 1 1;
  coil from X1673 to X1674;
!B2
  set planecolour hsb 0.4103 1 1;
  strand from X1674 to X1677;
  set planecolour hsb 0.359 1 1;
  coil from X1677 to X1685;
!B3
  set planecolour hsb 0.3077 1 1;
  strand from X1685 to X1690;
  set planecolour hsb 0.2564 1 1;
  coil from X1690 to X1695;
  set planecolour hsb 0.2264 1 1;
  coil from X1695 to X1700;
!A2
  set planecolour hsb 0.2051 1 1;
  helix from X1700 to X1709;
  set planecolour hsb 0.1551 1 1;
  coil from X1709 to X1712;
```

```
!B4
  set planecolour hsb 0.1538 1 1;
  strand from X1712 to X1715;
  set planecolour hsb 0.1026 1 1;
  coil from X1715 to X1716;
!A3
  set planecolour hsb 0.0752 1 1;
  helix from X1716 to X1726;
  set planecolour hsb 0.05128 1 1;
  coil from X1726 to X1735;
  set planecolour hsb 0 1 1;
  coil from X1735 to X1740;

end_plot
```

Molscript input file for 1JNX terminus:

```
! MolScript v2.1 input file

title "BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN"

plot
  shadows on;
  window 61.38;
  slab 32.30;

  read mol "1JM7-1.pdb";
  transform atom *
    by centre position atom *;
  set segments 2;

  set planecolour hsb 0.6667 1 1;
  coil from A1 to A7;
  set planecolour hsb 0.6275 1 1;
  helix from A7 to A22;
  set planecolour hsb 0.5882 1 1;
  coil from A22 to A33;
  set planecolour hsb 0.549 1 1;
  strand from A33 to A37;
  set planecolour hsb 0.5098 1 1;
  coil from A37 to A42;
  set planecolour hsb 0.4706 1 1;
  strand from A42 to A44;
  turn from A44 to A45;
  set planecolour hsb 0.4314 1 1;
  helix from A45 to A54;
  set planecolour hsb 0.3922 1 1;
  coil from A54 to A72;
  set planecolour hsb 0.3529 1 1;
  strand from A72 to A75;
  set planecolour hsb 0.3137 1 1;
```

```
  coil from A75 to A80;
  set planecolour hsb 0.2745 1 1;
  helix from A80 to A97;
  set planecolour hsb 0.2353 1 1;
  coil from A97 to A103;

  set planecolour goldenrod;
  coil from B1 to B34;
!  set planecolour hsb 0.1961 1 1;
  helix from B34 to B47;
!  set planecolour hsb 0.1569 1 1;
  coil from B47 to B74;
!  set planecolour hsb 0.1176 1 1;
  helix from B74 to B79;
!  set planecolour hsb 0.07843 1 1;
  coil from B79 to B98;
!  set planecolour hsb 0.03922 1 1;
  helix from B98 to B117;
!  set planecolour hsb 9.714e-17 1 1;
  coil from B117 to B122;

  set colourparts on;
  cpk in residue A123;
  cpk in residue A124;
  cpk in residue B125;
  cpk in residue B126;

  set colourparts off;
  set planecolour white;
  ball-and-stick require in residue A7 and backbone;
  ball-and-stick in residue A10;
  ball-and-stick in residue A85;
  ball-and-stick in residue A96;

  set colourparts off;
  set planecolour darkolivegreen;
  ball-and-stick require not peptide, not hydrogens and in
residue B36;
  ball-and-stick require not peptide, not hydrogens and in
residue B43;
  ball-and-stick in residue B117;
  ball-and-stick in residue B110;

  set linecolour cyan, linedash 4;
  line position res-atom A96 CB to position res-atom B36
NE2;
  line position res-atom A85 OE2 to position res-atom B43
NH1;
  line position res-atom A85 OE2 to position res-atom B43
NH2;
  line position res-atom A85 OE2 to position res-atom B43
NE;
```

```
  line position res-atom A7 CA to position res-atom B117
OD1;
  line position res-atom A10 OE2 to position res-atom B110
O;
  line position res-atom A10 OE2 to position res-atom B110
CA;
  line position res-atom A10 1HG to position res-atom B110
CB;

  set linecolour white;
  set labeloffset 2 0 0; label res-atom A123 ZN "%t";
  set labeloffset -5 0 0; label res-atom A124 ZN "%t";

end_plot
```