Aus dem Institut für
Pharmakologie und Toxikologie
der
Philipps-Universität Marburg

geschäftsführender Direktor: Prof. Dr. T. Gudermann

HERV-E. BRCA1

A Human Endogenous Retrovirus located in the Human BRCA1 Gene Locus

Inaugural-Dissertation zur Erlangung des Doktorgrades der gesamten Medizin dem Fachbereich Medizin der Philipps-Universität Marburg

vorgelegt von
Lars Hofmann
aus Landau

Marburg 2002

Angenommen vom Fachbereich Humanmedizin der Philipps-Universität Marburg am 12.12.2002 gedruckt mit Genehmigung des Fachbereichs

Dekan: Prof. Dr. med. B. Maisch
Referent: Prof. Dr. med. Frank Czubayko
Correferent: Prof. Dr. Radsak

Contents

1 Abstract

Breast cancer is the most frequent malignant disease in females worldwide. Of all cases 90 – 95% seem to occur sporadic, while 5 – 10% seem to have a hereditary origin. Sporadic breast cancer has an onset in the elderly patient, while about 30% of the hereditary cases have an onset at an age under 35. The majority of hereditary breast cancer cases are based on mutations in one of the two breast cancer susceptibility genes – BRCA1 and BRCA2. In BRCA1 alone, over 60 different mutations have been found of which most lead to truncated proteins or loss of transcript. Since BRCA1 was proven to be a tumor suppressor gene, the result can be fatal. The influence of BRCA1 in sporadic breast cancer is not clearly defined yet, but it is known that BRCA1 expression levels are downregulated in certain sporadic breast cancer cell lines. This leads to the conclusion that the regulation of BRCA1 gene expression has an impact on the development of breast cancer.

In 1998, Schulte et al. detected an 891 nt long, retroviral-like sequence in a duplicated region of the first three exons and introns of the BRCA1 gene (pseudogene), 35 kb upstream of the BRCA1 gene. Sequence comparisons of this region with GenBank entries revealed an up to 90% homology to a Human endogenous retroviral element (HERV) inserted in the human pleiotrophin gene (PTN). This retroviral element (HERV-E.PTN) generates an additional promoter with trophoblast specific activity [Schulte et al., 1998].

Based on this finding I defined the structure of the retrovirus-like element, inserted in the human BRCA1 pseudogene. Sequencing by primer walking exposed a complete retroviral-like element of 7455 bp length including both long terminal repeats. The high homology to HERV-E.PTN and the family specific $tRNA^{glu}$ binding site defined it as a member of the HERV-E family. I referred to it as HERV-E.BRCA1. Further sequence analysis showed us high similarities to the three defective retroviral coding genes *gag*, *pol* and *env*. Additionally I detected sequence homologies to retroviral elements from the RTVL-I family (retrovirus-like sequence-isoleucine), named after its primer binding site complementary to an isoleucine tRNA. This strongly suggests a retroviral recombination in the past.

A phylogenetic analysis with DNA samples of human, chimpanzee, gorilla and Rhesus monkey, defines the time point of insertion into the genome to at least 25 mio years ago.

An amplification of the HERV-E.BRCA1 5'LTR and transfection into one sporadic breast cancer cell line and three choriocarcinoma cell lines with high BRCA1 expression, revealed orientation independent, transcriptional activity of this element up to 45 fold over empty vector. In line with this result we speculated, whether the 5'LTR of the HERV-E.BRCA1 element could theoretically be able to initiate the expression of BRCA1-pseudogene antisense and / or of retroviral transcripts.

2 Introduction

2.1 Breast Cancer

2.1.1 Incidence and Mortality

The carcinoma of the female mammary gland is the most frequent malignancy for women in Germany and has one of the highest rates of incidence worldwide.
The National Cancer Institute estimates that one out of eight women in the US will develop breast cancer in her lifetime. About 45.800 new cases of breast cancer were diagnosed in Germany in the year 1997 (Robert-Koch Institut). Numbers from the Krebsregister Saarland / Germany for 1995 were as follows:

Incidence of breast cancer cases

age group	Incidence out of 100.000 women
30 – 35 years	**22**
40 – 45 years	**138**
60 – 65 years	**232**
75 – 80 years	**288**

Fortunately the mortality rate has been decreasing during the last decade. The age-adjusted mortality rate for white females of all ages in the US (standardized to the population of 1970), showed a 6.8% decline between 1989 to 1993. Nevertheless, the mortality rate for US white females in the age group 40-49 is 26,5 out of 100.000, for ages 50-59 it is 56 out of 100.000 and for ages 70-79 it is 126 out of 100.000 women (standardized to the 1970 US population) [Chu et al., 1996].
Although incidence and especially mortality rates have been dropping since the late 1980s [Smigel et al., 1995] due to better screening programs (mammography) and better treatments, breast cancer is still one of the highest medical and financial impact factors on medical systems in western nations. Therefore it is mandatory to

understand the mechanisms which cause breast cancer as well as the pathways that might lead to new treatments and earlier detection.

2.1.2 Pathogenesis

A series of risk factors are involved in the pathogenesis of breast cancer [as reviewed in Martin et al., 2000] such as:

1. **Hormonal risk factors**, e.g. increased estrogen exposure over a prolonged time [Begg et al., 1987]. Especially estrogen replacement therapy in menopause increases the incidence of breast cancer. After 15 years of estrogen replacement therapy a 30% increase in incidence rates could be proven, although these studies included pre-menopausal women and females which were using estradiol [Steinberg et al., 1991]. Studies researching the effect of oral contraceptives (OCs) show inconsistent results but it seems that the risk for breast cancer is slightly elevated. Especially when commencing the usage of OCs at a young age (<20 years) it presumably has a negative effect on the risk of developing breast cancer. 10 years after stopping the use of OCs the risk returns to normal levels [Cancer Net, National Cancer Institute, 2002]. Prospective controlled studies are needed to further elucidate the influence of hormone replacement therapy on the development of breast cancer [Marsden, 2002].

2. **Nonhormonal risk factors**, such as increased alcohol consumption. It is important that as many variables as possible are included in these studies, such as the age of alcohol consumption onset, its duration and intensity (grams per day). The age onset of drinking behavior seems not to influence the cancer risk, in opposition to the duration, although not when simultaneously included in the statistic with the amount per day. This result suggests that the important factor might be the intensity of drinking [Bowlin et al., 1997].

   Obesity in postmenopausal women is another risk factor, which is controversially discussed. The mechanism seems to be the metabolic conversion of androstenedione into estrone by adipose tissue [Pujol et al., 1997]. However a common problem in assessing nonhormal

risk factors for breast cancer development is the large number of variables like age, country of origin, nutrition, social background etc. A further risk factor seems to be exposure to radiation. Women who were exposed, e.g. to a nuclear weapons detonation in Japan or radiation treatment of Hodgkin's Disease seem to have an elevated risk for breast cancer [Goss et al., 1998].

3. **Genetic factors** – such as loss of function in the Breast Cancer susceptibility genes BRCA1 and BRCA2. Germline mutations in these genes could be correlated to the onset of hereditary breast cancer and are discussed as being involved in the development of sporadic breast cancer as well (as reviewed in [Yang et al., 1999] and [Krainer et al., 1997]). Furthermore it was reported that an additional gene on chromosome 13q (BRCAX) may also be associated with the development of breast cancer [Hopper et al., 2001]. Additionally there are at least five other genes linked to breast cancer (TP53, PTEN, ATM, HRAS1 and the androgen receptor gene p53), although with less numerical relevance than BRCA1 and BRCA2 [Easton et al., 1999].
4. **Other risk factors** are numerous such as late childbearing, breast tissue density, tobacco smoke etc. [Cancer Net, National Cancer Institute].

Breast cancer cases are classified based on their pathology, size and metastatic phenotype (TNM classification):

Stage 0: In situ carcinoma of the breast (DCIS and LCIS)

Stage I: Tumor size <cm2 and has not spread

Stage II:

IIA: The tumor is between 2 and 5 cm but there are no positive lymph nodes (LN) **or** the tumor is < 2 cm and has spread into axillary LNs.

IIB: The tumor is between 2 and 5 cm and has spread into axillary LNs **or** the tumor is up to 5 cm but has not spread.

Stage III:

IIIA: The tumor is smaller than 5 cm and has spread to the lymph nodes under the arm, and the lymph nodes are attached to each

other or to other structures **or** the tumor is larger than 5 cm and has spread to axillary lymph nodes.

IIIB: The tumor has spread to tissues near the breast (skin or chest wall, including the ribs and the muscles in the chest) **or** the tumor has spread to lymph nodes inside the chest wall along the breast bone.

Stage IV:

The cancer has spread to other organs of the body, most often to the bones, lungs, liver, or brain, or the tumor has spread locally to the skin and lymph nodes inside the neck, near the collarbone.

Survival rates depend on the stage of the disease. The five year survival rate for stage I is 90%, 65% for stage II, 45% for stage III and drops to 10% in stage IV of the disease. The number of involved lymph nodes and the tumor size are the best prognostic factors for the survival rate. Nevertheless, newer prognostic factors like hormone receptor status, number of mitotic figures, DNA index (grade of ploidy), number of blood vessels or the status of the genes p53, ERBB-2, EGF-R, TGFα are also indicators for the predicted survival rate.

Breast cancer can be distinguished in the hereditary or the sporadic form. Although the majority of breast cancer cases are thought to be sporadic, about 5% of the cases are estimated to be hereditary.

Hereditary breast cancer differs in a number of features from sporadic breast cancer:

1. early onset of the disease: about 30% of the women with a genetic pre-disposition develop breast cancer before the age of 35 [Claus et al., 1990]
2. increased prevalence for the development of a bilateral disease [Blackwood et al., 1998]
3. higher incidence of male breast cancer
4. association with other malignancies, e.g. ovarian, colon and prostate cancer [Marcus et al., 1994]
5. higher grades of detected BRCA1 associated cancer, which may be due to a faster growth rate [Breast Cancer Linkage Consortium, 1997]

About 5-10% of all breast cancer cases seem to be of a hereditary origin and the largest amount can be explained by mutations in one of the two susceptibility genes named BRCA1 and BRCA2. Germline mutations of the BRCA1 gene account for 40 - 50% of hereditary breast cancer [Claus et al., 1996].

2.1.3 Breast Cancer Therapy

Treatment of breast cancer depends on the type and stage of the disease, the patient's age, menopausal status, and overall health. Concerning surgery, the treatment of choice up to stage IIIB is either a breast conserving surgery, meaning the sole extirpation of the tumor or modified radical mastectomy. Modified radical mastectomy means the removal of the breast, of many of the lymph nodes (lymphadenectomy) under the arm, the lining over the chest muscles, and sometimes part of the chest wall muscles. Surgery is often followed by tangential X-ray radiation therapy (60 Gy) and / or systemic treatment with chemotherapeutic substances like cyclophosphamide (nitrogen mustard), methotrexate (folic acid analog) and 5-fluorouracile (pyrimidine analog) or hormones like Tamoxifen (antiestrogen), Raloxifene or estrogen.

Hereditary factors, such as mutations recently enter the clinical field. One study actually showed the possible benefit of a prophylactic salpingo-oophorectomy in women with mutations of BRCA1 and BRCA2, by reducing the risk for developing related gynecologic cancer [Kauff, et al. 2002].

These therapies underlie constant changes and vary from hospital to hospital. Additionally new therapy schemes are being introduced constantly.

(http://cancernet.nci.nih.gov/cancer_types/breast_cancer.shtml)

2.2 BRCA1 – A Breast Cancer Susceptibility Gene

2.2.1 The Human BRCA1 Gene

The hypothesis that a highly penetrant autosomal susceptibility allele could be responsible for the aggregation of inherited breast cancer, was first shown by Hall et al. In 1990 his research group was able to identify a region on chromosome 17q21 which seemed to be the location for inherited early-onset breast cancer

[Hall et al., 1990]. The linkage of a specific gene locus to a particular marker, provided the evidence that breast cancer predisposition is inherited in a Mendellian fashion in certain families [Vogelstein et al., 1994]. This region was mapped and the breast and ovarian susceptibility gene BRCA1 was isolated by positional cloning in 1994 by Miki et al. [Miki et al., 1994].

The human BRCA1 gene spreads over 100 kb and is composed of 24 exons. The open reading frame encodes for a nuclear protein of 1,863 amino acids [Miki et al., 1994]. Due to genomic rearrangements, the human BRCA1 locus contains a ~35 kb duplicated area which results in two copies of the first three exons of the human BRCA1 (pseudo BRCA1 exon 1a, 1b and 2) and of the upstream

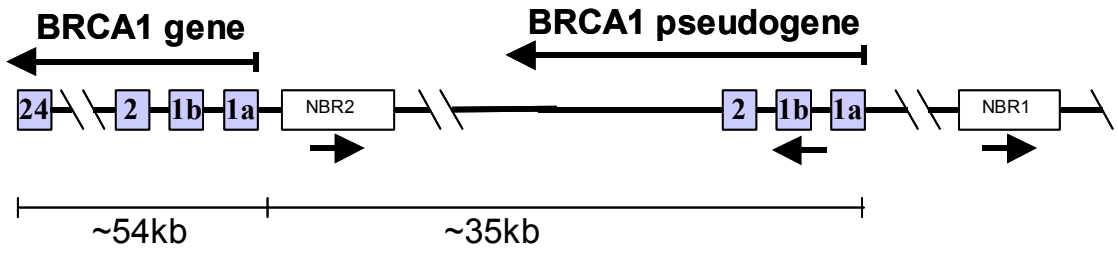

Figure 1:
The BRCA1 locus on chromosome 17q21. A duplication of at least the first three exons of the BRCA1 gene as well as the nearby NBR1 gene, has taken place and this fragment has reinserted in the locus ~35 kb upstream of the 5' end of the BRCA1 gene. The copy of the NBR1 gene is actually partially a new gene, thus named NBR2. Downstream of exon 2 in the BRCA1 pseudogene, we detected an insertion of retroviral DNA via computer sequence comparison with GenBank entries. Elements in this figure do not represent actual scale.

located NBR1 gene (pseudo NBR1 also named NBR2). Thus, a pseudo BRCA1 gene is located ~30 kb upstream of the human BRCA1 gene (s. figure 1). The murine BRCA1 locus does not bear this duplication, while it seems to be existing already in the BRCA1 locus of Rhesus monkey as our data indicates.

Expression of the BRCA1 gene is under the regulation of two promoters of which one (α) is acting bi-directional. The α-promoter is 267 bp long and covers the first 43 nt of exon 1a of the BRCA1 gene, 218 bp intergenic region and 6 nt of the first exon of the adjacent NBR2 gene (**n**ext to the **BR**CA1 gene). The α-promoter regulates the expression of the BRCA1 gene as well as the expression of the pseudo NBR1 gene. The β-promoter is located between exons 1a and 1b of the BRCA1 gene [Xu et al., 1997]. The murine BRCA1 gene is only under the regulation of one promoter, indicating possible species differences in the gene regulation and

potentially in the biology of the gene products.

The BRCA1 protein is a 1.863 aa (220-kDA) nuclear phosphoprotein with a highly conserved NH_2 terminal [Chen et al., 1995], a RING finger domain (zinc finger) and a COOH terminal including two copies of the BRCT domain [Koonin et al., 1996]. Mutations in these regions predispose for breast and ovarian cancer [Brzovic et al., 2001]. The BRCA1 protein participates in the regulation of cell proliferation, cell cycle progression, apoptosis induction and DNA repair and recombination, e.g. RAD51 and p53. Cells with nonfunctional BRCA1 protein do not arrest in G2 after DNA damage [Larson et al., 1997] and are supposedly deficient in transcription-coupled DNA repair [Gowen et al., 1998].

2.2.2 BRCA1 in Hereditary and Sporadic Breast Cancer

The BRCA1 Open Reading Frame encodes for a protein of 1883 amino acids. In 1996, already 63 mutations were found in the BRCA1 gene of hereditary breast cancer samples, which were distributed across the whole coding region of the gene. 84 % of the 63 mutations were predicted to cause truncated proteins or a loss of BRCA1 transcript [Xu et al., 1996]. Studies have proven that BRCA1 acts as a tumor suppressor gene (as reviewed in [Yang et al., 1999]):

1. Transfection of intact chromosome 17 bearing the BRCA1 gene into breast cancer cell lines lead to growth arrest [Negrini et al., 1994] and [Casey et al., 1993]
2. Tumors with mutations in the BRCA1 gene grow more rapidly [Marcus et al., 1994]
3. Lower BRCA1 expression levels in invasive breast cancer versus in CIS (carcinoma in situ) [Thompson et al., 1995]
4. Induction of BRCA1 expression in breast cancer cells triggers apoptosis through the induction of GADD45 by activation of the JNK/SAPK pathway [Harkin et al., 1999]. Additionally, BRCA1 induced apoptosis seems to involve the inhibition of the ERK 1 / 2 pathway as well [Yan et al., 2001]
5. BRCA1 seems to transactivate the expression of p27 (Kip1) and therefore induce growth inhibition [Williamson et al., 2002]
6. Certain mutations in the BRCA1 gene allow sustained activation of

telomerase (key enzyme in carciogenesis), via a pathway including a transcription factor complex containing BRCA1, c-Myc and NMI [Li et al. 2002]

Having established the link between BRCA1 and the development of hereditary breast cancer, it was questioned whether BRCA1 could also play a role in the development of sporadic breast cancer. Since sporadic breast cancer makes up to 95% of all cases it was of high importance to define the expression level of this tumor suppressor gene and its encoded protein. Indeed, several research groups could determine, that BRCA1 mRNA levels were decreased in sporadic breast cancer samples [Magdinier et al., 1998]. Furthermore inhibition of BRCA1 by antisense strategy accelerated the growth of malignant and normal mammary cells, [Thompson et al., 1995] indicating that loss of BRCA1 could play also play an important role in the development of sporadic breast cancer. Possible mechanisms of silencing the BRCA1 promoter include hypermethylation, especially in the presence of loss of heterozygosity (LOH) in particular histological subtypes [Esteller et al., 2000]. Additional mechanisms could be interruption of the BRCA1 pathway or mutations of the human BRCA1 gene.

2.3 Retroviruses

Retroviruses are RNA viruses. The virion RNA is 7-12 kb long, non-segmented and single stranded. They can be divided in simple and complex retroviruses. This is depending on the amount of coding regions they possess. All retroviruses contain at least three coding regions:

1. ***gag***, which directs the synthesis of internal virion proteins for the matrix, the capsid and the nucleoprotein structure
2. ***pol***, coding for reverse transcriptase and integrase enzymes
3. ***env***, carrying the information for surface and transmembrane components

Additionally all retroviruses have a domain which encodes for their protease (***pro***). Another way of subdividing retroviruses is by their relatedness, which creates seven subgroups. Five out of these seven groups have oncogenic potential. The replication of retroviruses is depending on the reverse transcription of its RNA genome in DNA sequence. The detection of retroviruses changed the scientific view on genetic information dramatically, due to the fact that the flow of information was considered at that time to be unidirectional from DNA to RNA to protein. Today we understand that a large percentage of the human genome is actually the product of reverse transcription, containing sequences whose characteristics point to RNA as a template precursor.

Retroviruses fuse with host cells after attaching to their membrane receptors which are detected by specific glycoproteins on the viral surface. After entering into the host cell, the virus RNA genome is still contained in a core complex of nonglycosilated proteins and attached to the virion reverse transcriptase. The reverse transcription takes place in the host cells cytoplasm. During this process, the virus coded reverse transcriptase has to perform two jumps from the 5' terminus to the 3' terminus of the template. This creates a duplication of the 5' and 3' ends of the virion

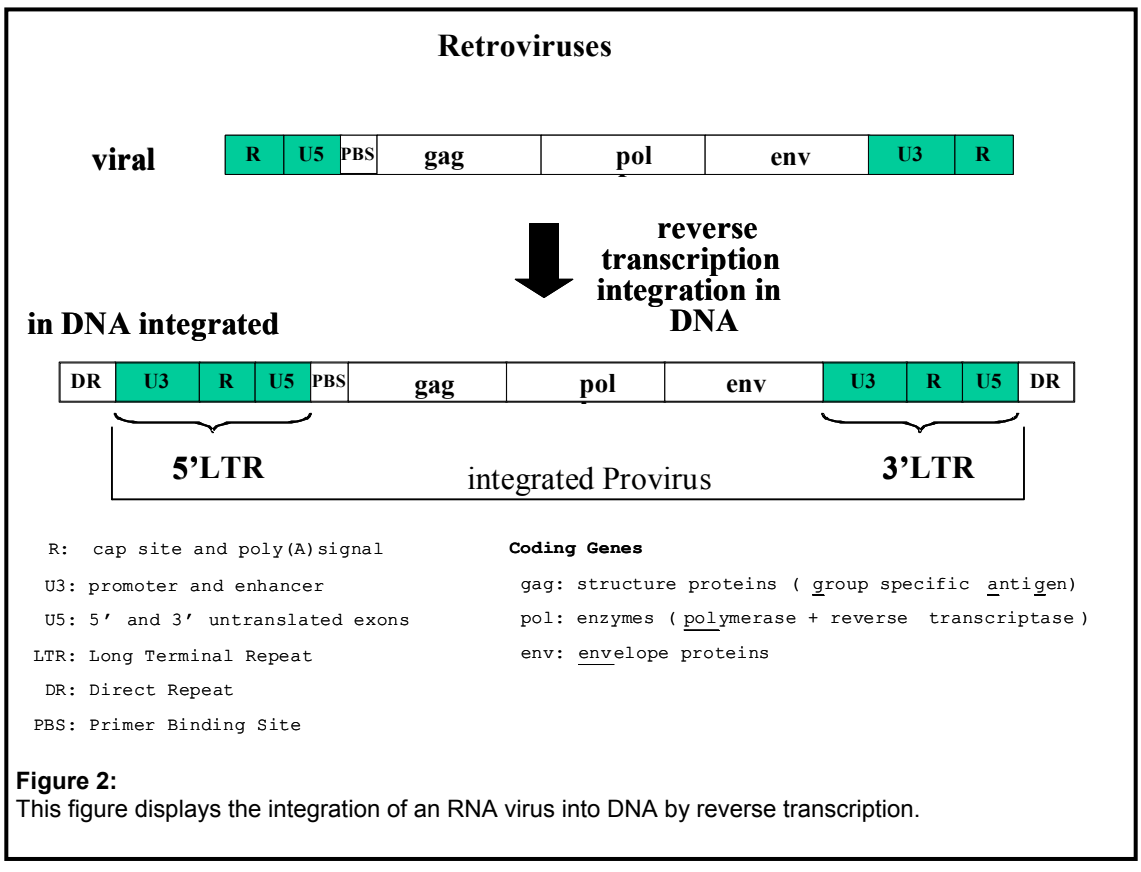

Figure 2:
This figure displays the integration of an RNA virus into DNA by reverse transcription.

RNA. At both ends of the viral DNA the duplicated regions are named Long Terminal Repeats (LTR). The LTRs regulate viral gene expression and viral replication (s. figure 2). The viral DNA integrates itself into the host DNA through its virion integrase. The provirus is now stable integrated and therefore expressed through the aid of cellular RNA polymerase II. Simple retroviruses control their transcription solely by interaction of cellular factors with their proviral LTRs. More complex viruses encode their own trans-activating factors to regulate their level of expression. Translation products combined with progeny RNA are assembled into viral particles. Then they fuse with the host cell membrane and are released.

Since retroviral particles are diploid, heterozygote virions form in cells which are infected by related retroviruses. Recombination between related retroviruses is a frequent occurrence [Temin, 1991], leading to insertion of different HERV fragments into the same gene locus.

2.3.1 Human Endogenous Retroviruses (HERVs)

Human endogenous retroviruses are remnants of prior infections with exogenous retroviruses and changes have occurred to make them no longer infectious [Levy et al., 1996]. They are scattered throughout the genome and make up to 1% of human DNA. HERVs have been identified over the past years, by a) screening the human genome with probes derived from animal retroviruses, b) screening with oligos derived from the viral primer binding site, or c) simply by chance. Two models of classifications have been developed, either by their homology to infectious animal retroviruses or by the identity of their tRNA which has homology to the minus strand primer binding site [Levy et al., 1996]. The first classification bases on the homology of the *pol* region in HERVs and the pol region in mammalian retroviruses:

Class I: HERVs with homology to mammalian type C retroviruses like the murine leukemia virus (MLV)

Class II: HERVs with homology to mammalian type B and D and avian type C retrovirus strains

The second classification compares the primer binding sites, a sequence complimentary to a particular tRNA which is located immediately downstream of the

5'LTR. The one letter code for the specific amino acid is added as a suffix to the acronym HERV [Lower et al., 1996]. This leads to the classification HERV-E for Glu-tRNA, HERV-I for Ile-tRNA, HERV-K for Lys-tRNA etc.

Most HERV fragments are defective in the genome and contain multiple termination signals and mutations throughout the sequence corresponding to the *gag*, *pol* and *env* genes, since selective pressure no longer applies to them [Levy et al., 1996]. Through cycles of evolution the majority of ORFs acquired random mutations and deletions. However, there are a few exceptions, e.g. HTDV/HERV-K and ERV-3, which express viral proteins [Lower et al., 1993] and [Venables et al., 1995]. Latter HERVs prove the point, that HERVs might posses a biological function by themselves, most likely when creating a selective advantage for the host.

HERVs can effect host genes by acting as insertional mutagens [Patience et al., 1997]. They are able to alter the expression patterns of adjacent genes through promoter interference. One example is the insertion of a retroviral element into the human amylase gene cluster. During evolution, insertion of a member of the HERV-E family into the amylase locus seems to be responsible for the expression of amylase in the human salivary glands [Samuelson et al., 1990] and [Ting et al., 1992].

Recently it was reported by Schulte et al., that the insertion of a HERV-E class virus (HERV-E.PTN) [Schulte et al., 1998] into the human growth factor pleiotrophin gene (PTN), generated an additional PTN promoter with trophoblast-specific activity. Activation from this promoter leads to the expression of HERV-PTN fusion transcripts in human trophoblast derived normal and tumor cells. Functional study has revealed, that depletion of HERV-PTN fusion transcripts in human trophoblast-derived choriocarcinoma cells almost prevented HERV tumorigenicity in an *in vivo* mouse model. Taken together we concluded, that insertion of the retroviral element into the human growth factor pleiotrophin gene, transferred intronic sequence into promoter active sequence and generated a phylogenetically novel promoter for this gene.

Based on the fact that retroviral elements have the ability to regulate adjacent host genes, thorough examinations of Human Endogenous Retroviruses is of great importance.

2.4 Goals of this Study

Preliminary data

- Expression of the BRCA1 gene is decreased in certain sporadic human breast cancer cell lines and tumor samples
- A duplication (pseudogene) of the first three exons and introns of the BRCA1 gene is integrated ~35 kb upstream of the 5' end of the BRCA1 gene.
- A 1.300 bp HERV-like sequence is located in the BRCA1 pseudogene and shows an up to 95% homology to the HERV-E.PTN element.
- The HERV-E.PTN element is known for its important role in the transcriptional regulation of the human pleiotrophin gene.

Goals

- Characterization of the HERV-like element in the human BRCA1 locus.
- Classification and phylogenetic analysis of the HERV-like element in the human BRCA1 locus.
- Evaluate whether the 5'LTR of the HERV-like element in the BRCA1 pseudogene possesses transcriptional activity.

Since we know that there is HERV-like sequence present in the duplicated region of the BRCA1 gene, it is obligatory to define the structure and sequence of this retroviral element. Additionally, based on our knowledge that HERVs can influence the expression of adjacent genes, we tested whether the retroviral element in the BRCA1 locus has transcriptional activity and the potential to influence the expression of the BRCA1 gene.

3 Material

3.1 Chemicals

Acrylamide / Bis-acrylamide	Sigma-Aldrich
Agarose	Invitrogen™
Ammonium persulfate	Sigma-Aldrich
Boric acid	Bio-Rad Laboratories
EDTA	Sigma-Aldrich
Ethanol	The Warner-Graham Company
Ethidium bromide	Stratagene
Formamide, high purity	Boehringer Mannheim
Glycerol	Sigma-Aldrich
Isopropyl alcohol	Fisher Scientific
1 kb ladder	Invitrogen™
Magnesium chloride	Fisher Scientific
Pellet paint	Novagen
Phenol/Chloroform/Isoamylalc.	Invitrogen™
Potassium chloride	EM Science
SDS (lauryl sulfate)	Sigma-Aldrich
Sodium acetate	EM Science
Sodium citrate	EM Science
Sodium hydroxide, 10N	EM Science
Scintillation fluid	Research Products Int. Corp.
Tris HCL	EM Science
Urea	Sigma-Aldrich
Xgal	EM Science

3.2 Enzymes

Sequenase	Amersham Life Science

*Bam*H I	New England Biolabs
Calf intestinal alkaline phosphatase	Invitrogen™
*Eco*R I	New England Biolabs
*Eco*R V	New England Biolabs
Hind III	New England Biolabs
Kleenow fragment	New England Biolabs
Kpn I	New England Biolabs
Not I	New England Biolabs
Sac I	New England Biolabs
Sma I	New England Biolabs
Taq – Polymerase	Invitrogen™
Tetramethylethylenediamine	Sigma-Aldrich
Terminator Ready Reaction Mix (containing AmpliTaq Polymerase)	Perkin Elmer
T4 DNA Ligase	New England Biolabs
T4 DNA PolynucleotideKinase	New England Biolabs
Xba I	New England Biolabs
Xho I	New England Biolabs

3.3 Radioactive Material

α-^{32}P dCTP	Amersham Life Science
γ-^{32}P ATP	Amersham Life Science

3.4 Molecular Biology Kits

Dual-Luciferase® Reporter Assay System	Promega
Expand™ Long Template PCR	Boehringer Mannheim
Qiaex II Gel Extraction Kit	Qiagen
Qiagen Plasmid Maxi Kit	Qiagen
QiAamp Blood Kit	Qiagen

QIAprep Miniprep	Qiagen
Rediprime II	Qiagen
T7 Sequenase v2.0	Amersham Life Science
TA Cloning Kit	Invitrogen™

3.5 Growth Media Ingredients

Agar	Difco Laboratories
Tryptone	Difco Laboratories
Yeast extract	Difco Laboratories
Ampicillin	Sigma
Zeocin	Invitrogen™

3.6 Buffer and Solutions

Denaturation solution:

- 1.0 M NaCl
- 0.5 M NaOH

Denhardt (50x Stock):

- 5 g Ficoll (Type 400 Pharmacia)
- 5 g polyvinylpyrolidone
- 5 g BSA (Fraction V, Sigma)
- add H_20 up to 500 ml

Hybridization membrane washing solution (low stringency):

- 2x SSC
- 0.1% SDS

Hybridization membrane washing solution (high stringency):

- 0.1x SSC
- 0.1% SDS

Neutralization solution:

- 1.5 M NaCl
- 0.5 M Tris-HCl (pH 7.2)
- 1 mM EDTA

Southern prehybridization solution I:

- 5x SSC
- 5x Denhardt
- 0,5% SDS
- add 100 µl denaturated salmon sperm (10 mg/ml) / 10 ml solution

Southern prehybridization solution II:

- 5x SSC
- 5x Denhardt
- 0,5% SDS
- 50% Formamide
- add 100 µl denaturated salmon sperm (10 mg/ml) / 10 ml solution

Prehybridization Solution for Genomic Southern:

- 5x SSC
- 5x Denhardt
- 0,5% SDS
- 50% Formamide
- 10% Dextran sulfate
- add 10 µl denaturated Salmon Sperm (10 mg/ml) / 10 ml solution

Sequencing gel matrix:

- 168 ml H_20
- 48 ml H_20
- 72 ml 38% Acrylamide / 2% Bis-Acrylamide solution
- 240 g Urea

This solution can be stored at 4°C.

SSC (20x):

- 3 M NaCl
- 0,3 M Na_3Citrat

TBE (10x):

- 1 M Tris
- 0,83 M Boric acid
- 20 mM EDTA

3.7 Growth Media for E.coli

Two liquid growth media were used throughout our studies:

Luria-Bertani (LB) Media:

- 1 % Tryptone
- 0.5 % Yeast extract
- 1 % NaCl

SOC media:

- 2 % Tryptone
- 0.5 % Yeast extract
- 10 mM NaCl
- 2.5 mM KCl
- 10 mM $MgCl_2$
- 10 mM $MgSO_4$
- 20 mM Glucose (dextrose)

3.8 Media for Human Cell Lines

Freezing medium (w. DMSO)	GibcoBRL™
Fetal bovine serum	GibcoBRL™
IMEM	GibcoBRL™
OPTIMEM®	Invitrogen™
PBS	GibcoBRL™

Trypsin-EDTA	GibcoBRL™

3.9 Transfection Material

LIPOFECTAMINE™	Invitrogen™

3.10 Working Material

Centrifuge Sorvall® RC-5B	DuPont Instruments
Chroma Spin Columns	Clontech
Chromatography Paper	Whatman
Eppendorf Pipettes	Continental Lab Products
Hyperfilm™ MP	Amersham Life Science
Micro Bio-Spin® 30 columns	Bio Rad
Microcentrifuge Collection Tubes	Bio Rad
Nitrocellulose Membrane	Schleicher & Schuell
Nylon transfer membrane	Magnacharge
Parafilm	American National Can
Quartz Cuvette 8 mm	Beckman
Spectrophotometer DU® 640	Beckman
Table Centrifuge 5415 C	Eppendorf
UV Stratalinker 1800	Stratagene

3.11 Primers

All primers are listed in 5' → 3' orientation.

3'LTR rev:

TCCATGACACCAACCATCC TM: ~58°C

4kb-1F:

GTGTAAGATAGACTGGCCAGC TM: ~64°C

4kb-1R:

GCTGGCCAGTCTATCTTACAC TM: ~64°C

4kb-2F:

GGGTACATGTGCACAACATGC TM: ~64°C

4kb-3F:

GTAAGGAGGCATAAGGATGTAC TM: ~64°C

4kb-4F:

CATACGAGAAGCACTGTTAAATG TM: ~64°C

A9/302-R:

GTGGCACATGTGCCCGTG AS TM: ~62°C

A9/477-F:

GAATAGTCAAGGCTCTGCCGG TM: ~66°C

A9/477-R:

CCGGCAGAGCCTTGACTATTC AS TM: ~56°C

A9/578-F:

GCATTGATCTGCTATTAC TM: ~50°C

A9/581-F:

GTGTCACTTACGCCACCTG TM: ~60°C

A9/D12:

CAACTAGGCTCTCTGATAC TM: ~56°C

A9/SP6-F:

GCCGTAGTTGGTCCAATGTTC TM :~54°C

A9/SP6-R:

GAACATTGGACCAACACGGC AS TM: ~54°C

A9/SP6-F2:

ACTTAGCGTTCCTTGGAGAC TM: ~60°C

B1:

GGAGCTTCGCTCTTGTGC TM: ~58°C

B11-F:

GTTCATTGGAACAGAAAGAG TM: ~56°C

B12-R:

GCAGAGTGGATGGAGAAC AS TM: ~56°C

B2:

GACTTACCAGATTGGACACTG AS TM: ~62°C

B3:

TCAGTCACTCCTCTGTAG TM: ~54°C

B4:

CAGATGGGACACTCTAAG AS TM: ~54°C

B5:

GAGTAGCTGGAGCGGCAC AS TM: ~60°C

B6:

CCGAAACTGGAGACCTC TM: ~58°C

B7α:

GAGACGCTTGGCTCTTTC TM: ~56°C

B8/LTR:

CTTCTTTAACTCGGTGTCTG TM: ~58°C

B9/SP6-R:

GAACATTGGACCAACTACGGC TM: ~64°C

B9/SP6-2R:

GGTGTGACGAGTCCATCC TM: ~58°C

B9/SP6-F:

GCCGTAGTTGGTCCAATGTTC TM: ~64°C

B9/SP6-F2:

ACTTAGCGTTCCTTGGAGAC TM: ~60°C

BR 1.0:

TGTCTTGATTGGTTCTGCACTGG TM: ~68°C

BR 1.2:

GGCAGGAATCTAGTTTAGATTAACTG TM: ~72°C

BR 1.3:

CTCAGGATCCGTGTTGCTCGG TM: ~68°C

D12/364-R:

GCTTCTTGTGCTAACAGGGC TM: ~62°C

D12/A9:

GCCCTGTTAGCACAAGAAGC TM: ~62°C

4 Methods

4.1 Sterilization of Solutions and Working Material

Solutions, microtubes, tips and other items were sterilized by using a standard autoclave.

4.2 Gel Electrophoresis of Nucleic Acids

4.2.1 Native Agarose Gels

Agarose / TBE gels were used to define the length of DNA.

After melting the agarose in 120 ml 1x TBE buffer, ethidium bromide was added prior to polymerization in a gel tray. The intercalation of ethidium bromide with DNA allows the visualization of DNA under UV light afterwards. Depending on the size of the to analyzing DNA, agarose amounts of 0,7% up to 1,4% were used.

The DNA samples were loaded on the polymerized gel together with a DNA ladder (1 kb) serving as a size reference. The separation was conducted at ~ 100V.

4.2.2 Sequencing Gel

Two glass plates, one large and one small one, were cleaned with water, 70% ethanol and then with isopropanol. The inner side of the small plate is coated with silicone to ensure that the gel will not stick upon removal. Spacers were placed on both sides and at the bottom and the plates were tightly clipped. The sequencing gel was prepared as described in [Sambrook et al., 1989].

The gel was poured immediately and a comb was placed on the top side. After a 45 min polymerization, the comb and bottom spacer were removed and the gel was placed in the apparatus between two chambers and after adding the comb separator, 1x TBE buffer was added to both chambers. We pre-ran the gel at 1550 V for about 45 min to remove “matrix worms” created by the presence of urea and to equilibrate the matrix.

After loading of the samples the separation was conducted for several hours at 1900 V.

4.3 Spectrophotometer Measurements

The spectrophotometer was chosen to measure the concentration of DNA in solution. For this purpose the optical density of a 1:50 dilution of our DNA sample in 100 µl purified water was measured in a Quartz cuvette at 260 nm. The wavelength corresponds to the average absorption maximum of the four bases ranging between 258 to 261 nm.

The concentration of the sample was determined by the fact that absorption of 1 at 260 nm is equivalent to a DNA concentration of 50 ng/µl.

4.4 Cultivating Bacteria

Different strains of *Escherichia coli* (DH5α and Top10F') were cultured in liquid growth medium (LB or SOC) or on agar plates depending on the experiment chosen.

4.4.1 Liquid Cultures

To reactivate bacteria obtained from a frozen culture 2-5 ml medium were inoculated with a small amount of frozen bacteria stocks and incubated at 37°C for at least 4-6 hours at 200 rpm. For preparative generation of bacteria, the starting culture was transferred into 500 ml growth medium and incubated overnight under the same conditions.

4.4.2 Plate Cultures

To analyze certain colonies, bacteria solution was plated in various dilutions on agar plates and incubated overnight at 37°C.

4.4.3 Frozen Cultures

Exponentially grown bacteria solution were mixed with sterile Glycerol in a 1 : 0.25 ratio, mixed and stored at -80°C.

4.5 Restriction Endonuclease Assay

DNA restrictions were used for either cloning purposes or to verify a recombinant clone due to its specific restriction pattern.

The enzymatic conditions were used as recommended by the vendor. Restriction assays were performed in at least 20 µl volume usually containing 1x restriction enzyme buffer, 1 µl enzyme (maximum is 1/10 of the volume) and DNA ranging from 500 ng to several µg.

4.6 DNA Extraction from Agarose Gels

In order to purify DNA samples from 280 bp up to 5 kb we extracted and purified the samples from agarose gels. The system in the QIAEX II kit is based on the adsorption of nucleic acids to silica particles, while impurities can be removed by several washing steps.

The DNA band was excised from the gel under UV light and transferred into a 1.5 ml microtube. Depending on the weight of the gel and the size of the DNA fragment, a suitable volume of Buffer QX I was added, following the manual. Buffer QX II was resuspended by vortexing it for 30 sec to ensure that the silica particles were in solution and an amount according to the manual was added. Following a vortexing step the tube was incubated for 10 min at 50°C, interrupted every 2 min by a short vortexing step to keep QIAEX II in suspension. After this, the sample was centrifuged at 13,000 rpm for 30 sec at room temperature and the supernatant was carefully removed. To start the washing procedure, we added 500 µl QX I buffer and resuspended the pellet. After another centrifugation (s. above for details), the sample was washed twice with 500 µl PE buffer to remove residual salt contaminants. The pellet was dried overhead until it turned white. To remove the nucleic acid from the silica particles, we added 100 µl H_20 and resuspended the pellet followed by a 30 sec centrifugation. The supernatant containing the DNA was carefully transferred into a new microtube and the concentration was determined by spectrophotometry.

4.7 DNA Purification

To remove proteins from DNA samples the procedure of choice is a Phenol : Chloroform extraction. The fact that nucleic acids are negatively charged and therefore will be found in the aqueous phase is used in this basic technique. We used PCI (phenol : chloroform : isoamylalcohol – ratio 25:24:1). The two organic solvents improve the deproteinization.

The volume of the sample was filled up to 500 µl with H_20. Then we added 500 µl cold PCI and vortexed for 30 seconds followed by a centrifugation step at 13,000 g for 5 min at RT. The phases were clearly separated and the aqueous phase containing the nucleic acid was carefully transferred into a new Eppendorf tube.

After defining the volume, we added 1/10 of the volume 3M Sodium acetate (pH 5,2) and 2.5x of the volume 100% cold Ethanol. The probes were vortexed and stored at –80°C for at least 30 min. After that, the DNA was sediment at 13,000 g for 30 min at 4°C. After removing the supernatant, 600 µl 70% cold Ethanol were added to reduce the salt content of the solution. The final washing step is a centrifugation at 13,000 g for 15 min at 4°C, after which we removed the supernatant and let the sample dry overhead for about 15 min. The sample was then resuspended in a suitable volume of H_20.

4.8 DNA Cloning

Cloning of DNA fragments into plasmid DNA was used during this study for various purposes. Either DNA samples had to be amplified for further procedures or promoter reporter gene constructs were generated to test for potential transcriptional activity in mammalian cells.

4.8.1 Ligation of DNA Fragments after Restriction Enzyme Modification

We used commercially available vectors during our studies (pCR and pXP) which contain multiple cloning sites with a variety of restriction sites. After the proper restriction sites were defined, vector and insert were digested in a preparative amount using the same restriction enzymes. A gel check of the cleaved vector DNA was

performed to prove a successful digestion. Insert and vector now have the same termini. To prevent the vector from religation, we always performed a dephosphorylation with calf intestinal phosphatase as described in the Sambrook / Fritsch / Maniatis laboratory manual. This enzyme removes the 5' phosphate groups from linear DNA.

To determine the optimal ligation efficiency we always used 4-5 different molar ratios of vector : insert DNA (s. experiments).

The molecular weight of 1 bp corresponds to 660 g / mol

vector size (bp) x 660 g/mol = (X) g / mol

= (X) µg / µmol

The amount of DNA product needed for each ligation reaction was calculated by using the following formula:

$$\mathbf{X}\text{ ng product} = \frac{(\mathbf{Y}\text{ bp of DNA product}) \times (50\text{ ng vector})}{(\text{size in bp of the vector})}$$

X represents the DNA product amount, which is supposed to be ligated in a 1 : 1 ratio.

We performed ligations in a final volume of 10 µl containing H_20, 1 µl T4 DNA Ligase Buffer, 0.02 pmol vector DNA, insert DNA in multiple ratios and 1 µl T4 DNA Ligase. The reactions were incubated overnight at 16ºC. One reaction was always performed without insert DNA as a background check to determine the yield of re-ligations.

4.8.2 Transformation

Ligation reactions where transferred into competent E.coli cells. For this purpose DH5α or Top10F'cells were thawed on ice after which, 50 µl cell solution and 2 µl ligation reaction were added together into a 10 ml tube. After 30 min incubation on ice the cells were heat shocked at 42°C for 45 sec followed by an additional 2 min on ice. This step opens pores within the bacterial cell membrane for the uptake of plasmid DNA. 500 µl growth medium were added to each tube and the cells were

incubated at 37°C for 45 min at 200 rpm allowing the antibiotic resistance gene to be expressed.

Following the incubation, 50 µl of each reaction were plated on LB – Ampicillin plates. The remaining cells were harvested after centrifugation at 4°C for 5 min at 3,000 rpm. The supernatants were discarded and the cell sediments were plated completely onto LB – AMP plates.

4.9 Transfer of DNA to Hybridization Membranes

4.9.1 Bacteria Colony Transfer

This methodology was chosen to define a clone of our interest in a subclone library via hybridization experiments.

The agar plates containing the subclone colonies were placed next to an open flame to minimize aerosol contamination. A membrane was then carefully lowered onto the agar surface and spread evenly by using a sterile glass spatula. It is important to mark the orientation on the membrane itself and on the plate.

The membranes were then transferred into a petri dish containing Whatman paper, barely covered with denaturation solution, for 10 min. Two neutralization steps (10 min each) followed in petri dishes containing Whatman paper soaked in neutralization solution. At the last step, the membranes were briefly submersed into a 2x SSC solution to remove the cell debris. The damp membrane was then crosslinked through either UV light or heat exposure.

4.9.2 Southern Blotting

Southern blotting was used to retrieve DNA from an agarose gel. The transfer was performed as described in the Sambrook/Fritsch/Maniatis laboratory manual .

Followed by electrophoresis, a gel picture was taken with a ruler to facilitate the size determination of hybridization signals afterwards. The gel was initially denatured for 15 min in denaturation solution (2x) followed by three neutralization steps, each for 10 min.

Before transfer, we cut the upper right corner of the gel for orientation. The gel was flipped and placed on two layers of soaked Whatman paper on a glass plate. The

lower one reaches into a container filled with 20x SSC solution. A gel size nylon transfer membrane was put on top of the gel and covered with two fitted and soaked Whatman papers. Three dry Whatman papers followed and a stack of fitted paper towels was put on top. Adding weights helps to evenly distribute pressure to the gel. The transfer was performed overnight.

4.10 Plasmid Preparation

4.10.1 Plasmid Mini Preparation with QIA-prep Spin Columns

For analytical use, only small amounts of DNA were needed. Thus overnight cultures in 8 ml LB media were sufficient. For harvesting, the cells were centrifuged at 3,000 rpm at 4°C for 5 min. The sediment was resuspended in 250 µl P1 buffer (containing RNase A) and transferred into a labeled microtube. After adding 250 µl P2 buffer (NaOH-SDS), the tubes were gently inverted 4-6 times. To stop the lysis reaction 350 µl neutralization buffer N3 (acidic K-Ac) were added to each tube after which they were inverted 4-6 times. The solution was centrifuged at 13,000 rpm for 10 min. In the meanwhile, the spin columns were placed in a 2 ml collection tube. The supernatant of the centrifugation step was carefully applied to the QIAprep column followed by a 45 sec centrifugation step. The flow-through was discarded and 750 µl washing buffer PE were added to the column. A 45 sec centrifugation step, and after discarding the flow-through, a 60 sec centrifugation step completed the washing procedure. A clean 1.5 ml microtube was placed under the column and the DNA was eluted with 50 µl H_20.

4.10.2 Plasmid Maxi Preparation by Qiagen

To generate a preparative amount of bacteria from a frozen sample or a colony, we usually prepared 2 ml starting cultures, containing growth media and the selecting antibiotic (e.g. LB and ampicillin). After 4-6 hours incubation at 37°C and 200 rpm, the solution was transferred into a sterile glass bottle containing 500 ml of the same media, followed by an overnight incubation under the same conditions.

At the next day, the cells were sedimented at 6,000x g for 15 min at 4°C and resuspended by adding 10 ml buffer P1 (containing RNase). The solution was

transferred to a 50 ml spin tube. After addition of 10 ml lysis buffer P2 (NaOH-SDS) the tube was inverted 4-6 times and incubated for 5 min at room temperature. Adding 10 ml P3 buffer (acidic K-Ac) neutralizes the reaction and inverting the tube 4-6 times precipitates the SDS, denaturated proteins, cell debris and chromosomal DNA through forming salt detergent complexes. The plasmid DNA stays in solution. We enhanced this reaction by using chilled P3 buffer and by incubating the samples on ice for 20 min, followed by a centrifugation at 20,000x g for 30 min at 4°C. The supernatant was then transferred into a new tube and re-centrifuged for 15 min under the same conditions. In the meanwhile the Qiagen-tip 500 columns had been equilibrated with 10 ml buffer QBT. We let the buffer flow through by gravity force. The equilibration ensures that the salt and pH conditions of the filter membrane allow only plasmid DNA to bind, while other products such as proteins or cell debris are not held back. The supernatant was then applied to the tip and allowed to flow through by gravity as well. Washing the membrane twice with 30 ml buffer QC is necessary to remove contaminants of the preparation and carbohydrates, which occur using certain bacteria strains.

After that, the column was mounted onto a sterile 30 ml glass spin tube and the retained DNA was washed off the membrane by using 15 ml elution buffer QF. Precipitation of the DNA was performed by adding 0.7 volumes of isopropanol to the eluted DNA solution followed by mixing and a centrifugation step at 15000x g for 30 min at 4°C. To eliminate precipitated salt, the DNA pellet was washed with 70% ethanol and centrifuged at 15,000x g for 15 min at 4°C.

As a last step the pellet was dried overhead for 5-10 min and dissolved in a suitable volume of H_2O, e.g., 100 µl.

4.11 Microspin Chromatography Columns

BIO-RAD Micro Bio-Spin® Chromatography columns were used prior to automatic sequencing to remove dye-labeled dideoxynucleotides from the sequencing reaction.

AMERSHAM PHARMACIA BIOTECH MicroSpin™ S200 Columns were used to purify random prime labeled reactions.

4.12 PCR

The Polymerase Chain Reaction was performed to amplify determined stretches of DNA by using two flanking primers – one in sense and one in antisense orientation.

4.12.1 Reactions

All reactions were performed in 0.7 ml microtubes using Invitrogen™ recombinant Taq DNA Polymerase kit.

Amplification of DNA fragments was performed in a final volume of 50 µl containing the listed components:

39.0	µl	sterile H_20
5.0	µl	10x PCR buffer
1.5	µl	$MgCl_2$ (0.75 mM)
1.0	µl	DNA of PAC1993 (50 ng)
1.5	µl	of each primer (150 ng)
1.0	µl	dNTP polynucleotide mix (10 mM each)
0.5	µl	Taq Polymerase (5 U/µl)

Reactions were overlaid with ~ 30 µl mineral oil

4.12.2 Cycling Conditions

A single 5 min denaturation at 94°C was followed by 30 cycles of denaturation for 45 sec at 94°C, 45 sec annealing at a temperature depending on the primers melting point and a 45 sec elongation period at 72°C. The final step is a single elongation period for 5 min at 72°C to ensure double strand character.

10µl of the product were then analyzed on a standard 1% agarose gel with a 1 kb DNA ladder for size determination.

4.13 Long Template PCR

4.13.1 Reactions

We used a long template PCR system to generate PCR products over 2000 bp, since recombinant *Taq* Polymerase does not have proof reading capacity. In absence of proofreading *Taq*, the polymerase would fall of the strand due to the accumulation of too many mismatches. Therefore this system uses an enzyme mix consisting of recombinant *Taq* Polymerase and *Pwo* Polymerase. The latter has proof-reading capacitiy. The disadvantage of the *Pwo* enzyme, cutting of the A on the 3' overhang, is minimal due to its very low concentration in the mix.

All reactions were performed in 0.7 ml microtubes using the Expand™ Long template PCR System (Boehringer Mannheim).

Amplification of DNA fragments was done in a total reaction mixture of 50 µl containing the listed components:

38.5	µl	sterile H_20
5.0	µl	10x PCR buffer (#1,2 or 3 depending on the reaction)
1.0	µl	template DNA (50-300 ng)
1.5	µl	of each primer (150 ng)
1.75	µl	dNTP polynucleotide mix (350 µM each)
0.75	µl	enzyme mix (Taq / Pwo approx. 2.5 U / reaction)

Reactions were overlaid with ~ 30µl mineral oil

4.13.2 Cycling Conditions

After 5 min denaturation at 94°C, 30 cycles with a denaturation step for 40 sec at 94°C, a 40 sec annealing step at a temperature three degrees below the primer melting point and a 6 min elongation step at 68°C followed. The final step is a single elongation period for 15 min at 68°C.

10 µl of the PCR products were then analyzed on a standard 0.8% agarose gel with a 1 kb DNA ladder for size determination.

4.14 Radioactive Labeling of Nucleic Acids and Hybridization

Radioactive nucleotides were used to label oligonucleotides or to perform Random Primed Labeling.

4.14.1 Labeling of Oligonucleotides

In order to detect a certain complementary DNA sequence, e.g. on a blot , radioactive labeled oligonucleotides were used as probes.

The labeling reaction itself was performed in a volume of 20 µl containing 1 µl PNKinase (PN = polynucleotide), 2 µl 10x PNKinase Buffer, 150 ng oligonucleotide and 50 µCi γ-^{32}P ATP. The enzyme was added last and the sample was incubated for 30 min at 37°C.

4.14.2 Random Prime Labeling for Genomic Southern

Random primed labeling was done by using the Rediprime™ II kit provided by Amersham Pharmacia Biotech. Feinberg and Vogelstein first introduced this technique. It takes advantage of the action of the Kleenow fragment of DNA polymerase I, which uses a primer DNA template (denatured DNA) complex as a substrate. In this kit the concentration of dCTP is 20x lesser than the concentration of the other nucleotides so that the newly synthesized DNA will incorporate radioactive α-^{32}P dCTP.

Reaction layout:

60 ng double stranded DNA and 1x TE buffer were mixed with sterile H_2O to a final volume of 45 µl and then denatured for 5 min at 95°C. The denatured DNA was then transferred into the reaction tube containing the kit. 5 µl α-^{32}P dCTP were added and the solution was mixed, by pipetting it up and down about 10 times. An incubation at 37°C for 20 min followed, before the mix was loaded onto a prepared gel permeation chromatography column to remove remaining free nucleotides. 1 µl of the solution was transferred into a glass vial containing 2 ml scintillation fluid to define the specific activity of the probe in a scintillation counter. Only probes with a total activity higher than 2 x 10^7 cpm were used during our experiments.

4.14.3 Pre- and Hybridization with Labeled DNA

The membrane containing the crosslinked DNA were carefully rolled and inserted into a hybridization tube, the surface with the nucleic acids facing inwards. Approximately 10 ml prehybridization solution (depending on the blot type) was added and the tube was incubated in a hybridization oven.

- Random primed probes:
  for at least three hours 42°C
- labeled oligonucleotides
  for two hours at 42°C

The radioactive probe was tested in a radiation counter and were used if the result was < 2 x 10^6 cpm / min / ml hybridization fluid.

After the prehybridization, the radioactive probe was added and the membrane was incubated overnight under the same temperature conditions.

4.14.4 Washing Conditions

4.14.4.1 Southern Blot and Bacterial Membranes

Membranes of these types where washed in two high stringency steps.

1. First the radioactive hybridization solution was transferred into a glass vial and stored for possible further use. Then ~ 15 ml 2x SSC / 0.1% SDS were added and the hybridization tube was rotated for 15 min at 42°C.
2. After discarding the washing solution, ~ 15 ml 0.1x SSC / 0.1% SDS solution was added and the rotating step was repeated under the same conditions.

4.14.4.2 Genomic Southern Membranes

Genomic Southern membranes were washed in three or more steps, depending on the background activity.

4.14.5 Autoradiography of Radioactive Membranes

After the necessary washing steps, the membranes were taken out of their hybridization tubes and sealed into plastic foil. After checking the activity with a

Geiger – Mueller counter, the sealed membranes were put into a KODAK™ film cassette and Hyperfilm MP™ was added in the darkroom. The cassette was then stored for exposure at -80°C for a time depending on the activity of the probe. During our studies exposure times ranged from 20 min to 5 days.

4.15 Sequencing after Sanger

Sanger's introduction of chain terminating dideoxynucleoside triphosphates (ddNTPs) in 1977 revolutionized enzymatic sequencing.

The lack of a hydroxyl residue at the 3' position deoxyribose in 2'3' ddNTPs prevents the formation of a phosphodiester bond with the succeeding dNTP. In the reaction mix for DNA synthesis a small amount of one 2'3' ddNTP type is present with conventional dNTPs, which leads to competition between termination (ddNTP) and extension (dNTPs). The products of the reaction are a variety of DNA fragments with different lengths. Each of the four ddNTPs is florescence labeled and reacts with a Rhodamine dye, so that each of the four bases can be defined by their typical light emission.

Sequencing reactions were analyzed on an ABI prism™ 377 DNA Sequencer (Perkin Elmer).

Procedure:

Reactions were performed in thin-walled 0.5 ml PCR microtubes. 400 ng template DNA were added to 10 ng primer, 4 µl Terminator Ready Reaction Mix™ (Perkin Elmer) containing the ddNTPs and filled up with sterile H_20 to a final volume of 10µl. Each reaction was overlaid with ~30 µl mineral oil.

Cycle sequencing conditions started with 30 sec denaturation at 96°C, followed by a rapid thermal ramp to 50°C. After 15 sec, the reactions were rapidly heated to 60°C. This cycle was repeated after 4 min on 60°C 25 times. A final thermal ramp cooled the reactions to 4°C.

The samples were purified over gelfiltration chromatography columns (Micro Bio-Spin® 30) and dried in a Speed Vac.

4.16 Cultivating and Transient Transfection of Human Cell Lines

4.16.1 Human Cancer Cell Lines used in this Study

- BeWo:
  Choriocarcinoma cell line
- JEG-3:
  Choriocarcinoma cell line
- JAR:
  Choriocarcinoma cell line.
- HS578T:
  Breast cancer cell line, derived from a 74 year old patient with ductal carcinoma. HS578T is a spontaneous breast cancer cell line.

4.16.2 Cultivating Human Cells

Two types of cultivating flasks were used during this study. 75 cm^2 flasks were usually filled with 12 ml growth media and 175 cm^2 flasks with 25 ml growth media. The flasks containing the media and the cells to cultivate were stored in standard humidified chambers at 37°C and 5% CO_2. The media was changed every 48 to 72 hours .

4.16.2.1 Media

The growth medium we used was IMEM with 10% fetal bovine serum (FBS)

4.16.2.2 Preparation of Cell Extracts

In order to use the same amount of cells for each transfection, the cells had to be harvested through trypsination and counted. 2 ml for the 175 cm^2 and 1 ml trypsin for the 75 cm^2 flasks were used for this purpose. First the growth media was sucked out of the flask with a sterile pipette. The cells were washed with 2 respectively 1 ml trypsin which was discarded afterwards. After adding the proper amount of trypsin, the flasks were incubated for 5 to 15 min at 37°C depending on the cell line. 10 ml growth media were added to the flask and pipetted up and down to get the cells into solution. A sterile tube containing 15-30 ml growth media was prepared and 10 ml of

the cell suspension was added. To define the amount of solution to be used in order to obtain 200,000 cells, the cells were counted using a standard counting chamber. The mean number of cells multiplied by 10,000 resulted in the amount of cells per ml contained in solution.

4.16.3 Transient Transfection

All transient transfections were performed in triplicates using 6 well plates. The wells were each filled with 2 ml IMEM + 10% FBS. The proper amount of cells (200,000 cells / well) was added and the cells were incubated under standard conditions for 12-16 hours.

Transfections were performed with lipofectamine reagent in concentration of 7 µl / µg DNA. Lipofectamine is positive charged, builds lipophilic complexes around the DNA and can penetrate the cell membrane. We prepared sterile tubes (A + B) containing the following solutions:

A: (for one triplicate)

- 30 µl plasmid DNA (100 ng / µl)
- 300 µl OPTIMEM media

B: (premix / each well)

- 7 µl lipofectamine
- 100 µl OPTIMEM
- 2.5 ng *Renilla Luciferase* Plasmid DNA

Tube B was vortexed after adding the contents.

To each prepared tube A, we added 300 µl of the contents from premix tube B, after which they were vortexed for a few sec and incubated for 30 min at RT to allow the lipofectamine / DNA complex to be formed.

In the meanwhile, six well plates were washed twice with OPTIMEM to clear the wells from FBS, which interferes with the lipofectamine. Afterwards 500 µl OPTIMEM media were filled in each well.

After 30 min, 900 µl OPTIMEM were added to each prepared tube with lipofectamine / DNA complex and the solution was pipetted up and down twice. 500 µl were added to each well, followed by a 5 hour incubation at standard

conditions. After five hours the solution was sucked out of each well and 2 ml IMEM + 10% FBS were added followed by a 24-28 hour incubation under standard conditions.

4.16.4 Harvesting the Cells

After the incubation period, the media was discarded and the wells were washed twice with 2x PBS. 150µl lysis buffer (contained in the kit) was added to each well, incubated at RT for ~10 min and scraped afterwards with a new spatula for each well. The scraped cell solution was added to prepared microtubes (1 / well) and centrifuged at 12000 rpm for 10 min. The supernatant contains the cell extract while the pellet is formed by cell debris.

4.16.5 Dual-Luciferase® Reporter Assay

During this study, a standard lumimometer was used to define light emission. The assay was performed as described in the Dual-Luciferase® Reporter Assay System manual provided by Promega.

First 100 µl Luciferase Assay Reagent II (LAR II) were pipetted into an empty glass tube. We added 2-4 µl cell extract and measured the luminescent signal created by the firefly luciferase. To quench this reaction and to define the activity of the *Renilla luciferase* as our internal transfusion control, 100 µl Stop&Glo® reagent were added. The sample was vortexed and measured immediately to quantify the *Renilla* luciferase activity. By using a dual reporter system, we could define the activity of the co-transfected *Renilla* reporter as an internal control for transfection efficiency and therefore were able to normalize the data received from the firefly reporter.

5 Results

5.1 Sequence and Structure of a Human Endogenous Retrovirus-like Element (HERV) located in the Human BRCA1 Pseudogene

5.1.1 Screening of the PAC clone 1993 and the *Hind* III and *Kpn* I Subclone Libraries

Based on our finding that HERV-like sequence is located downstream of exon 2 in the BRCA1 pseudogene, a PAC (P1 Artificial Chromosome) clone containing the BRCA1 gene and pseudogene region was obtained from Genome Systems Inc. The company defined the PAC clone (PAC1993) via PCR screening using a primer pair specific for the BRCA1 pseudogene (B1 and B2) and a primer pair specific for the BRCA1 gene (B3 and B4), which results in one 377 bp and one 451 bp fragment respectively.

The positive PAC clone was then afterwards digested with *Hind* III and *Kpn* I to receive two shotgun subclone libraries. However screening with various oligonucleotides retrieved from the 5' LTR of the known HERV sequence failed to identify positive subclones in those libraries.

Nevertheless, three PCRs performed in our lab could prove that the PAC clone 1993 contains the desired region of the BRCA1 locus (s. figures 3 and 4).

1. Primers BR1.3 and BR1.0 (pos. 3340 and 2583 GenBank # U77841) amplified the start of the retroviral element
2. Primers B1 and B2 (pos. 1822 and 2286 GenBank # U77841) amplified a fragment including exon 2, in the BRCA1 pseudogene
3. Primers B5 and B6 (pos. 2091 and 1201 GenBank # U37574) amplified a fragment of BRCA1 gene promoter region (s. figure 3)

Furthermore, digested PAC1993 DNA was hybridized with oligonucleotides specific for the BRCA1 gene (B3), the BRCA1 pseudogene (B2) and for the HERV element (B8-LTR) as shown in figure 5.

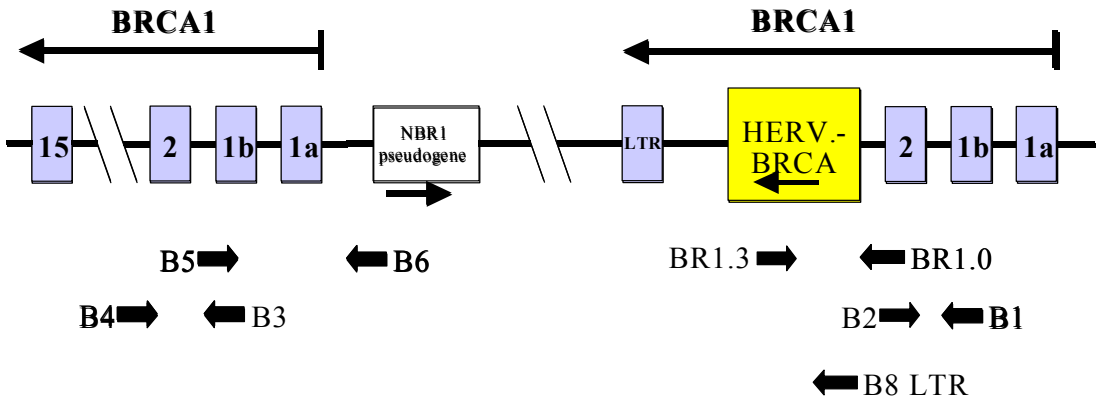

Figure 3:
Primer Locations (s. page 39) of the used in the characterization of the PAC clone 1993 and of its *Hind*III and *Kpn*I subclone libraries. This figure is not representing actual scale.

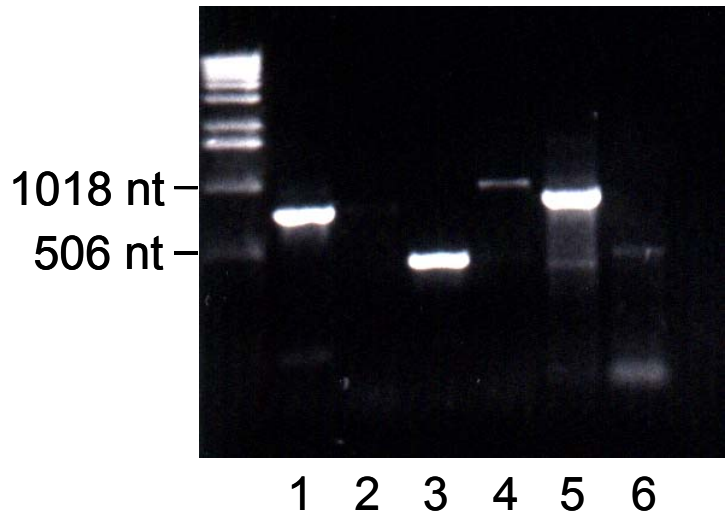

Figure 4:
PCRs with 60 ng PAC1993 DNA serving as template and primer combination BR1.3 and BR1.0 (lane 1) specific for the retroviral element in the BRCA1 pseudogene. This product is 802 nt long. Primer pair B1 and B2 (lane 3) is specific for the BRCA1 pseudogene and the product is 464 nt long. Primer pair B5 and B6 gives an 890 nt long product, specific for the BRCA1 gene. There are unspecific bands in the water controls (lanes 2, 4 and 6), but their length differs from the expected PCR products.

Additionally, sequencing of the PAC1993 clone with primer B1 resulted in BRCA1 pseudogene sequence as predicted. All these results lead us to three assumptions:

1. PAC1993 clone contains the region between the BRCA1 gene and the BRCA1 pseudogene.
2. A HERV-like sequence is located downstream of the exon 2 of the BRCA1 pseudogene and is therefore located on the PAC1993

clone.

3. Due to the striking similarity of the first 800 nt retroviral sequence to the recently reported HERV-E.PTN element, I speculated that a HERV-E.PTN similar element is located on the PAC1993 clone.

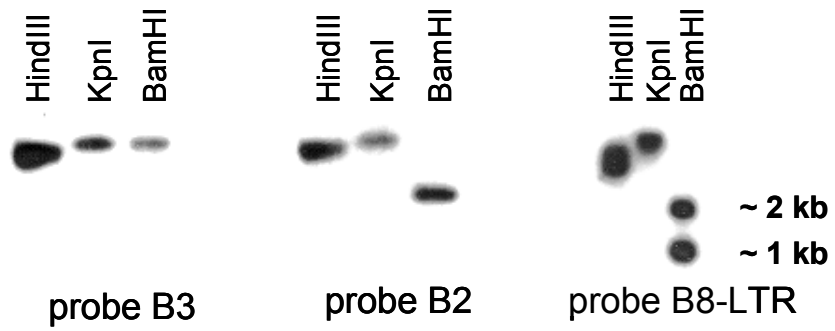

Figure 5:
Southern blot of PAC1993 DNA, digested with the endonucleases *Hind* III, *Kpn* I and *BamH* I. Oligonucleotides specific for the BRCA1 gene (B3), the BRCA1 pseudogene (B2) and the HERV element in the BRCA1 pseudogene, were labeled and used for screening. Positive signals appeared with all probes suggesting that the BRCA1 gene, pseudogene and the retroviral sequence are present in the PAC1993 clone.

I performed a MiniPrep plasmid preparation of each clone of the subclone libraries and developed a Southern blot for further screening. Based on my third assumption that a HERV-E.PTN similar retroviral element is located on the PAC1993 clone, I decided to hybridize the Southern Blot with 480 nt long fragment, derived from the HERV-E.PTN *env* region (shown in figure 6).

Five positive signals could be defined in the *Hind* III library and none in the *Kpn* I library. The five positive subclones were analyzed afterwards by gel electrophoresis and sequencing.

The result of the gelectrophoresis after restriction with endonuclease *Hind* III showed us that the subclones A9 and D10 as well as C12 and D12 are potentially identical. This could be confirmed by sequencing both ends of the clones with the vector specific insert flanking primers SP6 and T7.

Sequencing subclone B8 and comparing the sequence with GenBank entrez, did not reveal any similarity to HERV-E.PTN. Nevertheless, considering the size of the insert, this finding does not necessarily indicate, that there is no HERV-E.PTN sequence present on this subclone, since the hybridization probe was derived from

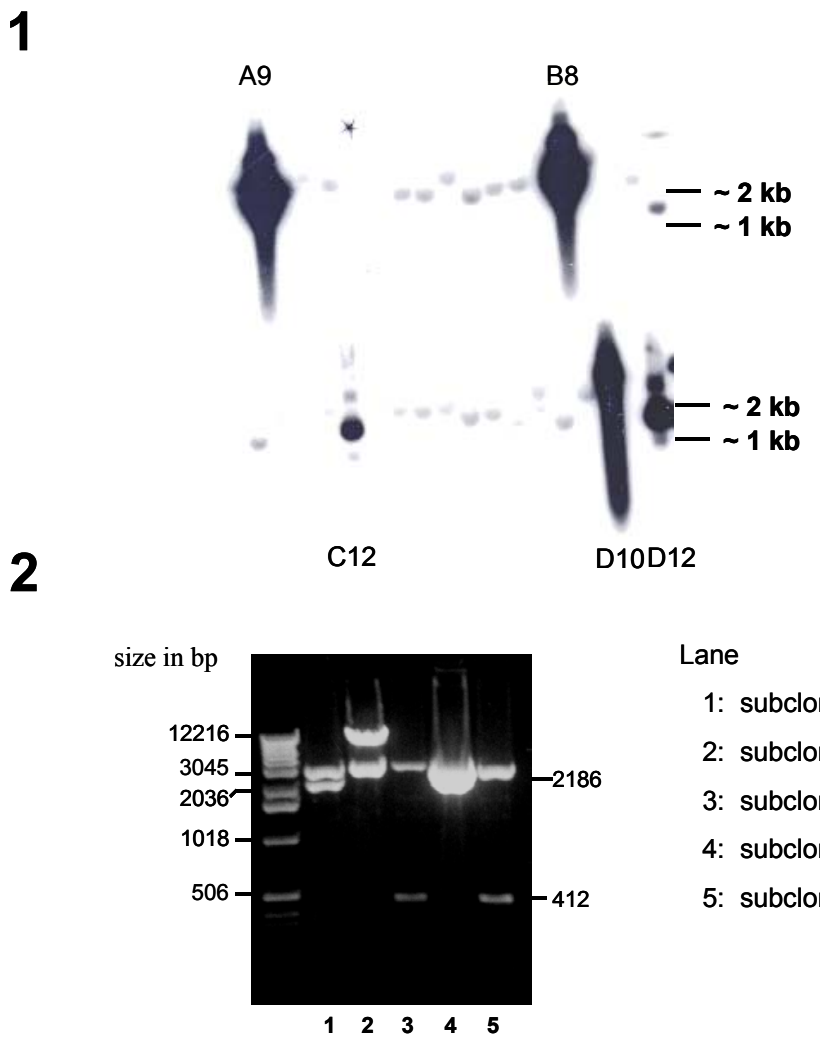

Figure 6:

1. Southern blot transferred from a gel containing all *Hind*III subclones derived from the shotgun subclone library of PAC1993. The used probe was derived from the *env* region of the HERV-E.PTN element. The probe was random primed labeled. Subclones A9, B8, C12, D10 and D12 seem to contain retroviral sequence. Size markers were not a priority, because the positive clones were to be defined, by PCR and sequencing afterwards.
2. DNA of the five positive subclones was digested with *Hind*III and loaded on a 1% agarose gel.. The band showing in all 5 lanes represents the pZErO™ vector backbone (2.8 kb). Lanes 1 and 4 (subclones A9 and D10) seem to be similar. The band representing the insert is 2186 bp long. Lanes 3 and 5 (subclones C12 and D12) seem to be similar as well. The insert is 412 bp long. The fragment inserted in subclone B8 (lane 2) is between 9 and 12 kb long.

the *env* region, it is possible that B8 contains a fragment of a different HERV element.

The subclones A9 and D12 were chosen for further sequencing analysis.

D12 is 462 nt long. GenBank comparisons of the sequence revealed an 88% similarity to HERV-E.PTN (GenBank accession # AF058907) throughout the complete subclone. The first ~60 nt in this region are also homologous to a member of the RTVL-I family. Two stretches which are covering two thirds of D12 are also

up to 100% homologous to the retroviral element located in the Human Hereditary Hemochromatosis locus (accession # U91328). Additionally the last ~380 nt of D12 are up to 89% similar to the HERV-E 4.1 element (GenBank # K02168).

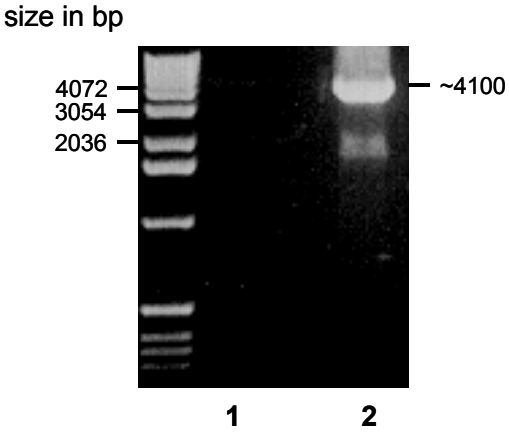

Figure 7:
Picture of a PCR product with a primer pair derived from subclone D12 (D12-364R) and the 5' LTR (B8-LTR). DNA of PAC1993 served as the template. The PCR resulted in a fragment of ~4.1 kb length, defining the localization of subclone D12 in the BRCA1 pseudogene. The fragment was loaded on a 0.8% agarose gel. In Lane 1, 30 ng PAC1993 DNA were used compared to 180 ng in lane 2. The second band (~2000 nt) appearing, is not specific as controlled by Southern.

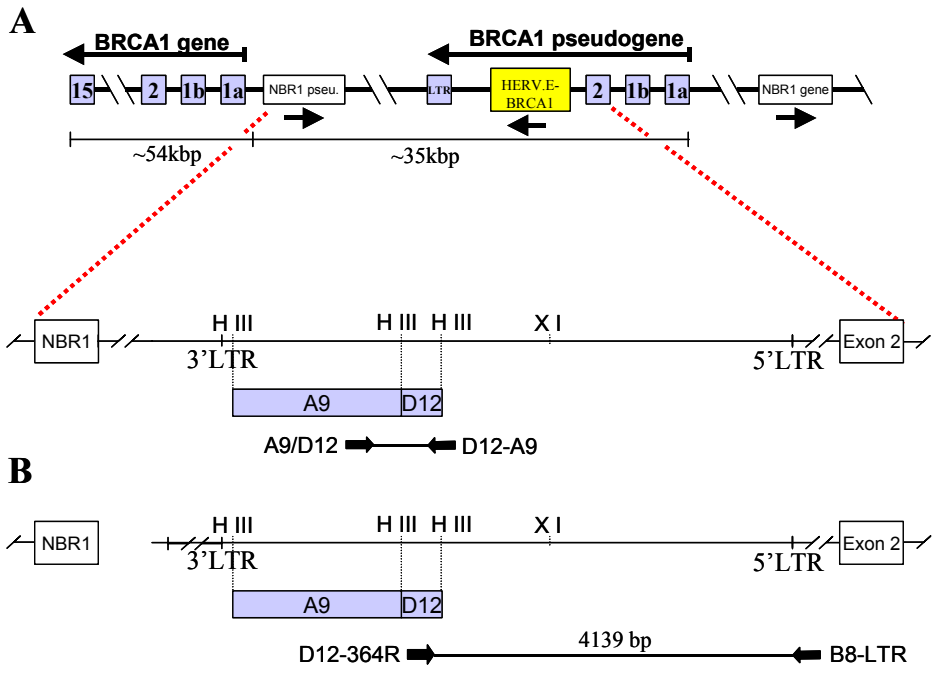

Figure 8:
This figure highlights the two subclones A9 and D12 and their position within the BRCA1 locus. Additionally we show the two PCRs, which were performed to prove the sequence continuation from subclone D12 into A9 **(A)** and to determine the location of the two clones within the BRCA1 pseudogene **(B)**.
H III stands for *Hind* III and **X I** for *Xho* I restriction site. NBR1 pseu. Stands for NBR1 pseudogene.

I defined the lenght of A9 with 2186 bp. Sequencing was performed by primer walking. I could determine two stretches (511 and ~800 bp) with an up to 88% similarity to HERV-E.PTN covering the beginning and the end of the subclone. While the interruption in the sequence is only ~300 nt in the HERV-E.PTN element it is almost 800 nt in subclone A9. Furthermore I could define two regions (~1100 and ~500 nt) with an up to 92% similarity to a retrovirus-like element (accession # D10450) describing the defective env gene region. Additionally we could detect, that the first ~700 nt are 88% homologous to the env region of HERV clone 8.4, which is a member of the HERV-E family as well.

Based on the sequence comparison and the similarity to the HERV-E.PTN element in both subclones, I assumed that D12 was located upstream of A9 and might be connected to it at the *Hind* III site.

This was proven by performing a PCR with a sense primer derived from the 5' start of A9 (A9/D12) and with an antisense primer derived from the 3' end of D12 (D12/A9). DNA of PAC1993 clone served as the template. Considering the distance of both primers from the *Hind* III site, the band was 224 nt long, as expected. To define the localization of the connected subclones A9 and D12 in the predicted HERV element of the BRCA1 pseudogene, a long template PCR was performed. I chose the primer D12-364R, located in subclone D12, as a sense primer and primer B8-LTR, in the 5'LTR of the HERV element as an antisense primer. The result was a ~4.1 kb band (figure 7 and figure 8).

Based on this result I concluded that the 5' start of subclone D12 is located approximately 4 kb downstream of the retroviral elements 5' LTR.

To elucidate the entire structure of the assumed retroviral element in the BRCA1 pseudogene, I obtained two additional subclone libraries from PAC1993 clone, generated after *EcoR* I and *Xho* I digestion.

5.1.2 *EcoR* I and *Xho* I Subclone Libraries from the PAC1993 clone

After a colony lifting, the new libraries were screened with the oligonucleotide B8-LTR. Five positive subclones were detected in the *Xho* I subclone library, picked and amplified by Qiagen Maxi plasmid preparation. The DNA was digested with *Xho* I, loaded on a 1% agarose gel and transferred via Southern blotting (figure 9).

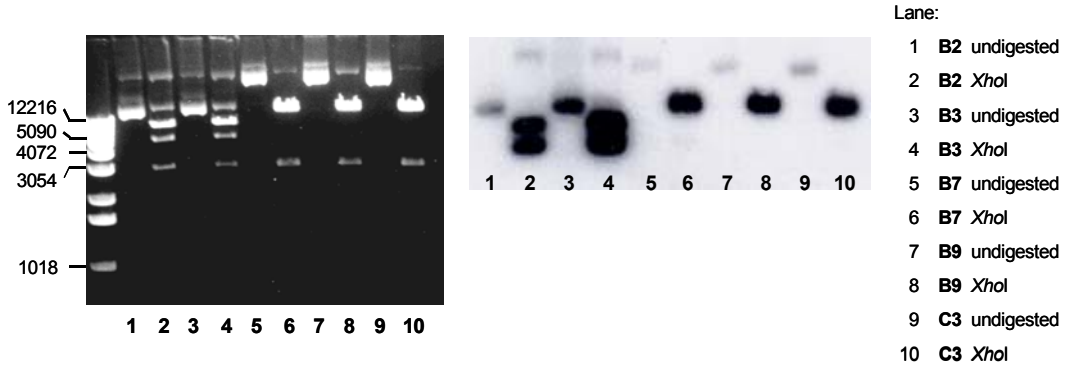

Figure 9:

Gel picture: Five positive subclones derived from the *Xho* I subclone library and digested with *Xho* I. The products were loaded on a 1% agarose gel.
The 2.8 kb band showing in all digested lanes, is the pZErO™ vector backbone. Subclones B2 and B3 (lanes 1-4) seem to be similar. The upper band in lanes 2 and 4 is the undigested vector. Both clones seem to have a triple integrant inserted, a ~5 kb, a 12 kb and a 15 kb long *Xho*I fragment. B7, B9 and C3 are similar as well. In Lanes 5-10, the upper band is representing the undigested vector. The single insert measures ~13 kb.
Southern blot of the same agarose gel, with oligonucleotide B8-LTR (specific for the 5' LTR of the retroviral element) serving as the probe. All signals represent the bands, which are shown in the agarose gel. For the subclones B2 and B3, we receive a signal in both inserts as well as in the undigested vector. The signal in subclones B7, B9 and C3 is strong in the insert, but very weak in the undigested vector. This might represent a less effective transfer of large DNA.

Subclones B2 and B3 as well as B7, B9 and C3 seem to be identical. This could be proven by sequencing them with the insert flanking primers SP6 and T7. I commenced sequencing via primer walking with the subclones B3 and B9.

B3 is approximately ~17.7 kb long. The T7 reading of B3 revealed that its 5' start is located in exon 1b in the BRCA1 pseudogene. A sense sequencing primer located ~1.2 kb downstream of the 5' LTR of the HERV element confirmed this fact by revealing BRCA1 pseudogene sequence (HSU77841 GenBank) downstream of exon 2.

Further sequencing by primer walking revealed high similarity (up to 89%) to HERV-E.PTN starting 124 nt downstream of the start of the 5' LTR. This similarity continues for 1.2 kb before I detected a gap of 351 bp followed by another 1.1 kb long stretch of high homology. The similarity to the gag coding region in

HERV-E.PTN starts approximately at position 1050 of the HERV element. The $tRNA^{glu}$ primer binding site, which is specific for the HERV-E family, could be detected at position 458 of the HERV-like element located 7 nt downstream of the end of the 5'LTR. Two point mutations have taken place (s. figure 15).

I could also validate a 86% homology to the Human Hereditary Hemochromatosis locus (pos. 64581 – 63849) ranging from position ~2300 to 2950 in the HERV element. Furthermore, sequencing showed a 88% similarity to a RTVL-Ib element from position 2836 to position 2950.

At position 2950 downstream of the 5' LTR, I identified an *Xho*I restriction site in the insert of the *Xho*I subclone B3. Downstream of this point, sequence comparisons to GenBank entrez did not reveal any similarities to HERV-like elements. Additionally, *Xho*I digestion of B3 showed also a triple band pattern. This fact combined with the length of the subclone (~17.7 kb) leads me to the conclusion, that B3 contains multiple fragments. The double integrant must contain a second LTR sequence, hence the positive signal with the LTR derived probe.

Subclone B9 is ~15 kb long. By sequencing with flanking primer SP6 I detected a 89% homology to position 63849 in the Human Hereditary Hemochromatosis region (accession # U91328). Homology to the same position was found the 5 kb *Xho*I fragment of subclone B3.

To prove whether the subclones B3 and B9 might be connected at the described *Xho*I restriction site, I performed a PCR with PAC1993 DNA as the template. Two primers were defined covering this region. Primer 4kb-3F (antisense) is located in subclone B3 702 nt upstream of the *Xho*I site and primer B9-SP6R (sense) is located in subclone B9 486 nt downstream of the 5' start. The product had the expected length of 1188 nt considering the distance of both primers to the *Xho*I site (Shown in figure 10 and 11).

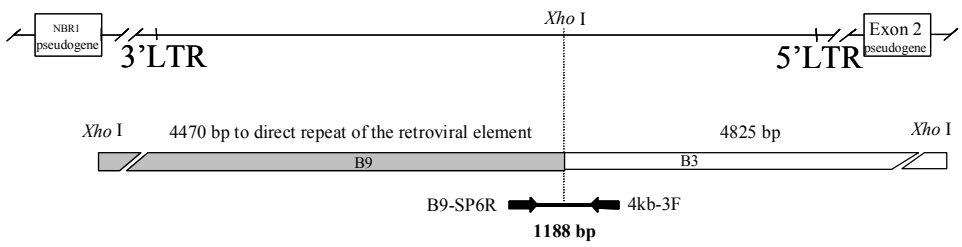

Figure 10:
Location of the subclones B3 and B9 within the BRCA1 pseudogene. B3 starts in Exon 1b of the BRCA1 pseudogene A PCR with primer pair 4kb-3F and B9-SP6R with PAC1993 DNA as template, proves the sequence continuation between the two subclones. The PCR product is 1188 bp long.

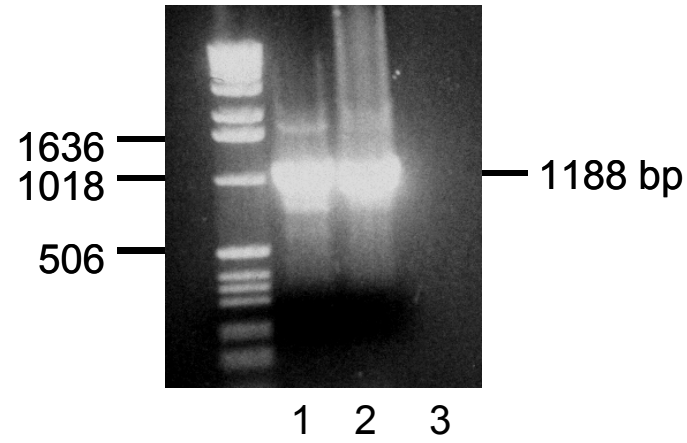

Figure 11:
PCR with primers 4kb-3F and B9-SP6R covering the region up-, and downstream of the *Xho*I restriction site, thus defining the continuation of the HERV sequence form B3 to B9. Pac1993 DNA served as template. We used 30 ng template in lane 1 and 90 ng in lane 2. Lane 3 is a water control. The products were loaded on a 1% Agarose gel and show the expected length of 1188 nt.

At position 3502 of the HERV-like element I detected an up to 91% homology to HERV-E.PTN (pos. 4233). Interestingly this sequence is the continuation of the HERV-E.PTN homology in the B3 subclone ending at position 2729 of the HERV element. This leads to the conclusion that the HERV element examined in my study, has an 771 nt long part, which is missing in the HERV-E.PTN element.

Furthermore could I detect an up to 87% homology over 376 nt to an RTVL-Ib element ranging from the 5' start of subclone B9 to position 3326 of the HERV element. Additionally, I found two short sequences up to 86% similar to RTVL-Ia element ranging from position 3562 to 3697 (135 nt) and from 3865 to 4093 (228 nt) of the HERV element.

Continuing the downstream sequencing of subclone B9, I detected a *Hind*III restriction site 1252 bp downstream of the *Xho*I site. The following sequence reading is identical with the 5' start of subclone D12. This finding was expected, since I had already proven the localization of D12 in the BRCA1 pseudogene region.

By deriving antisense primers from the 3' end of subclone A9, I could continue our sequence analysis. We could detect the start of the 3' LTR 142 nt downstream of the *Hind*III site at position 6990 of the HERV element. The 3'LTR is 465 nt long. Sequence comparisons in this region of B9 revealed high similarity to a retrovirus-like element describing the defective env coding region ending at position 7034.

By sequencing the downstream sequence of the 3'end, subclone B9 contains sequence up to 87% similar to BRCA1 gene sequence located downstream of exon 2. This leads to the conclusion that the duplication of the BRCA1 gene and therefore the BRCA1 pseudogene ranges at least to position 11821 of the BRCA1 gene.
Additionally, I could define a HERV-I like solitary LTR located ~4 kb downstream of the 3'LTR of the HERV element, by further sequencing of subclone B9.

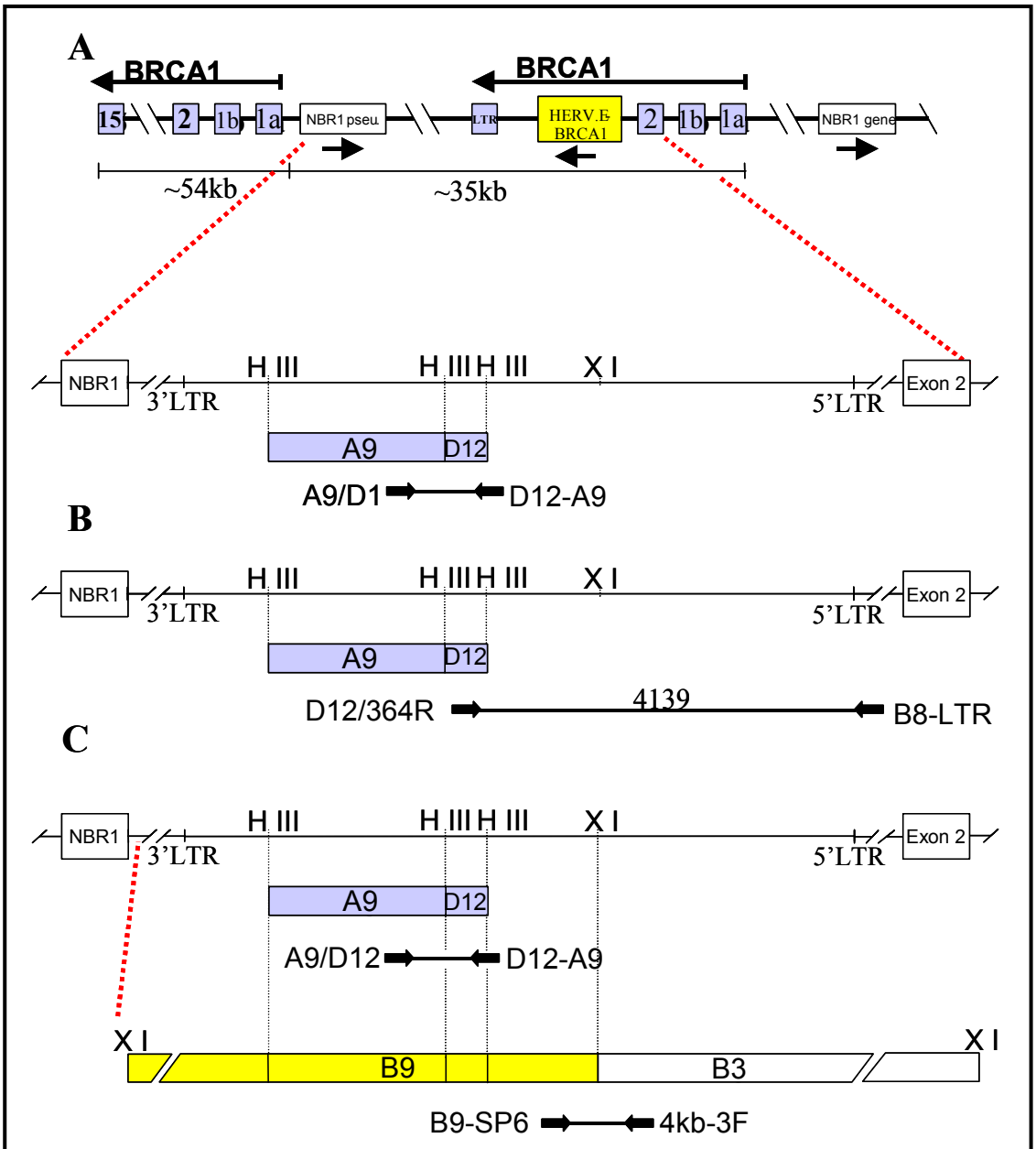

Figure 12:
This figure is not representing actual scale for easier visualization of the BRCA1 locus.
This figure describes the different steps which lead to the sequence and structure of the retroviral element located in the BRCA1 pseudogene. **A)** PCR to prove the continuation of subclone D12 to A9. **B)** The 4.1 kb long product of this long range PCR with PAC1993 DNA as template revealed the position of the subclones A9 and D12 in the BRCA1 pseudogene locus. **C)** After the detection of the subclones B9 and B3, the A9 to D12 PCR with B9 subclone DNA, proved the fact, that B9 contained the sequence of D12, A9 and sequence up-, and downstream of this position. The second PCR with the primers B9-P6R and 4kb-F covered the *Xho*I site in B3 and revealed the sequence continuation in B9.
X I stands for *Xho*I and **H III** for *Hind* III restriction site.

5.1.3 Characterization of HERV-E. BRCA1

The viral length, including the long terminal repeats (LTR), is 7455 bp. The long terminal repeats are 451 and 464 nt long. Both long terminal repeats are flanked by two identical, 18 nt long segments (Direct Repeats) representing the insertion sites into human DNA. The $tRNA^{glu}$ primer binding site is located 8 nt downstream of the end of the 5'LTR. It shows two point mutations at position 13 (C to T) and 16 (C to G) compared to the HERV-E.PTN element (s. figure 15). The LTR shows the typical retroviral organization, with U3-R-U5. The U3 region contains two transcription start sites - CCAAT at position 275 and the TATAA-Box at position 323. The polyadenylation signal (pos. 398) carrying R region starts at position 354, followed by U5 beginning at position 424. The U5 region contains a TTGT termination signal at position 438 (s. figure 13).

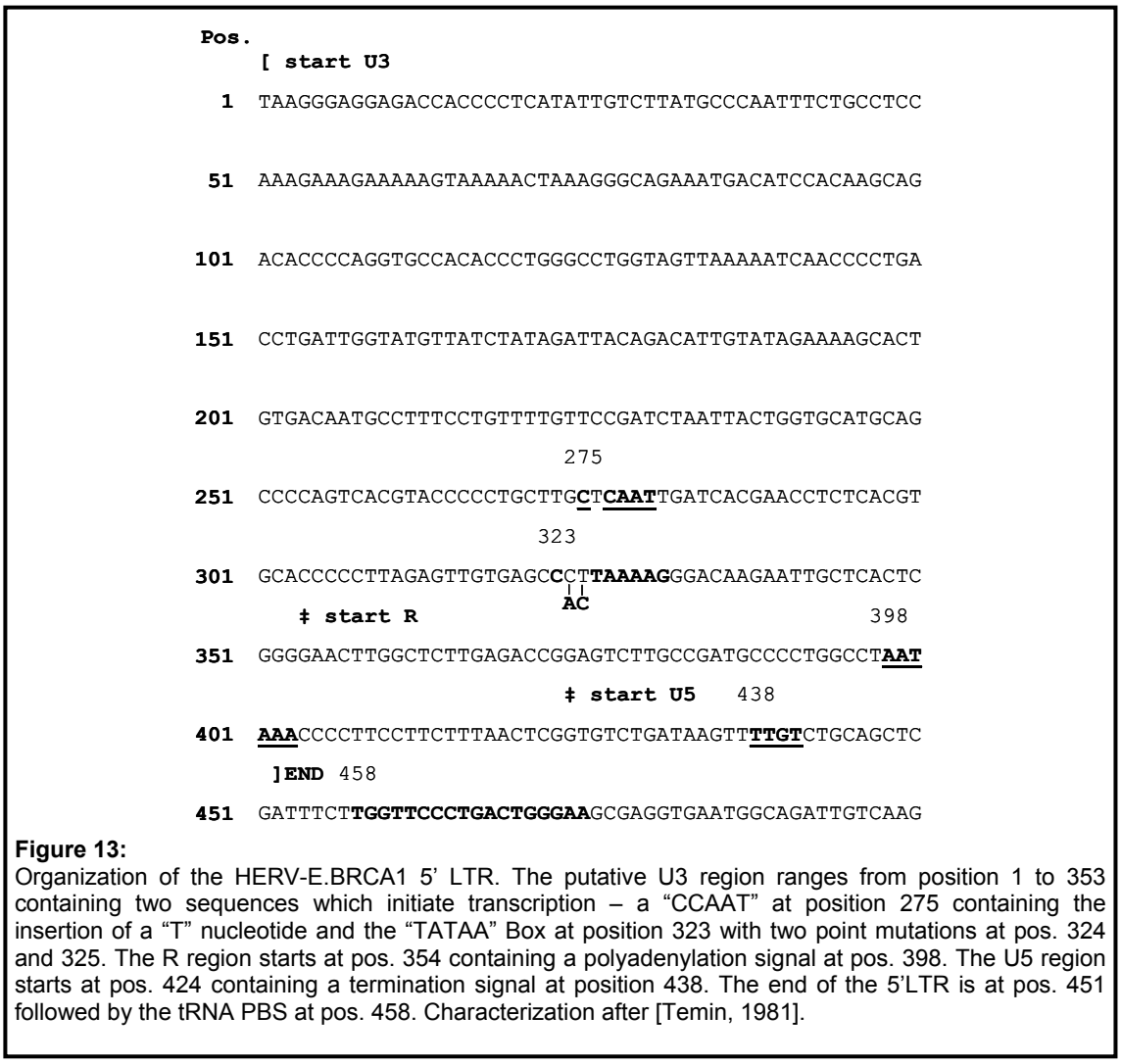

Figure 13:
Organization of the HERV-E.BRCA1 5' LTR. The putative U3 region ranges from position 1 to 353 containing two sequences which initiate transcription – a "CCAAT" at position 275 containing the insertion of a "T" nucleotide and the "TATAA" Box at position 323 with two point mutations at pos. 324 and 325. The R region starts at pos. 354 containing a polyadenylation signal at pos. 398. The U5 region starts at pos. 424 containing a termination signal at position 438. The end of the 5'LTR is at pos. 451 followed by the tRNA PBS at pos. 458. Characterization after [Temin, 1981].

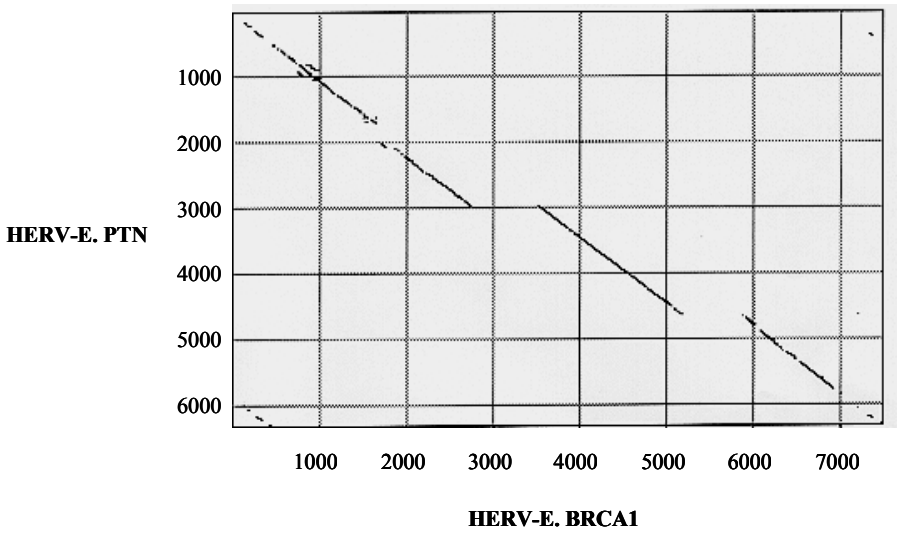

Figure 14:
This dot blot matrix shows the homology between HERV-E.BRCA1 and HERV-E.PTN. The minimal similarity in a window of 30 nt is 80%. The fact that HERV-E.BRCA1 is longer than HERV-E.PTN and the two gaps (~700 and 800 nt) in the sequence homology lead to the conclusion that HERV-E.BRCA1 is older than the latter.

The high homology to the HERV-E.PTN element (s. figure 14), to other HERV-E. members and the family specific $tRNA^{glu}$ binding site, define the retroviral element located in the BRCA1 pseudogene as a member of the HERV-E family. It will be referred to as HERV-E.BRCA1.

High similarities (up to 88%) to the three retroviral defective coding genes of HERV-E elements could be found (s. figure 15).

- **gag:** position 622 to 1350 in the HERV element
- **pol:** position 4313 to 4666 in the HERV element
- **env:** position 4666 to 7000 in the HERV element

Additionally, I could detect an insertion up to 86% homologous to the pol and env regions of a retrovirus of the RTVL-I family. From position 2836 to 3326 of the retroviral element located in the BRCA1 pseudogene, I detect an up to 86% similarity to RTVL-Ib and from position 3562 to RTVL-Ia. This suggests that this retroviral element is unlike HERV-E.PTN, a recombinant element (s. figure 15).

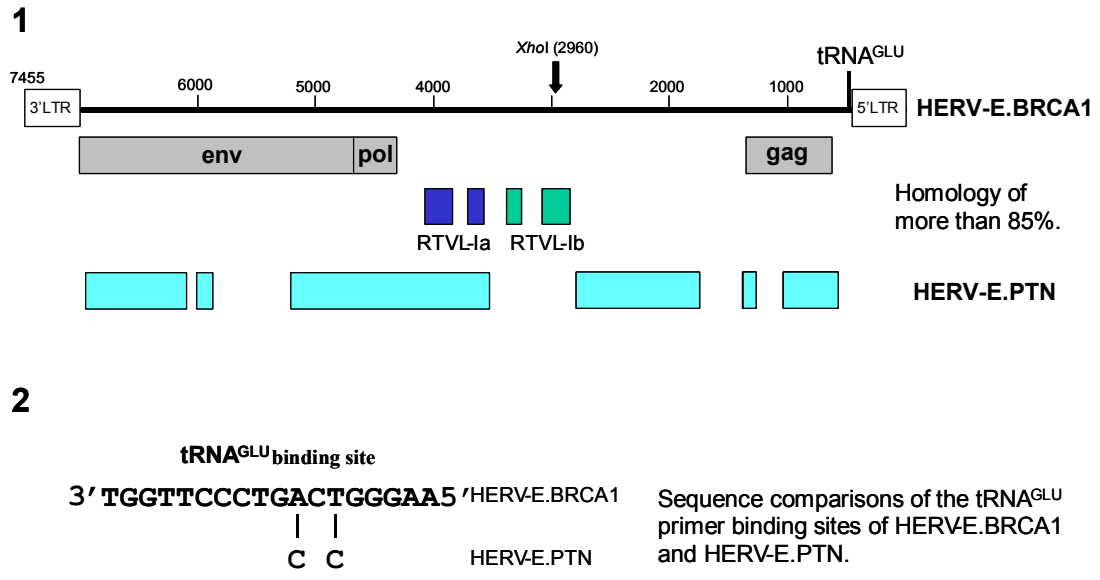

Figure 15:
1: Homologies between HERV-E.BRCA1 and HERV-E.PTN showing the similarities to the defective coding genes and to the RTVL-Ia env gene and RTVL-Ib pol gene viral element as well.
2: The sequence comparison of the $tRNA^{GLU}$ [Skiya et al., 1981] binding site between HERV-BRCA1 and HERV-E.PTN.

5.2 Phylogenetic Analysis of HERV.E-BRCA1

To define the evolutional time of insertion of HERV-E.BRCA1 element into the BRCA1 locus, I performed various PCRs from genomic DNA. To cover several time frames in evolution, genomic DNA of human, apes and Old World monkeys was used and the PCR fragments were compared to human DNA.

Three sets of primers were chosen, covering gene-specific, pseudogene-specific and HERV-specific regions in the BRCA1 locus. All oligonucleotides were compared to GenBank entries to ensure their specific binding abilities.

1. **Gene-specific:** B3 / B4 covering exon 2 of the BRCA1 gene from position 2956 to 2579 resulting in a 377 bp fragment.
2. **Pseudogene-specific:** B11F / B12R covering exon 2 and sequence from the upstream intron in the BRCA1 pseudogene ranging from position 2187 to 2479 resulting in a 292 bp fragment.
3. **HERV-specific:** B11F / BR1.3 covering exon 2 and 548 bp retroviral-like sequence (5' LTR) resulting in a 1153 bp fragment.
4. **HERV-specific in Rhesus monkey:** Since the PCR with primer combination 3 failed, a different primer pair was chosen. B1 / BR1.3 covering intronic sequence upstream of exon 2 and 548 bp retroviral-like sequence resulting in an 802 bp long fragment.(shown in figure 16)

All PCR products were loaded on a 1% agarose gel for gelelectrophoretic analysis and were transferred to a nylon membrane by Southern blotting. Three oligonucleotides were defined for screening the Southern blots. The BRCA1 gene-, and pseudogene-specific oligonucleotide (B11F) was derived from human sequence in exon 2. The second oligonucleotide (B2) is specific for the BRCA1 pseudogene as well as for the HERV-E.BRCA1 and cannot bind in exon 2 of the BRCA1 gene. The LTR probe B8-LTR was used for the HERV-specific Southern blot hybridization with Rhesus DNA.

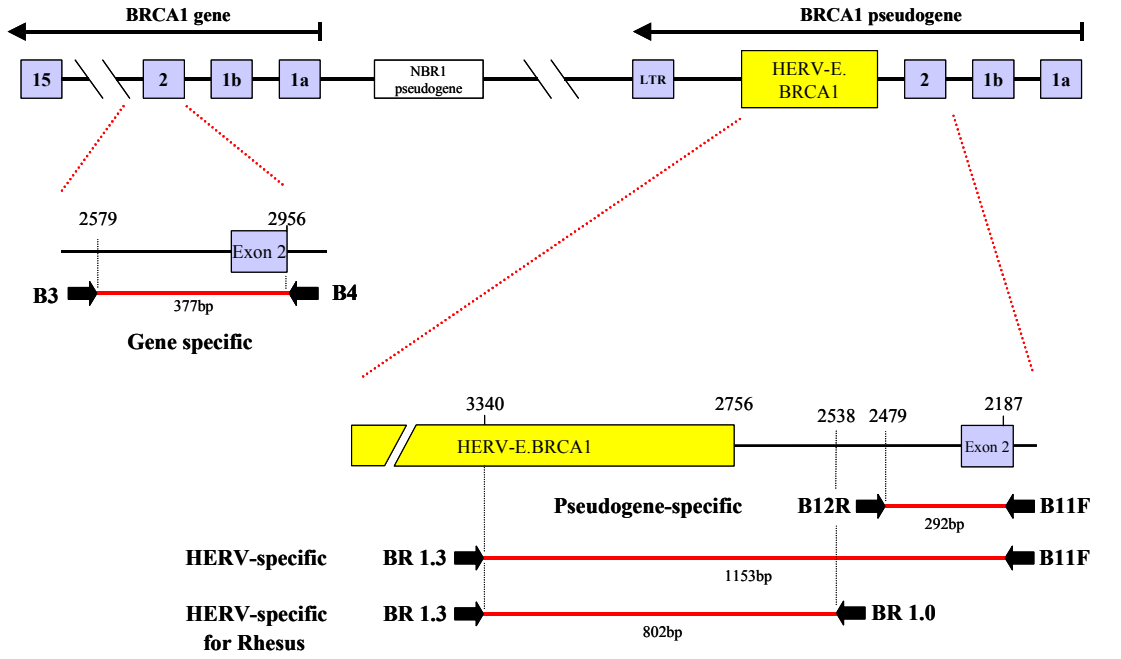

Figure 16:

The four PCRs were performed with Human, Chimpanzee, Gorilla and Rhesus DNA serving as template. The bottom PCR was only performed with the DNA of Rhesus monkey since the primer combination BR1.3 and B11F. The resulting products prove the presence of those specific sequences in the named species. Positions indicated in this figure represent the primer starting positions in the pseudogene sequence (GenBank # U77841) and in the BRCA1 gene locus (GenBank # U37574). Position 2756 in the BRCA1 pseudogene represents the start of the retroviral element.

The gene specific lower band (377 nt) shows up in all 4 DNA samples. The upper band showing up in the gene specific Southern blot (s. figure 17 A), is 662 nt long. Although primer B4 has two mismatches in the BRCA1 pseudogene, it apparently primed well enough to result in this pseudogene-specific product. It is missing in chimpanzee. This might be due to the in general low product amounts in the PCR performed with chimpanzee DNA or with a low amount of DNA on the gel.

The primer combination B12R and B11F, specific for the BRCA1 pseudogene, leads to the expected 292 bp long product in all 4 samples.

The first HERV-specific PCR results in the primer specific 1153 bp long product. Additionally, we see two bands of ~1000 (human) and 850 nt lengths (gorilla). These bands were as well visible on the agarose gel. I will further explain this fact in the discussion part of this thesis.

Since, I could not receive a product with the first HERV-specific PCR in Rhesus DNA, it was decided to perform a second PCR with the primer combination BR1.3 and BR1.0. I received a single 802 nt long band on the agarose gel. The Southern blot shows us a second band in human DNA with ~700 bp length.

Considering the evolutionary tree, the insertion of the BRCA1 pseudogene as well as of the HERV-E.BRCA1 element has occurred at least 25 million years ago. It is likely that the insertion of the HERV-E.BRCA1 into the BRCA1 locus occurred after the generation of the BRCA1 pseudogene, because there is no retrovirus-like sequence present in the BRCA1 gene (s. figure 16).

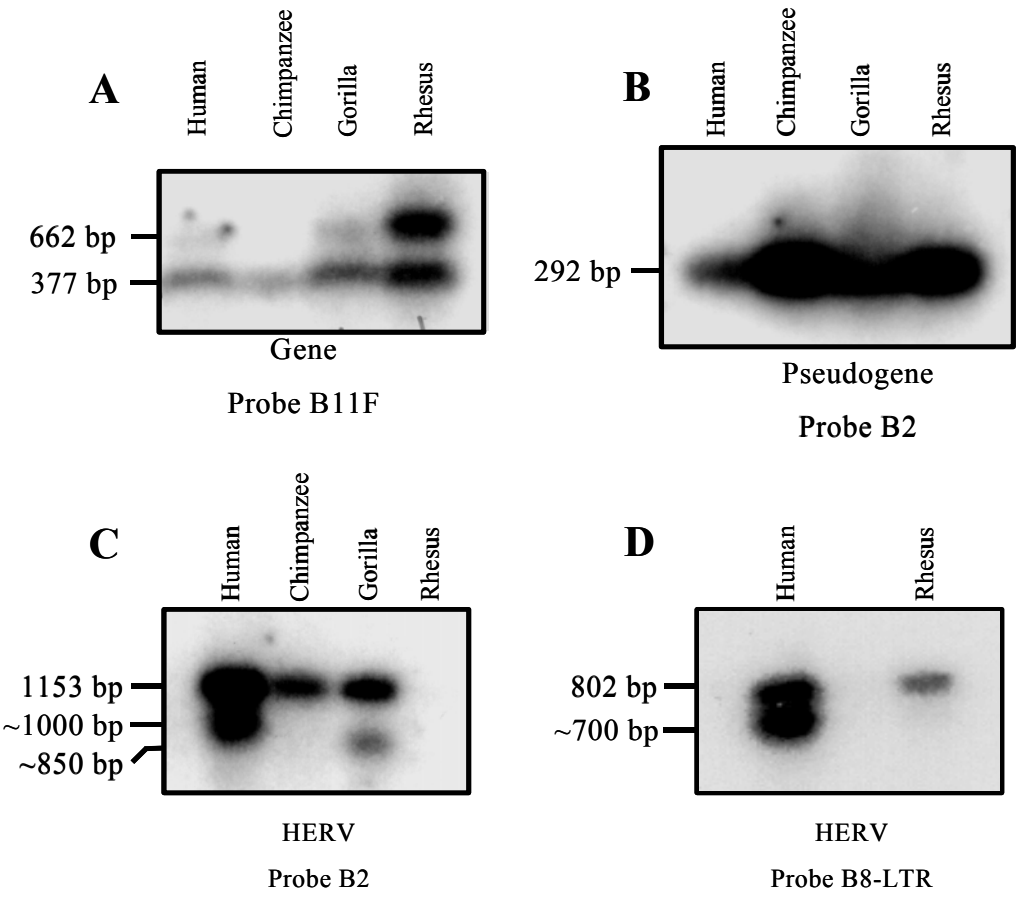

Figure 17:
The four Southern blots derived from a 1% Agarose gel. All four genomic PCR products are shown. **A** The gene specific PCR reveals two bands of which the lower (377 bp) one, is he specific gene product and the upper one (662 bp) is the specific product in the duplicated pseudogene. The probe we used, was derived from exon 2 in the BRCA1 gene. **B** The PCR product specific for the pseudogene is 292 nt long. The probe B2 is located in exon 2 and specific for the pseudogene. **C** The HERV element specific PCR exposes the expected 1153 nt long band. This product does not appear with Rhesus DNA as template. Additionally we can see several bands of unknown origin. The second band in human DNA is ~1000 nt and the lower band in gorilla, about 850 nt long.
D The second HERV specific PCR was performed with rhesus and human DNA. The expected product is 802 bp long. The lower ~700 bp long band in the first lane, is discussed later. Probe B8-LTR was derived from the 5'LTR in the BRCA1 pseudogene.

5.3 The Transcriptional Activity of the HERV-E.BRCA1 5'LTR and its Possible Influence on the Expression of the BRCA1 Gene and Pseudogene

5.3.1 Analysis of the Transcriptional Activity of the HERV-E. BRCA1 5'LTR

I hypothesized that HERV-E.BRCA1, referring to its high similarity to the transcriptional active HERV-E.PTN element, might have transcriptional activity. Based on the knowledge that the 5' LTR of retroviral elements usually contains sequences which ensure the synthesis of progeny viral RNA and mRNA, I decided to amplify this region via PCR. The fragment derived from the BRCA1 pseudogene (pos. +2538 to 3331 GenBank #U37574) contains 450 nt of the 5' LTR. It was amplified by PCR using the sense primer BR1.0 and the antisense primer BR1.3. The fragment was then cloned into a TA cloning vector for amplification and later isolated using the restriction endonuclease *BamH* I.

We chose a promoter-less dual reporter vector containing the luciferase gene (pXP-I) to test for transcriptional activity. The vector was digested with *BamH* I.

Since retroviral LTRs usually contain elements which act as orientation independent enhancer, the fragment was inserted into the vector facing both orientations, creating the vectors pXP/HERV3- and pXP/HERV2+ (in which plus stands for positive orientation of the HERV fragment) as shown in figure 18.

The two constructs plus the empty vector as control, were transiently transfected into several human cancer cell lines, all of which express BRCA1.

- JEG-3: choriocarcinoma cell line
- JAR: choriocarcinoma cell line
- BeWo: choriocarcinoma cell line
- HS578T: sporadic breast cancer cell line

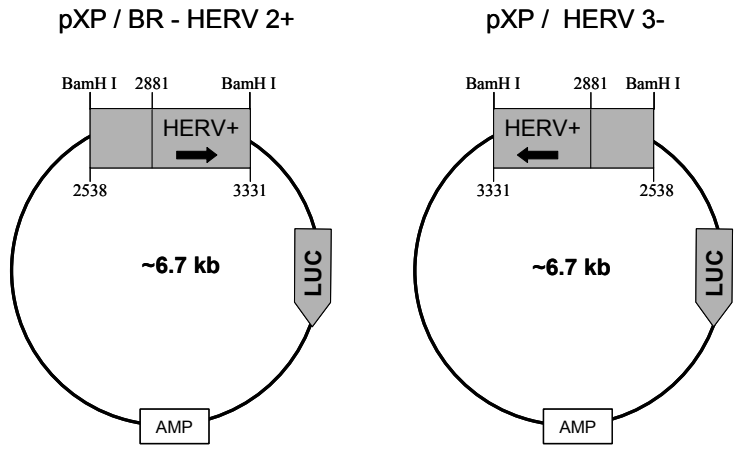

Figure 18:
The two clones containing a 793 nt long fragment retrieved from the BRCA1 pseudogene region. The positions indicated in the figure, represent the positions in the BRCA1 pseudogene sequence (U77841). 450 nt of this fragment represent the retroviral 5' LTR. The fragment was cloned into the *BamH*I restriction site in the multiple cloning site of a luciferase reporter vector. The distance to the luciferase gene (LUC) in the vector is 34 nt.

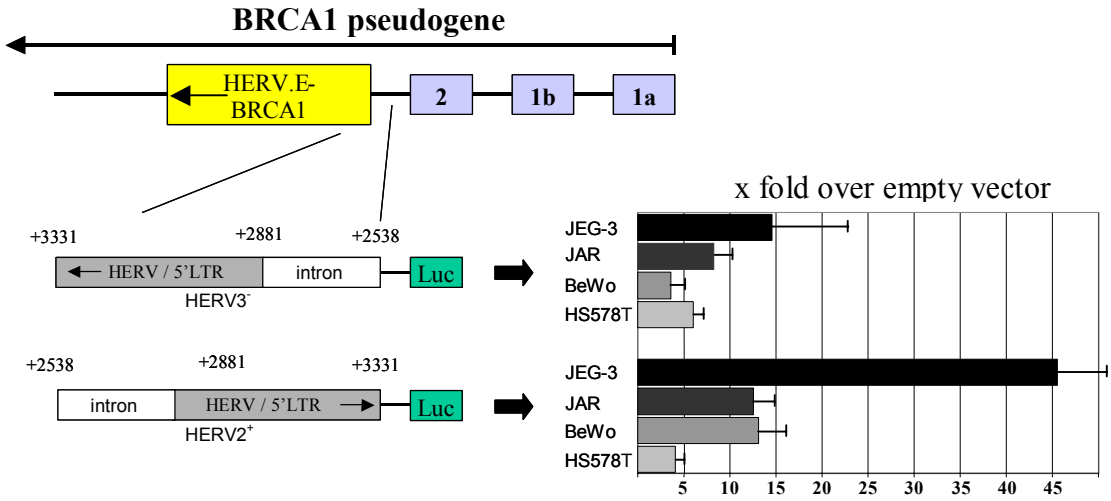

Figure 19:
Result of the transient transfection luciferase assay. The cloned fragment was 793 nt long and consisted of 405 nt retroviral LTR (pos. 2881 to 3331). The rest is intronic sequence. The fragment was cloned in both orientations into the multiple cloning site adjacent to the luciferase reporter gene. Activity is described as fold over empty vector.

All experiments were performed in triplicates.

The handling of the human cell lines as well as the luminescent assay was performed as described in chapter 4.16.

Transcriptional activity could be detected in all four cell lines (s. figure 19) in both orientations. Especially in JEG-3 cells, the HERV fragment shows the highest activity (46 fold) and thus the 5'LTR of HERV-E.BRCA1 could function as a

bi-directional promoter and therefore be theoretically able to initiate expression of BRCA1 pseudogene antisense transcript in addition to retroviral transcripts.
However, whether BRCA1 pseudogene antisense transcripts exist and whether they would be able to interfere with BRCA1 transcripts was not further evaluated in this thesis.

6 Discussion

In this thesis, I report the detection, structural analysis and evolutionary history of a human endogenous retrovirus (HERV) located in the human BRCA1 gene locus. Based on data received by GenBank comparisons with the recently published HERV-E.PTN [Schulte et al., 1998] sequence, we retrieved a region up to 85% homologous to the 5'LTR of HERV-E.PTN, located inside the BRCA1 pseudogene. This suggested the presence of a HERV-like element in this locus. By further sequence analysis, we detected a 7455 bp long retroviral element flanked by LTRs of 451 and 464 nt lengths. Based on its $tRNA^{GLU}$ primer binding site and sequence comparisons resulting in high similarity to other members of the HERV-E. family, we classified it as HERV-E.BRCA1. We could define three regions similar to the defective retroviral coding genes *gag*, *pol* and *env*. Additionally we could find a 1021 nt long stretch highly similar to members of the RTVL-I family. We conclude that HERV-E.BRCA1 is a noninfective, defective, retroviral element which has high homologies to other members of the HERV-E. family such as HERV-E. 4.1, HERV clone 8.4 and HERV-E.PTN. Additionally we could define regions in HERV-E.BRCA1 similar to up to 88% to the human hereditary hemochromatosis (hHH) region ranging over 650 nt and to the human X chromosome. The similarity to members of the HERV-E. family and the inserted stretch highly similar to RTVL-I elements, leads to the conclusion that HERV-E.BRCA1 is a recombinant retroviral element.

The HERV-E family belongs to the HERV-ERI superfamily. The first clones to be isolated were 51.1, 4.1 and 4.14. They were isolated by hybridizing the human genomic library with a probe derived from an African green monkey clone. This clone was found by cross-hybridizing the monkey DNA with a probe derived from the Murine Leukemia Virus (MLV). The detected clone was found to be highly similar to the *gag-pol* region of the baboon endogenous virus (BaEV) [Martin et al., 1981]. A thorough analysis of clone 4.1 revealed a full length HERV containing *gag*, *pol* and *env* and being flanked by two LTRs [Repaske et al., 1985]. The human genome contains supposedly about 50 full length and truncated copies of this HERV-element [Steele et al., 1984]. Members of the HERV-E. family have been detected in the genomes of Old World monkeys, apes and humans and more detailed studies of

clone 4.1 revealed, that the location in which it could be found was identical in all species [Shih et al., 1991]. This lead to the conclusion, that insertion of this retroviral element took place at a time point of evolution which dates back more than 25 mio years ago.

The RTVL-I family was detected by chance during sequence analysis of the human haptoglobin gene cluster. The prototype is flanked by two 500 bp long LTRs and contains sequences related to *gag*, *pol* and *env*. Based on its primer binding site's homology to the mouse tRNA for isoleucine, it was named RTVL-I [Maeda et al., 1985]. Three different retroviral elements were found in the haptoglobin gene cluster and were named RTVL-Ia, RTVL-Ib and RTVL-Ic of which humans lack of RTVL-Ib due to a DNA deletion which resulted from a crossover event [Maeda et al., 1990].

The organization of HERV-E.BRCA1 does not resemble normal retroviral organization. The order of sequence homologies to the defective retroviral coding genes is not *gag-pol-env*, instead there are two additional copies of coding genes inserted between *gag* and *pol*. This detection of sequence homologies to the *pol* region of RTVL-Ib and to the *env* region of RTVL-Ia in the middle section of HERV-E.BRCA1 (s. figure 12), are consistent with the description of reoccurring recombination events in the haptoglobin gene cluster. The *gag* region of RTVL-I is supposedly present in 8 copies inside the human genome, while the *pol*-related region is present in up to 25 copies. Additionally up to six tandemly repeated copies of the whole gene cluster have been described [Maeda et al., 1986]. Interestingly, by comparing HERV-E.BRCA1 to HERV-E.PTN, we loose sequence similarity from position 2729 to ~3502 exactly were a high homology to RTVL-Ib can be detected. There is a second gap between positions ~5200 to ~5900. This leads to the conclusion that HERV-E.BRCA1 is older than HERV-E.PTN but is as well the product of a retroviral recombination (compare figures 12 + 13). About the question whether the RTVL-I elements were already present in the BRCA1 pseudogene before they recombined with a HERV-E element or if recombination took place prior to insertion en bloc, one can only hypothesize. Fact is, that RTVL-I elements were already found in New World monkeys (~45 mio years ago) and recently it was reported that type I retroviruses were even found in reptiles [Martin et al. 1997]. As our group recently reported, it is unlikely that insertion of a HERV-E retroviral

element and insertion of a member of the RTVL-I family took place independently and parallel in the PTN gene locus, since there are HERV-E.PTN similar sequences (> 80%) on distant gene loci, the hHH region and on human chromosome X. Due to the striking similarity between HERV-E.PTN and HERV-E.BRCA1 (~ 90%), this event would be rather unlikely for the BRCA1 gene locus as well.

The HERV-E.BRCA1 insertion site is 17 nt long. This is rather long for a retrovirus (mostly 4-6 nt), but those repetitive sequences are not described to be similar and vary from virus to virus [Chen et al., 1984]. It appears that the viral integration site, may not be sequence specific and retroviral sequences are inserted into the genome at many different locations [Fitts et al., 1983]. The long terminal repeats are 451 nt, respectively 464 nt long. Defining the typical U3, R and U5 regions is rather difficult, because several mutations have taken place during evolution. Temin and his group described the typical structure of retroviral LTRs [Temin et al., 1981]. In HERV-E.BRCA1 all these typical LTR sequences can be found. The long terminal repeat starts with U3, a region in which supposedly promoter sequences regulate the expression of proviral DNA. A transcription initiation sequence (CCAAT) is located at position 275 and the TATAA box at position 323 of HERV-E.BRCA1. Both are located inside U3. R starts at position 354 and carries a polyadenylation signal at position 398. U5 starts at position 424 and carries a transcription termination (TTGT) signal at position 438. The 5'LTR ends at position 451 followed by the primer binding site at position 458 (s. figure 14). This LTR region has relatively low homology to HERV-E.PTN (~84%) suggesting that frequent mutations have taken place during evolution. Nevertheless most mammalian retroviruses differ in the length of their U3 region, while the lengths of R and U5 are rather similar (60-80 nt each). Interestingly in the case of HERV-E.BRCA1, R is 70 nt while U5 is only 27 nt long. In the contrast to other mammalian retroviruses, a deletion must have occurred, decreasing the length of U5.

To define the point of insertion of this retroviral element into the genome, we performed PCRs specific for retroviral sequence, the BRCA1 gene and the BRCA1 pseudogene using DNA of human, chimpanzee, gorilla and rhesus as template. Following the PCRs, the gel was transferred to a membrane and we performed a Southern Blot analysis, with sequence specific oligonucleotides (s. figure 17). Chimpanzees and gorillas belong to the apes which are the closest relatives to

humans. The gorilla is approximately 25 mio years old. The rhesus monkey dates back farther in evolution and belongs to the Old World monkeys ~25 – 45 mio years old. We know that HERV-E.PTN inserted into the genome about 25 mio years ago, because it could not be detected in the rhesus monkey. Since we expect HERV-E.BRCA1 to be older than HERV-E.PTN an insertion into the genome at the stage of the rhesus monkey was likely. The BRCA1 gene could be detected in all four species, which was obvious since the BRCA1 gene locus is already present in mice. The second band with a length of 662 nt is specific for the BRCA1 pseudogene. Although the gene specific primer pair B3 / B4 has two mismatches each in the BRCA1 pseudogene, it seems to prime well enough for PCR amplification. The oligo used for screening was B11F, which is specific for pseudogene and has one mismatch at the 3'end in the BRCA1 gene locus. The band does not appear in chimpanzee, which seems to be due to low product amount in the performed PCR.

It was difficult to detect primers solely specific for the BRCA1 pseudogene, since the duplicated region around exon 2 is highly similar to the BRCA1 gene region. We chose the primer B11F and B12R derived from the BRCA1 pseudogene sequence (accession No. U77841). Primer B11F has one mismatch in the BRCA1 gene compared to the BRCA1 pseudogene. An adenine base at the 3'end mutated to guanine in the pseudogene, making a positive PCR reaction from the BRCA1 gene region highly unlikely. The probe B2 derived from the BRCA1 pseudogene and used for hybridization of the 292 nt long PCR product, contains two point mutations compared to the BRCA1 gene. Conditions for the Southern blot were set to stringent conditions. The positive signal appears in all four species (s. figure 17), setting the time point of the insertion of the BRCA1 pseudogene to at least 25-45 mio years ago at the level of the New World monkeys. A BRCA1 pseudogene does not exist in the murine BRCA1 locus. To further examine the existence of the HERV-E.BRCA1 element, we chose a primer pair derived from exon 2 in the BRCA1 pseudogene and from the 5'LTR, located ~58 nt downstream from the direct repeat. We expected an 1153 nt long PCR product, specific for the BRCA1 pseudogene and the inserted HERV-E.BRCA1 as shown in our Southern analysis.

Additionally we received a ~1000 nt long band in human and a weak ~850 nt long band in gorilla. Since the chimpanzee is closer in evolution to the human race than

the gorilla and the additional ~850 nt long band does not appear in the chimpanzee and the human PCR, it could be specific for a retroviral element, which got mutated or deleted during evolution. Otherwise the PCR product would be the result of unspecific binding, which is rather unlikely. The double band appearing in human needs further explanation. Primer B11F was used as the primer derived from the pseudogene's exon 2. Furthermore it is not to be expected that this primer leads to a product inside the BRCA1 gene as sequence data showed. Additionally, the HERV-like sequence is not present in the BRCA1 gene locus. Sequence comparison with gene bank entries, clearly proved that the similarity between the BRCA1 gene and the pseudogene ends at position 3426 around 500 nt downstream from exon 2. Furthermore, GenBank comparisons revealed that 1153 nt long sequence is identical to the BRCA1 pseudogene. Its similarity to the BRCA1 gene stops downstream of exon 2 at the expected position, while sequence similarities to other entries start at the 5'LTR. There must be a second duplicated region of the BRCA1 gene locus in the human genome. This would be one explanation for the amplification of an additional 1 kb fragment with human DNA. The BRCA1 locus containing the HERV-.BRCA1 and exon 2 of the BRCA1 pseudogene could have inserted a second time somewhere within the human genome. This hypothesis is further strengthened by the second HERV specific PCR we performed with human and rhesus DNA.

Interestingly we get no PCR product in the rhesus monkey which first lead to the conclusion that the HERV-E.BRCA1 inserted into the genome after the duplication and re-insertion of the BRCA1 gene. We still decided to try a second PCR by exchanging primer B11F to primer BR1.0, which is situated 350 nt downstream in the BRCA1 pseudogene locus and identical to the equivalent position in the BRCA1 gene. This time we received the expected 802 nt long product in human and rhesus monkey DNA. A double band appears in human, strengthening the hypothesis of a second duplication of this gene locus in the human genome. The length difference between both products is ~100 nt as well.

The probe (B8/LTR), used to label the Southern blot, was derived from the 5'LTR and located 411 nt downstream of the direct repeat. GenBank comparisons with the probe sequence did not reveal any possible duplicated regions, affirming the specific binding in the applicated PCR product. We hereby proved the insertion of the HERV-E.BRCA1 into the genome at least at the evolutionary step of the rhesus and

thus the New World monkey about 25-45 mio years ago. The double band appearing in both human HERV specific PCRs lead to the hypothesis that a second duplication of this region might have taken place. Further experiments to prove this theory would be interesting to perform in the future.

We tried to elucidate possible physiological function of this HERV element. Since we know that retroviral elements are scattered throughout the genome and we could prove so far, that a complete recombinant retroviral element is inserted in the human BRCA1 gene locus, we focused on possible transcriptional activity of the HERV-E.BRCA1 5'LTR. Although, only a few examples of alterations in gene expression by a HERV have been reported, we assumed due to the high similarity to the transcriptional active HERV-E.PTN element and its influence on the expression of the human pleiotrophin gene, that the HERV-E.BRCA1 may be transcriptional active and could theoretically influence the regulation of the BRCA1 gene promoter.

Consequently we tested whether the 5'LTR of the HERV-E.BRCA1 has transcriptional activity. The fragment used for cloning contains the 5'LTR and thus the regulatory region of retroviral elements. Since it is known, that most retroviruses show orientation independent activity, the fragment was cloned in both orientations. Our work revealed that the 5'LTR of the HERV element shows transcriptional activity when tested in either orientation. In negative orientation, the activity ranged from 4 to 14 fold over empty vector and in positive orientation from 4 to 45 fold depending on the cell line. This finding reveals that the HERV-E.BRCA1 is theoretically able to initiate the expression of antisense transcripts from the BRCA1 pseudogene. These transcripts might interfere with BRCA1 transcripts or with the physiological regulation of the BRCA1 gene promoter. Overall, only 4 cell lines were tested in our study and we don't know if other BRCA1 expressing cell lines might lead to different results. This leaves room for further experiment, which would go beyond the scope of this thesis.

To come to a conclusion. In this thesis we defined a full length recombinant human endogenous retroviral element in the human BRCA1 gene locus named HERV-E.BRCA1. It is located in a duplicated region containing the first three exons of the BRCA1 gene named BRCA1 pseudogene and located 35 kb upstream from the BRCA1 gene promoter. We could elucidate its structure and sequence. It shows high homology (~90%) to the previously described HERV-E.PTN. Recombination

occurred with a member of the RTVL-I family. Our phylogenetic analysis proves the insertion of the human BRCA1 pseudogene into the genome, at least 25-45 mio years ago. The existence of the HERV-E.BRCA1 could be shown in the genome of New World monkeys, which proves an insertion at an evolutionary level 25-45 mio years ago. The appearance of an additional band in the HERV specific PCR in human and gorilla, suggests that possibly a second BRCA1 pseudogene duplication exist in the human genome.

The 5'LTR hosting the regulatory elements of a retrovirus, shows transcriptional activity in either orientation up to 45 fold over empty vector. Further studies are necessary to test for possible HERV-E.BRCA1 transcripts and the transcription of possible BRCA1 pseudogene antisense transcripts initiated by the HERV-E.BRCA1 5'LTR.

7 Bibliography

Begg, L., Kuller, L. H., Gutai, J. P., Caggiula, A. G., Wolmark, N., and Watson, C. G. (1987). Endogenous sex hormone levels and breast cancer risk, Genet Epidemiol *4*, 233-47.

Blackwood, M. A., and Weber, B. L. (1998). BRCA1 and BRCA2: from molecular genetics to clinical medicine [see comments], J Clin Oncol *16*, 1969-77.

Bowlin, S. J., Leske, M. C., Varma, A., Nasca, P., Weinstein, A., and Caplan, L. (1997). Breast cancer risk and alcohol consumption: results from a large case-control study, Int J Epidemiol *26*, 915-23.

Brown, M. A., Xu, C. F., Nicolai, H., Griffiths, B., Chambers, J. A., Black, D., and Solomon, E. (1996). The 5' end of the BRCA1 gene lies within a duplicated region of human chromosome 17q21, Oncogene *12*, 2507-13.

Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C., and Klevit, R. E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex, Nat Struct Biol *8*, 833-7.

Casey, G., Plummer, S., Hoeltge, G., Scanlon, D., Fasching, C., and Stanbridge, E. J. (1993). Functional evidence for a breast cancer growth suppressor gene on chromosome 17, Hum Mol Genet *2*, 1921-7.

Chen, Y., Chen, C. F., Riley, D. J., Allred, D. C., Chen, P. L., Von Hoff, D., Osborne, C. K., and Lee, W. H. (1995). Aberrant subcellular localization of BRCA1 in breast cancer, Science *270*, 789-91.

Chu, K. C., Tarone, R. E., Kessler, L. G., Ries, L. A., Hankey, B. F., Miller, B. A., and Edwards, B. K. (1996). Recent trends in U.S. breast cancer incidence, survival, and mortality rates [see comments], J Natl Cancer Inst *88*, 1571-9.

Claus, E. B., Risch, N. J., and Thompson, W. D. (1990). Age at onset as an indicator of familial risk of breast cancer [see comments], Am J Epidemiol *131*, 961-72.

Claus, E. B., Schildkraut, J. M., Thompson, W. D., and Risch, N. J. (1996). The genetic attributable risk of breast and ovarian cancer, Cancer *77*, 2318-24.

Consortium, B. C. L. (1997). Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium [see comments], Lancet *349*, 1505-10.

Easton, D. F. (1999). How many more breast cancer predisposition genes are there?, Breast Cancer Res *1*, 14-7.

Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I. C., Repasky, E. A.*, et al.* (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors, J Natl Cancer Inst *92*, 564-9.

Fitts, R., and Temin, H. M. (1983). Cellular DNA surrounding integration sites of an avian retrovirus, J Gen Virol *64 (Pt 2)*, 267-74.

Goss, P. E., and Sierra, S. (1998). Current perspectives on radiation-induced breast cancer, J Clin Oncol *16*, 338-47.

Gowen, L. C., Avrutskaya, A. V., Latour, A. M., Koller, B. H., and Leadon, S. A. (1998). BRCA1 required for transcription-coupled repair of oxidative DNA damage, Science *281*, 1009-12.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21, Science *250*, 1684-9.

Harkin, D. P., Bean, J. M., Miklos, D., Song, Y. H., Truong, V. B., Englert, C., Christians, F. C., Ellisen, L. W., Maheswaran, S., Oliner, J. D., and Haber, D. A. (1999). Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1, Cell *97*, 575-86.

Hopper, J. L. (2001). More breast cancer genes?, Breast Cancer Res *3*, 154-7.

Kauff, N. D., Satagopan, J. M., Robson, M. E., Scheuer, L., Hensley, M., Hudis, C. A., Ellis, N. A., Boyd, J., Borgen, P. I., Barakat, R. R.*, et al.* (2002). Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation, N Engl J Med *346*, 1609-15.

Kelsey, J. L., and Horn-Ross, P. L. (1993). Breast cancer: magnitude of the problem and descriptive epidemiology, Epidemiol Rev *15*, 7-16.

Koonin, E. V., Altschul, S. F., and Bork, P. (1996). BRCA1 protein products Functional motifs, Nat Genet *13*, 266-8.

Krainer, M., Silva-Arrieta, S., FitzGerald, M. G., Shimada, A., Ishioka, C., Kanamaru, R., MacDonald, D. J., Unsal, H., Finkelstein, D. M., Bowcock, A.*, et al.* (1997). Differential contributions of BRCA1 and BRCA2 to early-onset breast cancer [see comments], N Engl J Med *336*, 1416-21.

Larson, J. S., Tonkinson, J. L., and Lai, M. T. (1997). A BRCA1 mutant alters G2-M cell cycle control in human mammary epithelial cells, Cancer Res *57*, 3351-5.

Levy, J. A. (1996). The Retroviridae (New York, Plenum Press).

Li, H., Lee, T. H., and Avraham, H. (2002). A Novel Tricomplex of BRCA1, Nmi, and c-Myc Inhibits c-Myc-induced Human Telomerase Reverse Transcriptase Gene (hTERT) Promoter Activity in Breast Cancer, J Biol Chem *277*, 20965-73.

Lower, R., Boller, K., Hasenmaier, B., Korbmacher, C., Muller-Lantzsch, N., Lower, J., and Kurth, R. (1993). Identification of human endogenous retroviruses with complex mRNA expression and particle formation, Proc Natl Acad Sci U S A *90*, 4480-4.

Lower, R., Lower, J., and Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences, Proc Natl Acad Sci U S A *93*, 5177-84.

Maeda, N. (1985). Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene contains a retrovirus-like element, J Biol Chem *260*, 6698-709.

Maeda, N., and Kim, H. S. (1990). Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates, Genomics *8*, 671-83.

Maeda, N., McEvoy, S. M., Harris, H. F., Huisman, T. H., and Smithies, O. (1986). Polymorphisms in the human haptoglobin gene cluster: chromosomes with multiple haptoglobin-related (Hpr) genes, Proc Natl Acad Sci U S A *83*, 7395-9.

Magdinier, F., Ribieras, S., Lenoir, G. M., Frappart, L., and Dante, R. (1998). Down-regulation of BRCA1 in human sporadic breast cancer; analysis of DNA methylation patterns of the putative promoter region, Oncogene *17*, 3169-76.

Marcus, J. N., Watson, P., Page, D. L., and Lynch, H. T. (1994). Pathology and heredity of breast cancer in younger women, J Natl Cancer Inst Monogr *16*, 23-34.

Marsden, J. (2002). Hormone-replacement therapy and breast cancer, Lancet Oncol *3*, 303-11.

Martin, A. M., and Weber, B. L. (2000). Genetic and Hormonal Risk Factors in Breast Cancer, J Natl Cancer Inst *92*, 1126-1135.

Martin, J., Herniou, E., Cook, J., Waugh O'Neill, R., and Tristem, M. (1997). Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates, J Virol *71*, 437-43.

Martin, M. A., Bryan, T., McCutchan, T. F., and Chan, H. W. (1981). Detection and cloning of murine leukemia virus-related sequences from African green monkey liver DNA, J Virol *39*, 835-44.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., and et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, Science *266*, 66-71.

Negrini, M., Sabbioni, S., Haldar, S., Possati, L., Castagnoli, A., Corallini, A., Barbanti-Brodano, G., and Croce, C. M. (1994). Tumor and growth suppression of breast cancer cells by chromosome 17- associated functions, Cancer Res *54*, 1818-24.

Patience, C., Wilkinson, D. A., and Weiss, R. A. (1997). Our retroviral heritage, Trends Genet *13*, 116-20.

Pujol, P., Galtier-Dereure, F., and Bringer, J. (1997). Obesity and breast cancer risk, Hum Reprod *12 Suppl 1*, 116-25.

Repaske, R., Steele, P. E., O'Neill, R. R., Rabson, A. B., and Martin, M. A. (1985). Nucleotide sequence of a full-length human endogenous retroviral segment, J Virol *54*, 764-72.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual, 2nd edn (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory).

Samuelson, L. C., Wiebauer, K., Snow, C. M., and Meisler, M. H. (1990). Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution, Mol Cell Biol *10*, 2513-20.

Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T., and Wellstein, A. (1996). Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus, Proc Natl Acad Sci U S A *93*, 14759-64.

Schulte, A. M., and Wellstein, A. (1998). Structure and phylogenetic analysis of an endogenous retrovirus inserted into the human growth factor gene pleiotrophin, J Virol *72*, 6065-72.

Sekiya, T., Kuchino, Y., and Nishimura, S. (1981). Mammalian tRNA genes: nucleotide sequence of rat genes for tRNAAsp, tRNAGly and tRNAGlu, Nucleic Acids Res *9*, 2239-50.

Shih, A., Coutavas, E. E., and Rush, M. G. (1991). Evolutionary implications of primate endogenous retroviruses, Virology *182*, 495-502.

Smigel, K. (1995). Breast cancer death rates decline for white women [news], J Natl Cancer Inst *87*, 173.

Steele, P. E., Rabson, A. B., Bryan, T., and Martin, M. A. (1984). Distinctive termini characterize two families of human endogenous retroviral sequences, Science *225*, 943-7.

Steinberg, K. K., Thacker, S. B., Smith, S. J., Stroup, D. F., Zack, M. M., Flanders, W. D., and Berkelman, R. L. (1991). A meta-analysis of the effect of estrogen replacement therapy on the risk of breast cancer, Jama *265*, 1985-90.

Temin, H. M. (1981). Structure, variation and synthesis of retrovirus long terminal repeat, Cell *27*, 1-3.

Temin, H. M. (1991). Sex and recombination in retroviruses, Trends Genet *7*, 71-4.

Thompson, M. E., Jensen, R. A., Obermiller, P. S., Page, D. L., and Holt, J. T. (1995). Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression, Nat Genet *9*, 444-50.

Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., and Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene, Genes Dev *6*, 1457-65.

Venables, P. J., Brookes, S. M., Griffiths, D., Weiss, R. A., and Boyd, M. T. (1995). Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function, Virology *211*, 589-92.

Vogelstein, B., and Kinzler, K. W. (1994). Has the breast cancer gene been found?, Cell *79*, 1-3.

Williamson, E. A., Dadmanesh, F., and Koeffler, H. P. (2002). BRCA1 transactivates the cyclin-dependent kinase inhibitor p27(Kip1), Oncogene *21*, 3199-206.

Xu, C. F., Chambers, J. A., and Solomon, E. (1997). Complex regulation of the BRCA1 gene, J Biol Chem *272*, 20994-7.

Xu, C. F., and Solomon, E. (1996). Mutations of the BRCA1 gene in human cancer, Semin Cancer Biol *7*, 33-40.

Yan, Y., Haas, J. P., Kim, M., Sgagias, M. K., and Cowan, K. H. (2002). BRCA1-induced apoptosis involves inactivation of ERK1/2 activities, J Biol Chem *24*, 24.

Yang, X., and Lippman, M. E. (1999). BRCA1 and BRCA2 in breast cancer, Breast Cancer Res Treat *54*, 1-10.

```
   1  TAAGGGAGGAGACCACCCCTCATATTGTCTTATGCCCAATTTCTGCCTCCAAAGAAAGAAAAAGTAAAAACTAAAGGGCAGAAATGACATCCACAAGCAG
 101  ACACCCCAGGTGCCACACCCTGGGCCTGGTAGTTAAAAATCAACCCCTGACCTGATTGGTATGTTATCTATAGATTACAGACATTGTATAGAAAAGCACT
 201  GTGACAATGCCTTTCCTGTTTTGTTCCGATCTAATTACTGGTGCATGCAGCCCCAGTCACGTACCCCCTGCTTGCTCAATTGATCACGAACCTCTCACGT
 301  GCACCCCCTTAGAGTTGTGAGCCCTTAAAAGGGACAAGAATTGCTCACTCGGGGAACTTGGCTCTTGAGACCGGAGTCTTGCCGATGCCCCTGGCCTAAT
 401  AAACCCCTTCCTTCTTTAACTCGGTGTCTGATAAGTTTTGTCTGCAGCTCGATTTCTTGGTTCCCTGACTGGGAAGCGAGGTGAATGGCAGATTGTCAAG
 501  GCAGCTCTTTAGGCGGCTTAAACTTGCCTGTGGAACATCCTGCAGGGGACTCCAACCAGCCCGAGCAACACGGATCTTGAGGGCGCTCCTGGTTAGGCAT
 601  TTGCCCCATGGGACGCTTCGCCAGAGCACTGTATGGCAGGCCCCCATGGAGATCAACGCAGTGGCTGAACACCGGGAAGGAACGGGCACTTGGAGTCTTG
 701  GACATCTAAAACTTGGTAAGAGTAGTCCTTTGAAACTTGCCCACTCCATTTGAGTGGAAGCATGGCCTGATCACCCATGGCGTGCCTTTATTGGCACTTT
 801  GGTTTTGGTTTGAATTGCTTGACAGGATTGGTCTTGGGAACTTGCCTACTCCATTTGAGTGGAAGTGTGGCCTGATCACCCATGGTGTGCCTCTCCTGGC
 901  ACTTTGGTTTTTATTTTTGACTTGACTTGGATTGCTTGATACTTTGGTTTGGTTTTGACCTGGCTTGGATTTCTGGATACTCTGATTTTGGTTTGGTGCA
1001  AACTGCAAAACTGTGTGTGCGCCCTTTTTACCGGTTCTTTGTTTTGTGGTGTGTTTGTGGTGTGAGCGTGGTGTTTTGTCTTGAAGAAGCATAGGTCAGG
1101  CACGAATAAGCCCACCCCACTAGGAACTATGTTGACAAATTTCAAAAAAGAATCTGAGGGAGACTATGGAGTGCTATGACACCAGGAAAACTTAAAACTT
1201  TGTGTAAGATAGACTGGCCAGCATTAGAGGTAGGTTGGCCATTAGAAGGAAGCCTGGACAGGTCCCTTGTTTCAAAGGTATGGCACAAGGTAACCTATAA
1301  GCCAAGGAACCCAGACCAGTTCCCTTACAGAGACACTTGATTAGAGCTGGTTTTAGACCCCCCACCCCCAACACATAGTGGTTGAGAGAACAGCAGCATA
1401  AGTGGCTGGCAGAGGCAAGGAAAGACCAGCAGAGAGAGAGAAAGGAAAGAGACAGAGAGGAAAAGAGGAAAAGAGAGAGGAAGAGACAGGAAGAGACA
1501  GACAGAGAGTCAAGGATAGAGAGAGAGAGAGAGAGAGAGACACACACACACACACACACAGAGAGAGAGAGGAAGAGACAGAGGCAAAAGGAAAGTCAAA
1601  GAGAGAGACAAAGTCAAAGAGAGAAAGAAAGAGAGAAAGAGAGAGAGATACAAAGTTAAGAAAAAAAAGCCATCTATACCAATTCTAATTTAGACAAAAC
1701  AAGGTCTTATTAATAGCAAAGGATAATTAAAATCCCAAACTTACAAGGGTTTCTTTTTAGTATTATTATTATTATACTTTAAGTTTTAGGGTACATGTGC
1801  ACAACATGCATTTTTTAAATTTTTTTTAAAGGTTTTCAACTAAAGTTTGCTAAAAGTTAACAGTGTAACATGTATATAGTAACTTATATTCTTGTGGCCT
1901  AGTCCACAAGACACAAAAAAGGTTCGCTTTGGAAAAGAATGGTTATCATCTTCGAAAAAAAAAAGGAAAAAAAGGGGGGGCGGAATTTATATAAAAAGAG
2001  TATTATATGGTAAATTCTTGTCCTGAAACAAATTAACTGGTTGTTTAAAGAAATAAATATTTGTAATAAGTCAGAAAGTTGAGGCATATCGAAGAATTGT
2101  CTGCGAAAGTCACGAAAGAGAAAAATGTTATAAAAAAATTTATGCAAAAAATGTTGTATAATTTAAAAGTAACTAGGCCTCCTGAATGTAAAACTGTTGA
2201  AAAAAAAAAACAGTTTATATGCAAGGTGTATAAGAAAAGTAAAATATATCTTTGGTAAAAGAATTGTAAGGAGGCATAAGGATGTACATTTTCACCTACA
2301  TTAAGGGGTTAAAAATAATTATTGTTTTGAAGATTTAAGCAAGTTTCAAAATGTTAATTGTAAAGAAAATTCTGTGTGTAAACATATTAGCTAAAGTTAA
2401  AGAAGTATCATCCATTTTTTCTGTGAACTGGACATTAAAGTAAAAGCATAACAGGTTTTTTCTTAAAGCACCAACCTGCTCTTTTGCAAAAATTATAAAG
2501  GGTTAAAAAGAATCTATAAAATCTTACCTTATGGTCAAACATTAACAATTAGATACATATGTCTACAAGGTTTTATTAAAATTAAGTTTAATATTAATAA
2601  CACACTAATATAAAGATAAAATTTAGCTTATCGGGGTATAAAAATCATACGAGAAGCACTGTTAAATGTAAAATGGTATTTGGCTTTCTTTGGTTTAAAA
2701  ACTAATAAAAATAGGTGCTAAAGGAAATTTCTCAGGAAAAAGGCACTAAGGACTATAAAGTCCACTGCCAAGGTCCCCACATTTAAAACAAAAGGTCAGT
2801  TCCTTAGAATTATATGCTTGGTTTATTTTCCACTTTCCTTTCTCTCAAAAACTAAAAGTCTTTTAGCACATGTACCACCCCTAGAATTTCTAGTAAACCA
2901  GCACCAGCCTGAAGATCACATTCTCATCGAAAGGTGGAAAGAAGAAAAACTCGAGCCAGCCTGGAACGACCCTACCTTGTGTTGCTAACCACCAAGACTG
3001  CTGTTCGTACAGCAAAAAAGGATGGACTCGTCACACCAGAATCAAGAAAGTGCCACCCCCTCCAGAGTCATGGGCCACAGTCCCAGAGGAAAACCCTGCC
3101  AAACTAAAGCTACGAAAAATCTCTTTTCATCTATTCTATTACTCTTCTTTCTTTCCTCGTGCTATTGCTGACCATCTAGTTATTAACATAACCAAGTCAA
3201  TTTCACCTCAAACTATTGCATTTAATGCTTGCCTTGTTATACCCTGTGGGGACTTGCCAAGTCAAAGACAGCTCTCTACTTCAAGAAAAGTACTTCTGTC
3301  CCTCCTGACTCTCCTCAGACTGGACATTAGTAAACTAGGACCATTTAATCCAAGGAGATTTCGATAAAGACCCAATGCCAACCAAGGAGTCTTGTCCCCC
3401  TGATGTAGAGTTTTTATGCCGTAGTTGGTCCAATGTTCTGTGGACACTAAAGAGCAAGAATGGACTGCCCCAACCAGTTTTTGTAATTTCCTAAAACCAT
3501  ACATTCGTTTTACTAGAGGATCATAGAAGTTAAAGAGTTAAAACAAACTTTACCAATGAAGACAGGATACCAAGATGCAAATGCCTGGTTAAAATGCATC
3601  AAATATTCCATCTGCACGTTAAACAAAAGCAATTGTTATGCTTGTGCACATGGCAGGCCAGAGGCCCAGATTGTCCCCTTTCCACTAAGGTGGTCCTCCA
3701  GTCGACCAGATGTGGGCTGCCTGGTAGCTCTTTTCCAGGATTCTACAGCCTGTATTAATAAGTCATGCCAAGCTCTCTCTGCTATATCCCAAAGTCCAGC
3801  ACCCTGCGGGTCAGCCCCCGAGGGCCATCCATCCTCCATCTCCCAACACTAAGTTCACTTCATGTCTCTCACAACAGGGAGGAAACTTAGCGTTCCTTGG
3901  AGACTGAAAGGATGCAGTTAGCTTAAGAATTTCAAGAGCTTATCAATCAGTCAGCCCTTGTTCATCCCCGAGCAGATGTGTGGTGGTATTGTGGTGGACC
4001  TTTACTGGGCACTCTGCCAAATAACTGGAGTGGCACTTGTACTTTAGTCCAATTGGCTATCCCTTTCACCCTGGCATTTCATCAACCAGAAGGAAAAAAA
4101  GAAAGACATTGTAAAGCGAGAGAAGCCCCTTATGGGTCTTTCGACTCTCATGTCTATTTATATGCAATTGGAGTCCCACGAAGAATACCAGATCAATTTA
4201  AAGCTTGAAATCAAATAGCTGCAGGATTTAAGTCAATATTTTAGTGGGTGACAGTTAATAAAAATGTAGATTAGATAAACTACATCTATTACAACCAACA
```

```
4301 ACAACGAGGTTTTCATGAGTTAAAAGAAAACCTCATGTTGGTCCCAGCCCTGAGGCTACCTGACCTGACAAAACTCTTTACACTCTATGTGTCAGAAAGA
4401 GAAAAAATGGCAGTTGGAGTTTTAACCCATACTGTGAGGCCCTGGCCAAGGCCAGTGGCCTATCTCTCAAAACAACTAGACAGGATTTCCAAAGGCTGGC
4501 CCCCAGGTCTAAGGGCCCTAGCAGCAACAGCCCTGTTAGCACAAGAAGCAGATAAACTAACCCTTAGGAAAACCTGAATATAAAGGCCCCCCATGCTGTG
4601 GTAACTTTAATGACTACCAAAGGACCTCATTGGTTACAAATGCTAGATTAACCAAGTACCAAAGCTTGCTATGTGAAAATCCCCACATAACCACTGAAGT
4701 TTGCAACACCCTAAACCCCACCACCTTGCTCCTGGTATCAGAGAGCCTAGTTGAACATACTGTGTAGAGGTGTTGGACTCAGTTTATTCTAGCAGGCCCA
4801 ACCTCCGAGACTATCCTTAAACATCAGTAGACTGTGAGGGGTACGTGGACGGGAGTAGCTTCGCCAACCCCTGCAAAGTGACTCTGAAGAAGACGACAAG
4901 CTCTGCTCCAGTCACACCCAGAAGCTGACTGGTCCACGCACGGCCGAAGCATGAGAAAACTCATTGCGGAACTCATTTTCCTTAAAATTTGGACTTTTAC
5001 AGTAAGGACTTCAACTGACCTTCCTCAGACTGAGGACTGTTCCAGTGTATACATCAAGCAAGTCACTGTAGGTAGGACAAAAGGTTGCTACAGTCCTATT
5101 ATTTTATGGTTACTACAATGTGTACTGGAACTCGAAAAATAACTTGTTTGTATAATGTTATTCTGTACAAGGTATGTCGCCCATGAAGTGACCAACCTGA
5201 TGTGTGTTATGACCCAACTGAGCCTCCCATGACCACGGTTTTTAAAATAAGATTAAGGACTGAGGACTGGTGGGGGCTCACATATGATATGAGTAAAGTG
5301 TTAGGCAACACACAACATTCAGGGGTGCCCAAACAAGTCACCCTAAAATTTGAGGCCTGTGCTGTCACTAATAGTAATAAGTTAGGAATAAACTGTGGTT
5401 CCCCTAATTAGAAAAGAGGCTATATGGCATAAAATACATACATCTGTCATAAATTACGACTGTGTGGAAATAAATGTAAATACTGGTCTTGTGTCACTTA
5501 CGCCACCTGGTTTTACAAAAAATCAAAAGGATCCAGTCCACCTTCAGAAAAGAAATTCCTGGCCTTCCTGTGCTAAGGGACAATGTAACCCCTTAACAGC
5601 TAGTACAATCACTCCCCTTGATCCACGCTGGGAAAAAAGAGTAGCGTGTGACCTTAAGATTCAATGGGGCTGGGGGGCTGGACTGAATCCCGAGTAAATA
5701 TCTTGGTTCGAGGAGAAGTTTACAAATGCTCTCTTGAGCCAGTGTTTCAAACTTTCTATGATGAACTAAATGTGCCAATAACAGAATTTCCAGGAAAAAC
5801 AAGAAATTTGTTTTTGCAATTAGCGGAGCATGTAGCCCAGTCTCTCAATGTCACTTCATGTTATGTATGTGGAGGAACTTGTAATAGCAGATCAATGCCC
5901 ATTGGAAGCCCGGGAATTTGTACCTACAGACCCAGTTCCTGATGAATTCCCAGCTCAAAAGAATCACCCTGATTCTGGGTCCTAAAAGCCTCAATCATTA
6001 GACAATACTATATAGCAAGAGTGGAGAAGGACTTCTGTAGGAAGACTACATGGTGGAGTTCAAACCACACAGAGAAAAATCCATTCAGTAAATTTCCAAA
6101 GTTGCAGACCGT17AGGCCCACCCAGAATCCCACCGGGACTGGACACCCACTGGGCTATACTGGATATGTGGACATAGAGCCTACACTAAGCTGCCTGAC
6201 AAGTAGTTGTGTTATTGGCACTATTAAACCATCTTTCTTCCTACTGTCCATAAAAACAGGAGAACTCCTGGGCTTCCCTGTCTATGCTTCCCGTGAAAAG
6301 CATAGCTATAAGAAATTGAAACAATGATAAATGGCCCCCTGAGAGAATCATACAGTATTATGGGCCTGCTACTTAGGCACAAGATGGCTCGTGGGGATAC
6401 CGGATCCCCATTTATATGATCAACCGAATCATACGGTTACAAGCTGTCTTAAAAATAATCACTGCAACCGGCAGAGCCTTGACTATTCTTGGCCCAGCAA
6501 GAAACTCAGATGAGAAATGCTATCTATCAAAATAGATTAGCTCTCGACTACTTGCTAGCAGCTGAAGGAGAGGTCTGTAGGAAATTTAACCTTACTAATT
6601 GCTGCCTACACATAGATAACCAAGGGCAAGTAGTTGAAGACATAGTTAGAGATATGACAAAAGTGGCACATGTGCCCGTGCAAGTGTGGCATGGATTTGA
6701 TCCTGGGGCCATGTTTAGAAAATGGTTCCCAGCGCTAGGAGGATTTAAAACTCTTATAATAAGAGTTATAATAGTAATAGGAACCTACTTACTGCTCCCT
6801 CGTTTGCTACCTGTACTTCTTCAAATGATAAAAAGCTTCATCGCTACCTTAGTGTACCAAAATGCATCAGCACAAGTGTACTATATAAATCACTATTGAT
6901 CTGTCTAGCAAGAAGACATGGGTAGTAAACATGAAACAACTCCTGCTATTGAGTGAGAGTCTCAAAGTGGGGGAATAAGGGAGGAGACCACCCCTCATAT
7001 TGTCTTATGCCCAATTTCTGCCTCCAAAGAAAGAAAAAGTAAAAACTAAAAGGCAAGAAATGAAATCCACAAAGCAGACAGCCCGGGCGCCACANCCTGG
7101 GACTGGGTAGTTTGAAAATCAATCCCGGACCTAATCGGTTATGTTATCTATAAGATTACAGACATTGTATAGAAAAGCATTGTGAAAATCTCATTCCTGT
7201 TTTGTCCAATCTAATACTGGTGCGTGCAGCCCCCAGTCACGTACCCACGTACCTCCTGCTTGCTCAATCGATCAGACCCTCTCACGCACACCCCCTTAGA
7301 CTTGTGAGTCCTTAAAAGGGACAGGAATTGCTCACTCGGGGCGGGGGCTCGGCTCTTGAGACAGGAGTCTTGCTGATGCCTCCGGCCTAATAAACCCCTT
7401 CCTTCTTTAATCGGTGTCTGAGGAGTTTTGTCTGCAGCTCGTCCTGCTACAGGTG
```

Acknowledgement

The practical work of this thesis was performed from September 4th 1998 to April 1st 1999 in the laboratory of:

Anke M. Schulte, Ph. D.
Assistant Professor
Institute of Oncology
Lombardi Cancer Center, NRB E 312
Georgetown University Medical Center
Reservoir Road 3970 NW
Washington, DC 20007-2179
USA

I hereby wish to thank Anke for allowing me to be part of her lab and for teaching me the techniques in molecular biology needed for this research project. It was an honor for me to work with her, since she was an excellent teacher not only in science, but in understanding the relevant aspects and approaches to work on a project. She made me gain an enthusiasm in this fascinating field of research and working with her was as challenging as it was fun. Thank you Anke for being patient enough with me and for all your support!

I also wish to thank my parents, family and friends for their constant support. It is very encouraging to have someone who backs you up, nevertheless what situation you are in.

Additionally I want to show appreciation for all the people that contributed to this thesis either by their moral or by their scientific expertise and support. I especially want to point out Dr. Claudius Malerczik, Dr. Achim Aigner and Prof. Dr. Frank Czubayko. I thank you for answering the numerous questions I had and for the hours you spent helping me to correct this thesis, first in Georgetown and then in Marburg. Additionally I would like to mention Christine Cottichia, Violaine Harris, Ph.D. and Jason with whom I had the pleasure to work with.